

Essays in Forecasting and Empirical Macroeconomics

Saeed Zaman

Presented in the fulfilment of the requirements
for the degree of Doctor of Philosophy

Department of Economics
University of Strathclyde

August 2021

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree. The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

Acknowledgements

This research was supported financially by the Federal Reserve Bank of Cleveland. I am very grateful for this support. I am highly indebted to Ellis Tallman and Edward Knotek II, my colleagues at the Cleveland Fed, for their valuable support and encouragement throughout the PhD studies.

This thesis would not have been possible without the excellent support, guidance, and encouragement from my advisors, Professor Gary Koop (primary) and Professor Julia Darby (secondary). I found Gary's instructional and writing style quite inspirational. As a result, I learned a lot from him, including Bayesian econometrics and, more broadly, the necessary intellectual rigor to successfully carry out empirical research. I sincerely thank Professor Koop and Professor Darby for the valuable feedback on this thesis's chapters, particularly chapter 4. And I look forward to maintaining a research relationship with them. I would also like to thank Dr. Alex Dickson, who provided prompt help with administrative matters related to travels between the US and the UK.

I am very grateful to Professor Ana Galvão and Dr. Aubrey Poon for their thoughtful and helpful comments on my thesis.

Many people have directly or indirectly (via discussions) provided comments that I found beneficial in shaping my thesis chapters, and here I acknowledge them. I thank Todd Clark, Kurt Lunsford, Elmar Mertens, James Mitchell, Randal Verbrugge, Benjamin Wong, and Willem Van Zandweghe for detailed comments and suggestions on chapter 4 of this thesis. In addition, I am grateful to Grant Allan, Fabio Canova, Siddhartha Chib, Olivier Coibion, Sharada Davidson, Jesus Fernandez-Villaverde, Luca Gambetti, Greg Ganics, Domenico Giannone, Yuriy Gorodnichenko, Dimitris Korobilis, Nimantha Manamperi, Mike McCracken, Stuart McIntyre, Juan Rubio-Ramirez, Shaun Vahey, and Ping Wu.

I also acknowledge helpful comments from the participants at various conferences and seminars that include, 11th and 12th International Conference on Computational and Financial Econometrics, International Institute of Forecasters' 38th International Symposium on Forecasting, 2nd Central Bank Forecasting Conference at the Bank of England, 3rd Central Bank

Forecasting Conference at the Bank of Canada 2019, Midwest Econometrics Group Meetings 2018 and 2019, Fall 2018 Midwest Macroeconomics Meetings, 94th Annual Conference of the Western Economics Association, the 2019 Conference on Real-Time Data Analysis, Methods and Applications, the 2020 Joint Statistical Meetings, the Society for Nonlinear Dynamics and Econometrics 28th Annual Symposium, ESCoE Conference on Economic Measurement 2021, the 6th RCEA Time Series Econometrics workshop, 27th International Conference Computing in Economics and Finance, NBER-NSF SBIES 2021, the Federal Reserve System Day-Ahead Inflation workshop, and the Ohio University.

I am grateful to Andrea Carriero and Todd Clark for sharing Matlab code estimating large BVAR with stochastic volatility. I thank Knut Aastveit for sharing his Matlab code for density forecasts with MIDAS models. I have benefitted from the Matlab code on optimal density combinations using density calibration metrics shared on Greg Ganig's website, density calibration diagnostics shared on Tatevik Sekhposyan's website, and modeling, unobserved components models shared on Joshua Chan's website. Lastly, chapter 3 has benefitted from the PRObability FORecasting (PROFOR) Matlab toolbox developed by researchers at the Norges Bank, the Bank of England, and the University of Warwick.

I am also very grateful to my family, including my parents, my wife, Ammara, and my two beautiful daughters, Aliza and Minahil, who have been my constant source of motivation and drive to complete this dissertation in a reasonable amount of time. I am genuinely indebted to Ammara, who at times, which were many, had to single-handedly care for the kids so that I can continue to make progress on the dissertation.



Finally, because I am a central banker, the usual disclaimer applies: The views expressed herein are mine and or of my co-authors (where applicable) and do not necessarily represent the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

Statement of Contribution of Co-Authors

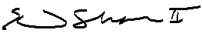

Whereas Chapter 4 is solo authored, Chapters 2 and 3 (of my dissertation) are joint work. Chapter 2 is co-authored with my colleague at the Cleveland Fed, Ellis W. Tallman, and that chapter is now published in the International Journal of Forecasting. Chapter 3 is co-authored with my Cleveland Fed colleague, Edward Knotek II, and is available as a Cleveland Fed working paper. The undersigned hereby certify that:

1. they meet the criteria for co-authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the working paper in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the working paper;
3. there are no other authors of the publication according to these criteria; and
4. there are no potential conflicts of interest that require to be disclosed to (a) granting bodies, (b) the editor or publisher of any journal or other publication.

In the case of **Chapter 2**, contributions to the work involved the following:

Contributor	Statement of contribution	Publication title and date of publication or status
Ellis W. Tallman	Co-writing;	Tallman, E.W. and Zaman, S.
 Signature	Methodology; Validation;	(2020). Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy, <i>International Journal of Forecasting</i> , 36(2): 373-398 https://doi.org/10.1016/j.ijforecast.2019.04.024
Saeed Zaman	Conceptualization;	
 Signature	Co-writing; Methodology; Data Curation; Coding and all estimation;	

In the case of **Chapter 3**, contributions to the work involved the following:

Contributor	Statement of contribution	Publication title and date of publication or status
Edward S. Knotek II  Signature	Conceptualization; Co-writing; Validation;	Knotek, E.S., II and Zaman, S. (2020). Real-Time density nowcasts of US inflation: A model-combination approach. Federal Reserve Bank of Cleveland, Working Paper No. 20-31. https://doi.org/10.26509/frbc-wp-202031
Saeed Zaman  Signature	Conceptualization; Co-writing; Methodology; Data Curation; Coding and all estimation;	

Abstract

This thesis consists of three self-contained essays that contribute to the literature on macroeconomic forecasting and empirical macroeconomics. The first essay establishes the importance of good starting conditions (i.e., nowcasts) and terminal conditions (i.e., steady-states or "stars") in obtaining accurate forecasts from vector autoregressive (VAR) models estimated with quarterly data. It does so by proposing the technique of relative entropy to tilt the VAR forecast both in the near term with the survey nowcast and in the long run with the survey long-run projection. Doing so leads to meaningful gains in multi-horizon forecast accuracy. The gains in accuracy are made possible because our proposal is an indirect approach to accommodating structural change and moving end points.

The second essay develops a framework based on the model and density combinations that generate highly accurate point and density nowcasts of inflation at a daily frequency. We adopt a novel flexible treatment in the use of the aggregation function to combine density estimates from a range of mixed-frequency models. The framework permits dynamic model averaging via weights that are updated based on learning from past performance. Together these features allow non-Gaussian densities. The accuracy of the density and implied point nowcasts are significantly more accurate than the nowcasts from the survey of professional forecasters.

The third essay develops a large-scale unobserved components model to estimate a range of macroeconomic stars (i.e., terminal points). The model is motivated by economic theory and empirical features such as time-varying parameters and stochastic volatility. The model allows for a direct link between the model-based star and long-run survey expectations, which significantly improves the precision of the model-based estimates of stars. The by-products are the time-varying estimates of the wage and price Phillips curves, passthrough between prices and wages, which provide new insights into these empirical relationships' instability in the US data.

Contents

1	Introduction	10
1.1	Motivation	10
1.1.1	Importance of reasonable steady states	11
1.1.2	Estimating steady states (stars)	12
1.1.3	Importance of good nowcasts and their role in forecasting	13
1.2	Contributions and Unifying Themes	14
1.2.1	Contributions of this thesis	14
1.2.2	Unifying themes across chapters	16
1.2.3	Layout of chapters and status	17
2	Combining Survey Long-Run Forecasts and Nowcasts with BVAR Forecasts Using Relative Entropy	19
2.1	Introduction	19
2.2	Data and the Empirical Model	22
2.2.1	Data	22
2.2.2	Bayesian VAR models	24
2.3	Real-time Long-horizon Survey Forecasts versus BVAR Forecasts	29
2.4	Methodology for Tilting Forecasts	31
2.4.1	Relative entropy	31
2.4.2	Determining the forecast horizon for tilting	34
2.5	Results	36
2.5.1	Forecasting exercise	36
2.5.2	Inflation forecast accuracy of the tilted VAR compared to univariate benchmarks	44
2.6	Conclusions	46
3	Real-Time Density Nowcasts of US Inflation: A Model-Combination Approach	59
3.1	Introduction	59
3.2	Mixed-Frequency Models	62
3.2.1	Deterministic Model Switching (DMS)	62

3.2.2	Mixed Data Sampling (MIDAS)	64
3.2.3	Dynamic Factor Model (DFM)	65
3.2.4	Mixed-Frequency Model Space	65
3.3	Combination Methods	66
3.3.1	Functional Forms for Aggregation	66
3.3.2	Weighting Schemes	68
3.4	Real-Time Data	71
3.5	Nowcast Evaluation	72
3.6	Empirical Results Using Real-Time Data	74
3.6.1	Density Nowcasts from Mixed-Frequency Model Classes	75
3.6.2	Comparison across Grand Combinations	76
3.6.3	Comparing the Grand Combination with Its Underlying Component Densities	78
3.6.4	Time-Varying Properties of the Grand Combination: Weights, Uncertainty, Skewness, and Kurtosis	79
3.7	Comparison with the Survey of Professional Forecasters	80
3.8	Conclusion	81
4	A Unified Framework to Estimate Macroeconomic Stars	97
4.1	Introduction	97
4.2	Empirical Macro Model and Variants	101
4.2.1	The econometric notion of a long-run equilibrium	101
4.2.2	The role of survey expectations	103
4.2.3	Unemployment block	104
4.2.4	Output block	107
4.2.5	Productivity block	109
4.2.6	Price inflation block	111
4.2.7	Wage inflation block	114
4.2.8	Interest rate block	116
4.2.9	Base model and its variants	119
4.3	Data and Bayesian Estimation	119
4.3.1	Data	119
4.3.2	Bayesian estimation	120
4.4	Full Sample Estimation Results	122
4.4.1	Estimation results for u-star	122
4.4.2	Estimation results for g-star and the output gap	126
4.4.3	Estimation results for p-star	130
4.4.4	Estimation results for π -star	132
4.4.5	Estimation results for W-star	135
4.4.6	Estimation results for r-star	139

4.5	Real-time Estimates and Forecasting	143
4.6	The implications of COVID-19 Pandemic on Stars	147
4.7	Conclusion	148
5	Conclusions	165
5.1	Summary of contributions and policy implications	165
5.2	Further research	169
	Bibliography	170
A	Chapter 2 Appendix	185
A0.	Technical appendix	185
A1.	Results for Medium VARs	190
A2.	Density Forecast Calibration Diagnostics	198
A3.	Density Forecast Assessment based on Log-Score metric	203
A4.	Ranking the Models: Before tilting vs. Post-tilting	206
A5.	Effect of Tilting on Small VAR with loose priors	209
A6.	Are there benefits to utilizing survey information for additional horizons?	211
A7.	Hybrid vs. Federal Reserve’s GreenBook	215
A8.	Hybrid vs. Time-Varying VAR (Baseline)	219
A9.	Steady-State BVAR vs. Hybrid (Small BVAR est. 1960)	222
A10.	BVAR in Gaps vs. Hybrid (Small BVAR est. 1960)	225
A11.	Evolution of Long-Run Forecasts from Other Surveys	228
A12.	Gaussian example: Illustrating the Spillover Effects of Tilting	230
A13.	Sampling from Tilted Predictive Density: Multinomial Algorithm	231
A14.	Evolution of Forecast Horizons for Tilting	231
A15.	Sensitivity to Getting the Horizon Wrong	233
B	Chapter 3 Appendix	239
B1.	Description of Mixed-Frequency Models and Simulation Procedures	239
	B.1.1. MIDAS Model	239
	B.1.2. DFM Model	240
	B.1.3. DMS Model	242
B2.	Mechanics of Density Combination and Graphical Illustration	248
B3.	Comparing Properties of Grand Combinations across Weighting Schemes	249
C	Chapter 4 Appendix	273
C1.	Bayesian Estimation Details	273
	C1.a. Base Model equations	273
	C1.b. Prior Elicitation	276
	C1.c. MCMC Algorithm	278

C1.d Marginal likelihood computation	318
C2. Prior Sensitivity Analysis	318
C3. MCMC Convergence Diagnostics	320
C4. Additional Forecasting Results: Base vs. Benchmarks	323
C5. Additional Forecasting Results: Steady-State BVAR, Base stars vs. Survey	325
C6. Additional Real-time Estimates Stars	329
C7. Estimated relationship between Survey and Stars	332
C8. Additional COVID-19 Pandemic Results	333
C9. Backcast: Survey R^* from 1959-1982	339
C10. R^* : Additional Full Sample Results	341
C10.a. Role of data vs. prior in shaping r-star	341
C10.b. Base vs. External models	342
C10.c. Sensitivity of r-star to the prior setting	342
C10.d. The usefulness of the Taylor-rule equation	344
C11. Π^* : Additional Full Sample Results	345
C11.a. Pi-star comparison Base vs. outside models	345
C11.b. Sensitivity of pi-star to modeling assumptions	347
C11.c. Pi-star estimates for some variants of the Base model	349
C12. P^* : comparison with Kahn and Rich (2007)	350
C13. P^* : Additional Full Sample Results	353
C13.a. Cyclical Productivity based on Output gap	353

List of Figures

2.1	Real-time Long-run Forecasts	48
2.2	Results of the Small VAR est. 1960	49
2.3	Results of the Small VAR est. 1985	50
2.4	Results of the Small TVP-VAR SV	51
3.1	PITs across Single Specifications of Mixed-Frequency Model Classes	88
3.2	PITs across Stage 1 Combinations within Model Classes	89
3.3	PITs across Grand Combinations	90
3.4	Density Performance Comparisons across Grand Combinations	91
3.5	Point Nowcasting Performance across Grand Combinations	92
3.6	Density Performance of Grand Combination vs. Its Components	93
3.7	Point Nowcasting Performance, Grand Combination vs. DMS	94
3.8	Weights and Higher-Order Moments (CPI and Core CPI)	95
3.9	Weights and Higher-Order Moments (PCE and Core PCE)	96
4.1	Overview of Interactions Between Blocks	150
4.2	Full Sample Estimates for Unemployment Rate block	154
4.3	Full Sample Estimates for Cyclical Unemployment	155
4.4	Full Sample Estimates for Output block	156
4.5	Model-based Output Gap vs. CBO	156
4.6	Full Sample Estimates for Productivity block	157
4.7	Full Sample Estimates for Price Inflation block	158
4.8	Full Sample Estimates for Nominal Wage block	159
4.9	Wage Persistence, Wage Phillips Curve, Passthrough from Prices	160
4.10	Full Sample Estimates for Interest Rate block	161
4.11	More Estimates for Interest Rate block	162
4.12	Real-time Recursive Estimates of Stars: Base model	163
A1	GreenBook vs. Baseline and Hybrid (Small VAR 1960)	217
A2	Real-Time Long Run Forecasts	229
A3	Forecast Horizons at which model combined with long-run survey	232

A4	Univariate approach: Forecast Horizons at which model combined with long-run survey	234
B1	Illustration: Linear vs. Log Opinion Pools	250
B2	Example Stage 1 DMS Combination	251
B3	Single Specifications vs. Stage 1 Combinations	252
B4	Real-Time Density Nowcasts	253
B4	Real-Time Density Nowcasts (continued)	254
B5	Weights Underlying Grand Combination: Ganics Weighting Scheme	255
B6	Grand Combination vs. Its Components: Month-Over-Month Inflation	256
B7	Point Nowcasting Performance, Grand Combination vs. DMS: Month-Over-Month Inflation	257
B8	Density Performance of Grand Combination vs. Its Components: Quarterly Inflation	258
B9	Point Nowcasting Performance, Grand Combination vs. Other Combinations and Single DMS Specification: Quarterly Inflation	259
B10	Weights for Stage 1 DMS Combinations, Log Score Weighting Scheme	260
B11	Weights for Stage 1 DFM Combinations, Log Score Weighting Scheme	261
B12	Weights for Stage 1 MIDAS Combinations, Log Score Weighting Scheme	262
B13	Time-Varying Uncertainty Estimates for Density Nowcasts of Quarterly Inflation	263
B14	Stage 2 Grand Combination of DMS, DFM, and MIDAS Combinations	264
B15	Weights and Higher-Order Moments, CMG Weighting Scheme	265
B15	Weights and Higher-Order Moments, CMG Weighting Scheme (continued)	266
B16	Weights and Higher-Order Moments, Ganics Weighting Scheme	267
B16	Weights and Higher-Order Moments, Ganics Weighting Scheme (continued)	268
B17	Weights and Higher-Order Moments, CRPS Weighting Scheme	269
B17	Weights and Higher-Order Moments, CRPS Weighting Scheme (continued)	270
B18	Comparison of Weights within the DMS Model Class, Log Score Weighting Scheme vs. CMG Weighting Scheme	271
C1	MCMC Diagnostics of Base Model	321
C2	MCMC Diagnostics of Base-NoSurv Model	322
C3	Real-time Recursive Estimates of Output Gap: Base Model	329
C4	Real-time Recursive Estimates of Stars: Base-NoSurv Model	330
C5	Real-time Recursive Estimates of Stars: Base vs. Base-NoSurv	331
C6	Estimated Link Between Survey Forecasts and Stars	332
C7	Estimates of Stars pre- vs. post-COVID Recession	336
C8	Estimates of Stars pre- vs. post-COVID Recession (more)	337
C9	Estimates of Stars post-COVID Recession: Base vs. Outside	338
C10	Survey R* and G*	340

C11	R* estimates	343
C12	The Usefulness of Taylor Rule equation	344
C13	Pi* estimates: Base vs. External models	346
C14	More Estimates for Price inflation block	348
C15	Pi* estimates: Base vs. Base model variants	349
C16	P* consistent with narrative from 2-Regime Markov-Switching Model	352
C17	Base-P*CycOutputGap model	353

List of Tables

2.1	Out-of-sample Point Forecasting Performance: Small BVAR est. 1960.	52
2.2	Out-of-sample Density Forecasting Performance: Small BVAR est. 1960.	53
2.3	Out-of-sample Point Forecasting Performance: Small BVAR est. 1985.	54
2.4	Out-of-sample Density Forecasting Performance: Small BVAR est. 1985.	55
2.5	Out-of-sample Point Forecasting Performance: Small TVP-VAR.	56
2.6	Out-of-sample Density Forecasting Performance: Small TVP-VAR.	57
2.7	CPI inflation Real-time Out-of-sample Point Forecasting Performance.	58
3.1	Model Space: Mixed-Frequency Model Classes and Specifications	83
3.2	Representative Dates for Monthly Nowcasting Performance	83
3.3	Calibration Diagnostics	84
3.4	Calibration Diagnostics (continued)	86
3.5	Nowcasting Comparison with the Survey of Professional Forecasters	87
4.1	Description of Model Specifications	151
4.2	Parameter Estimates	152
4.3	Bayesian Model Comparison: Main Models and selected variants	153
4.4	Model Comparison: Variants focused on unemployment rate	153
4.5	Model Comparison: Variants focused on GDP	153
4.6	Model Comparison: Variants focused on labor productivity	153
4.7	Model Comparison: Variants focused on nominal wages	153
4.8	Model Comparison: Variants focused on price inflation	158
4.9	Model Comparison: Variants focused on interest rate	162
4.10	Real-Time Forecasting Accuracy: Base vs. Base-NoSurv	164
A1	Out-of-Sample Point Forecasting Performance: Medium BVAR est. 1960	191
A1	Cont.: Out-of-Sample Point Forecasting Performance: Medium BVAR est. 1960 .	192
A2	Out-of-Sample Density Forecasting Performance: Medium BVAR est. 1960 . . .	193
A2	Cont. Out-of-Sample Density Forecasting Performance: Medium BVAR est. 1960	194
A3	Out-of-Sample Point Forecasting Performance: Medium BVAR est. 1985	195
A3	Cont.: Out-of-Sample Point Forecasting Performance: Medium BVAR est. 1985 .	196
A4	Out-of-Sample Density Forecasting Performance: Medium BVAR est. 1985 . . .	197

A5	Calibration Assessment of Density Forecasts: Small VAR est. 1960	199
A6	Calibration Assessment of Density Forecasts: Small VAR est. 1985	200
A7	Calibration Assessment of Density Forecasts: Small TVP-VAR	201
A8	Calibration Assessment of Density Forecasts: Medium VAR est. 1960	202
A9	Density Forecasting using Log-score metric: Small BVAR est. 1960	204
A10	Density Forecasting using Log-score metric: Small BVAR est. 1985	205
A11	Density Forecasting using Log-score metric: Small TVP-VAR	206
A12	Forecast Accuracy Assessment: Before Tilting (Raw) and Post-Tilting (Hybrid)	208
A13	Forecast Accuracy Assessment: Before Tilting (Raw) and Post-Tilting (Hybrid): Added Small VAR with loose priors	210
A14	Real-Time Forecasting Performance: Hybrid vs. Survey of Professional Forecasters	214
A15	Real-Time Forecasting Performance: Hybrid vs. Green Book	218
A16	Real-Time Forecasting Accuracy Hybrid (from Small VAR) vs. TVP-VAR SV . .	221
A17	Real-Time Out-of-Sample Forecasting Performance: Small BVAR (Hybrid) vs. Steady-State BVAR	224
A18	Real-Time Out-of-Sample Forecasting Performance: Small BVAR (Hybrid) vs. Small BVAR in Gaps (Baseline)	227
A19	Out-of-Sample Point Forecasting Performance: Small BVAR est. 1960 (horizon determined using univariate approach)	235
A20	Out-of-Sample Point Forecasting Performance: Small BVAR est. 1985 (horizon determined using univariate approach)	236
A21	Out-of-Sample Point Forecasting Performance: Small BVAR est. 1960 (horizon dogmatically set at 7yrs)	237
A22	Out-of-Sample Point Forecasting Performance: Small BVAR est. 1960 (horizon dogmatically set at 10yrs)	238
B1	Representative Dates for Quarterly Nowcasting Performance	272
C.1	Prior Settings	276
C2	Parameter Estimates: Comparison	319
C3	Out-of-Sample Forecasting Performance: Base vs. Benchmarks	324
C4	Out-of-Sample Forecasting Performance: Steady-State BVAR	328

Chapter 1

Introduction

1.1 Motivation

Macroeconomic forecasts play a crucial role in the conduct of monetary policy. Monetary policy is assumed to affect the economy with long and variable lags, and so policymakers need to be forward-looking, which necessitates the reliance on forecasts. Over time, research has shown the usefulness of clear communications and transparency in the effectiveness of the monetary policy. An important component of clear communications involves communicating the outlook to the public to support policy actions. Accordingly, central banks now routinely publish their forecasts and give regular press conferences to explain their policy actions. In explaining their actions, they often refer to their projections. The publication of the forecasts and their focal role in the deliberations of policy discussions and as a communication device has increased considerable interest in research related to forecasting and nowcasting macroeconomic variables.

In an influential work, Faust and Wright (2013) illustrate that in forecasting US inflation, two critical ingredients to obtaining accurate forecasts for inflation are good “jumping-off” point (i.e., an accurate nowcast) and a reasonable terminal point (i.e., steady-state). Similarly, Wright (2013) shows using a vector autoregressive (VAR) model that the same two key ingredients, accurate nowcast and reasonable estimate of steady-state matter, for improving the forecast accuracy for a host of other macroeconomic variables in addition to inflation. Thus, in a nutshell, “good” nowcasts and “reasonable” steady states are crucial ingredients to obtaining accurate macroeconomic forecasts, a point also emphasized by Wright (2019). In addition, surveys of professional economists, such as Survey of Professional Forecasters (SPF) and Blue-Chip Economic Indicators (BC), two well-known surveys in the case of the United States, have shown to be reliable sources for obtaining estimates of both the nowcasts and the (time-varying) steady states.

1.1.1 Importance of reasonable steady states

Macroeconomic forecasters often use atheoretical models for forecasting. For example, Banbura, Giannone, and Reichlin (2010) and Koop (2013) show that fixed-parameter large VARs containing more than 100 variables can work effectively, a finding that has contributed to a resurgence in the use of VARs in forecasting and policy analysis by both central banks and private forecasters.

Although computational advances and developments in Bayesian estimation methods have made solving time-varying parameter VAR models convenient, constant parameter medium-scale VAR models remain popular due to their ability to generate accurate forecasts. More important, Aastveit et al. (2017) show that the forecasting accuracy of constant parameter medium-scale VAR is competitive with both a small-scale time-varying parameter BVAR with stochastic volatility (similar to Primiceri 2005), with a medium-scale time-varying BVAR (built along the lines of Koop and Korobilis, 2013) and with a regime-switching VAR (Barnett, Mumtaz, and Theodoridis 2014). This finding lends credibility to the use of medium-scale fixed-parameter VAR models for forecasting, especially given their computational ease relative to the alternatives.

In fixed-parameter VAR models, the unrestricted long-run forecasts converge to or nearly to the unconditional mean of the estimation sample (i.e., implied trend or terminal points), which at times differs substantially from economists' views of the long-run values for particular variables. Contributing to this divergence in views is the use of history for model estimation that may reflect an outdated characterization of the macroeconomic relationships, including the unconditional means. Furthermore, the entire forecast trajectory would be biased in the direction of the implied trend estimated in the model over the full sample. This is because the implied trend of the model increasingly influences the forecasts beyond five quarters. Therefore, a badly estimated trend (i.e., the steady-state) is the primary source of forecast errors for medium-term forecasts (as was emphasized by Clark and McCracken, 2008; Clements and Hendry, 1999).

The long-run forecasts in published surveys of professional forecasters have shown to be reasonable proxies for the underlying trends (steady states) because they adjust to any exogenous and or underlying shifts in the economy more quickly than the unrestricted long-run model forecasts (see Kozicki and Tinsley, 1998; Wright, 2013). The quicker adjustment of the survey expectations stems from the fact the survey participants have at their disposal indicators that typically are not included in the information set fed into the models, such as information about the value of the inflation target, central bank communication, and demographic factors.¹ The fact the survey long-run projections are reasonable proxies of steady states motivates Wright (2013) to rely on survey projections as estimates of the steady-states for his VAR model. The

¹It is worth noting that even if these indicators were included as part of the model's information set, the time-invariant models such as VARs would extrapolate forward the trends prevailing over the entire estimation sample, which may not necessarily align with recent developments.

insights of Wright (2013) motivate our **chapter 2**, and the results we find echo many of those reported by Wright. However, imposing steady states on VAR models using the approach of Wright (2013) requires specifying VAR in a certain way, i.e., deviation from means. As I discuss shortly, chapter 2 proposes an alternate (arguably more flexible) approach to influencing the VAR model(s) implied steady states. Specifically, the technique of relative entropy is utilized to alter the medium- to long-horizon VAR forecast to match the real-time survey long-horizon forecast.

The interest in the estimates of steady states of macroeconomic variables, often denoted with “stars” notation, goes beyond forecasting. Their estimates are used for a variety of purposes. For example, let’s take the steady-state unemployment rate (u-star); its estimates or equivalently of the potential real GDP (gdp-star) are used to infer the economy’s cyclical position. The information about the economy’s cyclical position is input in a range of policy decisions, business decisions about future spending, and the construction of cyclical indicators. For example, fiscal agencies use the information of steady states to separate fiscal balances into cyclical and structural components. Similarly, an assessment of the cyclical position is essential in constructing cyclical indicators, such as cyclical inflation measures (e.g., Stock and Watson, 2020; Zaman, 2019). And combined with information about estimates of the steady-state real rate of interest (r-star) and steady-state inflation (pi-star), it plays a direct role in the conduct of monetary policy.

1.1.2 Estimating steady states (stars)

The long-term expectations reported in surveys are a potential proxy for stars. However, to infer the estimates of the stars, macroeconomists have applied a range of statistical and econometric methods to observable historical data. These methods range from univariate statistical filters to multivariate models, including semi-structural time-series models and fully structural dynamic stochastic general equilibrium (DSGE) models. Economic theory posits that the structural aspects of the economy, which inform the values of the stars, change slowly. Therefore, methods that produce estimates of stars that change only gradually have more traction than methods that give less smooth estimates. According to this criterion, multivariate unobserved components (UC) models, which are statistical models that use economic theory to frame the empirical specification, have been shown to provide reasonable estimates of the stars (e.g., Kuttner, 1994; Laubach and Williams, 2003; Chan, Koop, and Potter, 2016). Hence, they are the dominant methods for obtaining time-varying estimates of the stars.

However, with few exceptions, the popular multivariate UC models that provide estimates of time-varying stars focus on a small number of observables, often just two or three, and have a minimal structure (e.g., Laubach and Williams, 2003). As I discuss later, **chapter 4** takes on the challenge of jointly estimating several macroeconomic stars using a semi-structural time series model (aka multivariate UC model with a particular structure informed from economic theory). In principle, proceeding with the joint estimation of a framework that explicitly models the

objects of interest and permits interactions among them (e.g., stars) can provide more reliable estimates of the objects compared to approaches that ignore them.

1.1.3 Importance of good nowcasts and their role in forecasting

Like long-run survey projections, which have good forecasting properties, the nowcast estimates of macroeconomic variables contained in surveys have shown to have good nowcasting properties. However, since the seminal work of Giannone, Reichlin, and Small (2008) on nowcasting GDP, a growing literature has focused on developing quantitative methods, which utilizes high-frequency data in conjunction with low-frequency data, to nowcast macroeconomic variables. In this literature, the success of the proposed method is typically assessed relative to the survey nowcasts, in that whether the nowcast accuracy of the proposed method matches the survey. If it does, then the method is favorably viewed because there are advantages to using a model-based approach to computing nowcasts. There is an implicit recognition in the literature that survey point nowcasts are hard to outperform, so methods that rival surveys and or hard to beat univariate benchmarks are viewed favorably (e.g., Carriero, Clark, and Marcellino, 2015 in the case of real GDP). And methods that outperform surveys attract greater attention and become the new baseline for subsequent studies (e.g., Knotek and Zaman, 2017 in the case of inflation).

Forecasters have at their disposal a rich set of empirical models to produce macroeconomic forecasts. Often these models are estimated with quarterly data to match the frequency of key variables such as real GDP. However, forecasts are produced at a frequency greater than once per quarter to account for the flow of high-frequency information. The standard practice to updating outlook involves augmenting model-based quarterly forecasts with updated nowcasts informed from external sources. These include surveys, which are updated more than once per quarter, or external nowcasting models that explicitly link high-frequency data to low-frequency variables, e.g., real GDP; see Del Negro and Schorfheide (2013), and Knotek and Zaman (2019).

More recently, Kruger, Clark, and Ravazzolo (2017) show that conditioning quarterly forecasting models on both the nowcast mean and variance, i.e., uncertainty around the mean, leads to improvements in multi-horizon forecast (point and density) accuracy compared to the typical approach of conditioning with nowcast mean only. This improved accuracy translates into an improved assessment of probability events (and balance of risks around point forecasts), such as the risk of a recession or probability of deflation in the next X quarters or so.

The past two decades have seen considerable growth in density forecasting and density nowcasting literature. However, recent work on density nowcasting (based on mixed-frequency models) has generally focused on real GDP growth and other indicators of real economic activity (e.g., Aastveit et al., 2014; Carriero et al., 2015). As discussed shortly, one of the contributions of this thesis (i.e., **chapter 3**) is to provide a modeling framework to produce density (and point) nowcasts of inflation at a trading day frequency.

1.2 Contributions and Unifying Themes

1.2.1 Contributions of this thesis

In light of the discussion above, which emphasizes the usefulness of macroeconomic forecasts and the estimates of the stars as inputs into the forward-looking monetary policy setting, this thesis seeks to improve forecasting and allow for a more credible assessment about the “location” of the stars. Specifically, the thesis develops frameworks to generate accurate forecasts for a range of macroeconomic variables, highly accurate nowcasts of inflation measures, and credible estimates of stars, which are of broader interest to macroeconomists. The forecasts and estimates of stars from these frameworks are shown to be competitive to popular benchmarks. The expectation is that the frameworks developed in this thesis would prove valuable to anyone interested in macroeconomic forecasting, including central banks, private institutions, fiscal agencies, and the public. We note that even though the empirical focus of the thesis is on US data, the modeling frameworks developed are general enough and so can be applied more broadly. We expect that the contributions of the essays will appeal to a broad group of empirical macroeconomic researchers, and we find some evidence of this effect via increasing citations to the working paper versions of the chapters.

Chapter 2 is titled “Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy.” This chapter constructs hybrid forecasts that combine forecasts from vector-autoregression (VAR) model(s) with both short and long-term expectations from surveys. Specifically, we use relative entropy to tilt one step ahead, and long-horizon VAR forecasts to match the nowcast and long-horizon forecast from the Survey of Professional Forecasters. Kruger, Clark, and Ravazzolo (2017) [KCR] use relative entropy to generate conditional forecasts with moment conditions that match survey forecasts using the short-term forecast from the survey as the mean condition on a one-step-ahead VAR forecast. They construct variance conditions around their mean conditions using ex-post real-time survey nowcast errors. The second chapter of this thesis extends the work of KCR: in addition to tilting the VAR forecast in the nowcast quarter, we also tilt the medium-to long-horizon forecast from the VARs to match the long-horizon forecast reported in the external survey of forecasters, the SPF. We consider a variety of VAR models ranging from simple fixed-parameter to time-varying parameters. The results across models indicate meaningful gains in multi-horizon forecast accuracy relative to model forecasts that do not incorporate long-term survey conditions. Accuracy improvements are achieved for a range of variables, including those not directly tilted but affected by spillover effects from tilted variables. The accuracy gains for hybrid inflation forecasts from simple VARs are substantial, statistically significant, and competitive to time-varying VARs, univariate benchmarks, and survey forecasts. We view our proposal as an indirect approach to accommodating structural change and moving endpoints.

We note that the by-product of our chapter is a rich set of forecasting results confirming evidence documented in earlier papers: (1) Banbura, Giannone, and Reichlin (2010) and Koop

(2013), in that medium-scale VARs generate more accurate forecasts than small-scale VARs; (2) Clark (2011) and Koop and Korobilis (2013), in that allowing for stochastic volatility helps improve forecast accuracy; (3) D’Agostino, Giannone and Gambetti (2013), in that stochastic volatility helps more than time-varying parameters in forecasting macroeconomic variables; (4) Kruger, Clark, and Ravazzolo (2017), conditioning on nowcast mean and nowcast variance leads to spillover effects in the form of improved accuracy for further out horizons.

Chapter 3 is titled “Real-Time Density Nowcasts of US Inflation: A Model-Combination Approach.” Recent work on density nowcasting (based on mixed-frequency models) has generally focused on real GDP growth and other indicators of real economic activity. This chapter develops a flexible modeling framework to produce density nowcasts for US inflation at a trading-day frequency. Our framework: (1) combines individual density nowcasts from three classes of parsimonious mixed-frequency models; (2) adopts a novel flexible treatment in the use of the aggregation function; and (3) permits dynamic model averaging via the use of weights that are updated based on learning from past performance. Together these features provide density nowcasts that can accommodate non-Gaussian properties. We document the competitive properties of the inflation nowcasts generated from our framework using high-frequency real-time data over the period 2000-2015. As shown in chapter 2 and KCR, conditioning macroeconomic models with nowcast means and nowcast densities (informed from external sources) lead to improvements in the accuracy of multistep point and density forecasts, especially inflation. The framework developed in chapter 3 provides a potential source of nowcast mean and nowcast density estimates for inflation.

Chapter 4 is titled “A Unified Framework to Estimate Macroeconomic Stars.” This chapter takes on the challenge of developing a flexible semi-structural time series model to estimate jointly several macroeconomic stars, i.e., unobserved long-run equilibrium levels of output (and growth rate of output), unemployment rate, the real rate of interest, productivity growth, price inflation, and wage inflation. For each star, we formulate a rich structure whose elements are motivated by economic theory and, in part, by the empirical features necessitated due to the changing economic environment. Following recent literature on inflation and interest rate modeling, we explicitly model the links between survey long-run expectations and stars to improve the stars’ econometric estimation. We allow for time variation in the important macroeconomic relationships and the error variances (aka stochastic volatility). Clark (2011), among many others, highlights the importance of allowing for stochastic volatility in macroeconomic models. Permitting time-varying relationships between model components will potentially lead to more credible estimates and are arguably less susceptible to the Lucas critique. The by-products are the time-varying estimates of the wage and price Phillips curves, passthrough between prices and wages, changing procyclicality of labor productivity, and persistence in price and wage inflation dynamics, which provide new insights into the instability of these empirical relationships in US data. To tractably estimate our large multivariate model, we use a recently developed precision sampler that relies on Bayesian methods. Generally, the contours of our stars echo

those documented elsewhere in the literature, which employs smaller models, but at times are different, and these differences can matter for policy. Furthermore, our estimates of the stars are among the most precise. Lastly, we document the competitive real-time forecasting properties of our model and, separately, the stars' estimates if they were used as steady-state values in external models.

1.2.2 Unifying themes across chapters

Even though chapters 2, 3, and 4 of this dissertation are independent, they all share several themes: improving inflation prediction and or trend measurement, accounting for model uncertainty, providing role to survey expectations, and modeling changing economic environment.

The three chapters of the thesis contribute to inflation forecasting and trend estimation: (1) the contribution of chapter 2 includes proposals to generate highly accurate (point and density) forecasts of inflation using a range of VAR models; (2) Chapter 3 focuses on producing high quality (point and density) inflation nowcasts using a range of mixed-frequency models; and (3) the contribution of chapter 4 includes production of credible estimates (both point and precision) of trend inflation using UC models.

In macroeconomic analysis, accounting for model uncertainty is crucial to the credibility of the analysis. In the literature, several strategies have been entertained to account for model uncertainty, and the same is the case for the three chapters of this thesis. In chapter 2, the efficacy of the proposed method to construct hybrid forecasts (i.e., the combination of model and survey forecasts) is examined across a range of VAR models, about 10 in total reflecting the numerous models in use. In chapter 3, model uncertainty is accounted for by combining the estimates of density nowcasts from many different models to construct a combined (composite) estimate. In chapter 4, model uncertainty is highlighted by estimating the baseline model and fifteen additional model specifications to assess the empirical support for numerous features embedded in the baseline model.

Another unifying theme is the role of survey projections. In both the second and fourth chapters, the information from the survey of professional economists plays a crucial role. In the second chapter, the model-based forecasts for a select number of variables are augmented in the near-term and the longer-term with survey expectations, resulting in improved forecast accuracy (via spillover effects) for all the forecast horizons considered and for a range of macroeconomic variables. Hence, information from surveys is an essential input to the forecasting models. In the fourth chapter, survey long-term expectations play a vital role in improving the precision and plausibility of the model-based estimates of stars, i.e., the long-run equilibrium values. In the third chapter, the inflation nowcasts reported in the survey (SPF) – or estimated using the variance of the past historical errors – are treated as a benchmark; the nowcasting accuracy of the framework developed is assessed relative to this benchmark.

There is widespread recognition that macroeconomic relationships evolve, and so methods that either implicitly or explicitly account for these changing relationships have greater appeal.

The strategy employed to account for the changing economic relationships varies across the three chapters. In chapter 2, our (hybrid) proposal, which anchors model-based forecasts towards survey nowcasts and long-term projections, is a post-estimation procedure to accommodating structural change and moving endpoints in the model-based forecasts. The fact that survey participants have access to a large information set, the patterns gleaned from it can help shape their opinions, including any perceived structural change, which can immediately influence their expectations about the long-run. In chapter 3, we combine evidence (i.e., density nowcasts) from a range of models that in principle should provide insurance in the face of “*future uncertainties*” about specifications as it will be more robust to “*structural instabilities*” (see Durlauf and Vahey, 2010). In empirical studies, approaches that combine estimates from a rich set of models in a dynamic fashion (i.e., where weights assigned to individual models are allowed to vary over time) explicitly recognize that economic relationships, which models seek to capture, are evolving. And at any given point in time, some models would do a better job than others. In chapter 4, the modeling framework developed explicitly models time-variation in parameters assumed to capture important macroeconomic relationships and error variances.

1.2.3 Layout of chapters and status

The three subsequent chapters are packaged such that they are self-contained. Each chapter has its introduction, results with tables and figures, and footnotes. The corresponding supplementary appendices are relegated to the end. In addition, all the references corresponding to each of the chapters are compiled into one section labeled Bibliography. As of June 2021: chapter 2 is published in the International Journal of Forecasting; chapter 3 is under review and is available both as a Cleveland Fed working paper and a Strathclyde discussion paper; chapter 4 is under preparation for submission to a journal.

Chapter 2 (tilting VAR forecasts to surveys) is a co-authored work with Ellis W. Tallman. I did most of the work, including writing the first draft and all the empirical work. My co-author, Ellis, assisted with valuable edits to improve the exposition of the draft. Ellis taught me the technique of relative entropy by meticulously going over his Matlab code from his paper, Robertson, Tallman, and Whiteman (2005). Chapter 3 (density nowcasting inflation) is co-authored work with Edward Knotek II. Again, I did most of the work and wrote the first and second drafts of the chapter. My co-author, Edward, who is particular about writing, contributed by editing and validating the draft to streamline the paper’s exposition. This chapter substantially builds on my earlier work with Edward on inflation nowcasting (Knotek and Zaman, 2017). Edward proposed the need to construct density estimates around our point nowcasts from our earlier work. How to go about doing it, i.e., nowcasting via combining density nowcasts from many different models, is what I contributed to in the conceptualization phase. I was familiar with the density combination work by the Norges bank economists (e.g., Aastveit et al., 2014) and by Anthony Garratt, James Mitchell, and Shaun Vahey (2014). I thought that applying their methods to density nowcasting inflation (using mixed-frequency models)

would be a fruitful avenue. Chapter 4 (joint estimation of macroeconomic stars), which is solo authored, is under preparation for submission to a journal and soon will be available as a Cleveland Fed working paper and a Strathclyde discussion paper.

Chapter 2

Combining Survey Long-Run Forecasts and Nowcasts with BVAR Forecasts Using Relative Entropy

Based on the paper: Tallman, E.W. and Zaman, S. (2020). Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy, *International Journal of Forecasting*, 36(2): 373-398

2.1 Introduction

Macroeconomic forecasters often use atheoretical models for forecasting. Banbura, Giannone, and Reichlin (2010) show that large fixed-parameter VARs that contain more than 100 variables can work effectively, a finding that has contributed to a resurgence in the use of VARs for forecasting and policy analysis by both central banks and private forecasters.

This paper proposes a technique for adjusting the forecasts of the implied trends from a VAR toward (plausible) values proposed by judgmental forecasters. Specifically, we utilize the technique of relative entropy to alter the medium- to long-horizon VAR forecast to match the real-time survey long-horizon forecast.

The long-run forecasts contained in published surveys of professional forecasters are reasonable proxies for the underlying trends because they adjust to any exogenous and/or underlying shifts in the economy more quickly than the unrestricted long-run model forecasts (e.g. Kozicki & Tinsley, 1998; Faust & Wright, 2013; Wright, 2013). The quicker adjustment of the survey expectations stems from the fact the survey participants have at their disposal indicators that typically are not included in the information set fed into the models, such as information about the value of the inflation target, central bank communication, and demographic factors.

In fixed-parameter VAR models, the unrestricted long-run forecasts converge to or nearly to the unconditional mean of the estimation sample (i.e., implied trend or ‘end points’ in the

terminology of Kozicki & Tinsley, 1998), which at times differs substantially from economists' views of the long-run values for particular variables. Contributing to this divergence in views is the use of history for model estimation that may reflect an outdated characterization of the macroeconomic relationships, including the unconditional means. Furthermore, the entire forecast trajectory would be biased in the direction of the implied trend estimated in the model over the full sample. This is because the forecasts beyond five quarters are increasingly influenced by the implied trend of the model, and therefore a badly estimated trend is the primary source of forecast errors for medium-term forecasts (as was emphasized by Clements & Hendry, 1999; Kozicki & Tinsley, 2001a,b; Clark & McCracken, 2008).

Econometricians have addressed the misspecification issues that arise from structural changes by introducing various innovations to standard VAR models, such as time-varying parameters (e.g. Cogley & Sargent, 2005; Primiceri, 2005). These models appear to do well for forecasting variables that exhibit notable structural shifts (e.g. D'Agostino, Giannone, & Gambetti, 2013). These models have become popular as the computational demands of operating them have become less binding, due to both the availability of greater computing power and the introduction of newer methods. The latter advantage has also allowed the potential for estimating these models with larger information sets (e.g. Koop & Korobilis, 2013). However, these models are complex (as there are so many moving parts involved) and require a certain level of sophistication, which limits their wider use.

Some practitioners in the US and other advanced economies estimate VAR models using data starting in 1985, i.e., after the well-documented structural change. Empirical evidence documenting the good forecasting ability of VAR estimated with shorter samples (e.g. Aastveit, Carriero, Clark, & Marcellino, 2017) supports this practice.

We account for the numerous models that are in use by considering the efficacy of our proposal in a range of VAR models. Specifically, we consider a constant-parameter VAR estimated with both a longer sample (starting in 1959) and a shorter sample (post-1985), and a time-varying VAR. We consider model specifications both with and without stochastic volatility, and perform our assessment on both small-scale and medium-scale VARs. All in all, our model space consists of 10 models.

We use relative entropy to combine the VAR forecasts with the forecasts reported in the Survey of Professional Forecasters (SPF). The technique has gained widespread use in combining model-based forecasts with external information since its application to economic forecasting by Robertson, Tallman, and Whiteman (2005). This increasing usage stems mainly from its ease of use, computational ease, and flexibility; it allows the forecaster to combine appropriately both the mean condition and the modeler's confidence in that mean condition (i.e., variance), as illustrated by Krüger, Clark, and Ravazzolo (2017; KCR hereafter).

KCR use relative entropy to generate conditional forecasts with moment conditions that match survey forecasts using the short-term forecast from the survey as the mean condition on a one-step-ahead VAR forecast. They construct variance conditions around their mean

conditions using ex-post real-time survey nowcast errors (using a rolling window over a pre-forecast evaluation sample). This chapter extends the work of KCR: in addition to tilting the VAR forecast in the nowcast quarter, we also tilt the medium- to long-horizon forecast from the VARs to match the long-horizon forecast reported in the external survey of forecasters, the SPF. Our methodology parallels that of Altavilla, Giacomini, and Ragusa (2017), who also use the relative entropy to tilt the segments of the yield curve forecasts from the term structure models so as to match survey expectations.¹

Two popular approaches to influencing VAR forecasts directly using survey long-run forecasts are: (1) a steady-state BVAR (developed by Villani, 2009), which Wright (2013) uses to show that prior beliefs regarding the unconditional mean of the variable that are informed by a survey’s (Blue Chip) long-run forecasts lead to systematic improvements in forecast accuracy for a range of U.S. macroeconomic variables, especially for inflation; (2) modeling some or all VAR variables in ‘gaps’, where gaps are computed as deviations from the trends informed by either survey long-run projections or univariate moving average methods (e.g. Kozicki & Tinsley, 2001a,b; Clark & McCracken, 2010; Zaman, 2013). The crucial difference between our approach and these existing approaches is that our approach influences the forecasts’ post-model estimation. The insights of Wright (2013) motivate our study, and the results in this chapter echo many of those reported by Wright.²

Our main question of interest is whether we can achieve any meaningful gains in the forecast accuracy of the VAR variables over the forecast horizons that are of interest to policymakers by forcing the medium-term to long-horizon forecasts of a select number of VAR variables to match the published survey forecasts. Essentially, the forecast of interest is a hybrid forecast that consists of a *survey nowcast*,³ a *BVAR forecast*, and a *survey long-horizon forecast*.

Our empirical forecast evaluation results provide evidence of notable improvements in both the point and density forecast accuracies of hybrid VAR forecasts for several macroeconomic variables. All of the models in our set benefit from being combined with survey information; not surprisingly, the accuracy gains are largest for the model specifications estimated with a longer sample and smallest for time-varying VARs, but all gains are statistically significant. This result indicates that tilting helps to mitigate misspecification issues more in models that are thought to have a higher degree of misspecification (caused in part by badly estimated unconditional means). On a related note, we find that all models benefit from tilting in the post-crisis period, a period that is associated with structural change (see Aastveit et al., 2017).

The accuracy gains are achieved not only for variables that are tilted directly, but impor-

¹The empirical application in this chapter fits within the broader context in the use of relative entropy for ‘theory-coherent forecasting’, as proposed by Giacomini and Ragusa (2014). In our case, the use of survey long-run forecasts as a proxy for trends can be viewed as adjusting the forecasts from atheoretic VAR models toward long-run equilibrium values that are informed partly by economic theory.

²In a related work, Giannone, Lenza, and Primiceri (2019) propose a natural conjugate prior (denoted PLR) for influencing the joint long-run behavior of the VAR variables modeled in levels.

³There is a long list of papers documenting the usefulness of nowcasts for helping improve the forecasting accuracy for future horizons (e.g., KCR; Knotek & Zaman 2019).

tantly also for variables that are influenced indirectly through the spillover effects of the tilted variables (e.g. wage inflation and payroll employment). Over the forecast horizons that are of interest to monetary policymakers (i.e., 1 to 12 quarters ahead), the gains in forecast accuracy are strongest for inflation and the federal funds rate. We infer that the apparent structural break in inflation and the extended experience of the federal funds rate near the zero lower bound (ZLB) are the crucial challenges, especially in VAR models estimated with fixed coefficients.

We summarize additional findings as follows. First, we show that the point forecast accuracy of hybrid CPI inflation forecasts from our modeling approach is competitive with those of tough-to-beat univariate benchmark models. The forecasts are also competitive with those of SPF and the Federal Reserve’s Greenbook. Our results make a practical contribution to policymakers. Monetary policymakers desire to use multivariate model(s) which allow for feedback effects from policy to the real economy and inflation. However, often these multivariate models are unable to match the forecasting performances of the univariate forecasting models. The application in this chapter provides one potential path.

Second, we show that hybrid forecasts generated from simple VARs estimated with post-World War II data are competitive with the forecasts generated from computationally-intensive time-varying VARs. For inflation, the hybrid forecasts are more accurate and the gains are statistically significant. This result is of practical importance for practitioners who are reluctant to use time-varying VARs for a variety of reasons, including computational reasons, complexity issues, and shorter data samples (a primary concern in the case of developing and emerging market economies).

The chapter is structured as follows. The next section describes the data and empirical models. Section 2.3 compares the survey long-run forecasts with the BVAR model long-run forecasts. Section 2.4 details our methodology for generating the hybrid forecasts. Section 2.5 reports the forecasting results. Section 2.6 concludes. Supplementary sets of results are reported in the companion appendix A.

2.2 Data and the Empirical Model

2.2.1 Data

Our empirical examination uses real-time data at a quarterly frequency. Our model space contains small VARs (denoted Small VAR) that consist of five variables and medium-sized VARs (denoted Medium VAR) that consist of 10 variables. The Small VAR consists of real GDP, the CPI, the unemployment rate, the effective federal funds rate, and the credit spread (defined as the difference between the yield on Baa corporate bonds and the yield on the 10-year Treasury note). Both real GDP and the CPI enter in annualized quarterly rates, and the remaining three variables (the unemployment rate, the effective federal funds rate, and the credit spread) are defined in units of percentage points. We use a Small VAR because several papers on VAR forecasting employ it as a benchmark VAR that contains core variables

that are of interest to central bankers. The Medium VAR adds to the Small VAR variables that have been shown to be useful in improving forecasts of the core variables. Forecasts of the additional variables may be of their own interest to central bankers, such as productivity growth and wage inflation measures, for example. Specifically, these additional five variables include real personal consumption expenditures, nonfarm business productivity, the employment cost index–wage and salary of private workers (ECI), nonfarm payroll employment, and the core CPI (i.e., the CPI excluding food and energy). All of these variables are transformed to quarterly annualized growth rates. We compute the growth rates using 400 times the log difference formula.

We construct our real-time quarterly data set using both the Federal Reserve Bank of Philadelphia’s real-time data set for macroeconomists and the Federal Reserve Bank of Saint Louis’s ALFRED database. Quarterly data on financial variables, which are real-time by construction, are downloaded from Haver Analytics.

All vintages of real-time quarterly data coincide with the survey date of the SPF, which is a quarterly survey that is released approximately in the middle of the second month of each quarter. That is, each forecast origin coincides with the SPF survey release date, so whatever quarterly data are available are utilized for estimation. For example, consider the forecast origin in February 2016: for real GDP, the advance estimate for 2015.Q4 would be used (reported by the Bureau of Economic Analysis in the last week of January 2016); for monthly variables such as the unemployment rate, it will be the second estimate for 2015.Q4 (reported by the Bureau of Labor Statistics in early February 2016). The real-time vintages start in 1994.Q1 and end in 2016.Q4. In each real-time vintage, the data sample begins in 1959.Q4; in the 1994.Q1 vintage, the data sample ends in 1993.Q4; in the last vintage (i.e., 2016.Q4), the data sample ends in 2016.Q3. The first real-time forecasting run is performed with estimation data ending in 1993.Q4, and out-of-sample forecasts are generated one to twelve quarters ahead, i.e., for 1994.Q1 to 1996.Q4.

For the purpose of forecast evaluation, we treat the ‘truth’ as the third quarterly estimate (as per Tulip, 2009).⁴

We collect the SPF nowcasts and SPF long-horizon forecasts for real GDP growth, CPI inflation, and the unemployment rate. The SPF does not report nowcasts or long-run forecasts for the federal funds rate, but it does report long-run forecasts for the 3-month Treasury bill. Accordingly, we treat the long-run forecast for the 3-month bill as the long-run estimate of

⁴For example, in the case of real GDP, the third estimate for 2015.Q4 would correspond to the actual value available as of early 2016.Q2. Similarly, for the unemployment rate, which is a monthly variable, the third quarterly estimate for the reference quarter (e.g. 2015.Q4) would coincide with the third revision to the final month of the quarter (e.g. December 2015), which will be roughly the first week of the second quarter following the reference quarter (e.g. 2016.Q2). We emphasize that our results for relative scores would have been similar both qualitatively and quantitatively if we had treated the ‘truth’ as the latest available estimate instead. For the results for the Medium VAR, please refer to the working paper version of this chapter (Tallman & Zaman, 2018). We use Haver Analytics to collect the most revised quarterly data for forecast evaluation (i.e., the vintage available as of August 2017).

the federal funds rate.⁵ For the nowcast, we construct an estimate using the intra-quarterly daily data on the federal funds rate along with the simple daily random walk model of Knotek and Zaman (2019).⁶ We collect the daily data starting from January 1, 1994, to December 31, 2016, using Haver Analytics. For illustrative purposes, we also collect the long-run projections from Blue Chip Economic Indicators (and Blue Chip Financial Forecasts), the Federal Open Market Committee’s (FOMC) Summary of Economic Projections (SEP). For the formal forecast evaluation exercises reported in the chapter, we use nowcasts and long-horizon forecasts from the SPF.

Finally, we assess the accuracy of the Federal Reserve Board’s Greenbook forecasts by collecting forecasts for real GDP, CPI inflation, and the unemployment rate (from the January 1994 FOMC meeting to the December 2012 meeting) from the Philadelphia Fed’s website (Real-Time Data Research Center).

2.2.2 Bayesian VAR models

All empirical examinations in this chapter use VAR models estimated with Bayesian methods. We consider both fixed-parameter and time-varying parameter models (with and without stochastic volatility). Following a long list of papers on forecasting with VARs, the lag order (p) is set to four in the case of fixed-parameter VARs and two in the case of time-varying VARs. Our model specifications follow Clark and Ravazzolo (2015).

VAR with constant volatility

This model is defined as

$$Y_t = A_0 + \sum_{i=1}^p A_i Y_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma), \quad (2.1)$$

where $t = 1, \dots, T$, Y_t is an $n \times 1$ vector of n observed variables, A_0 is an $n \times 1$ vector of intercepts, A_1, \dots, A_p are $n \times n$ matrices of coefficients, and ε_t is an $n \times 1$ vector of error terms distributed normally with zero mean and variance-covariance matrix $\Sigma = E\varepsilon_t\varepsilon_t'$.

For estimation details, please refer to the technical appendix A0.

VAR with stochastic volatility

The stochastic volatility process is modeled using the estimation procedure of Carriero, Clark, and Marcellino (2016). The algorithm assumes a Kronecker structure for the multivariate

⁵Historically, there is a small gap between the two, with the federal funds rate averaging roughly 30 basis points higher than the 3-month Treasury bill on a quarterly basis.

⁶The procedure involves using the available daily reading as of the SPF survey date to fill the missing trading days of the quarter. The average of the daily readings (which includes the daily data and the random walk forecast) within the quarter constitutes our nowcast estimate.

stochastic volatility, and thus makes it feasible computationally to estimate large VAR models.

$$\begin{aligned}
Y_t &= A_0 + \sum_{i=1}^p A_i Y_{t-i} + \varepsilon_t \\
\varepsilon_t &= B^{-1} \Lambda_t^{0.5} \mu_t, \quad \mu_t \sim N(0, I_n), \quad \Lambda_t \equiv \text{diag}(\lambda_{1,t}, \dots, \lambda_{n,t}) \\
\log(\lambda_{j,t}) &= \log(\lambda_{j,t-1}) + e_{j,t}, \quad j = 1, n \\
e_t &\equiv (e_{1,t}, \dots, e_{n,t})' \sim N(0, \Phi) \\
\Sigma_t &\equiv \text{Var}(\varepsilon_t) = B^{-1} \Lambda_t B'^{-1},
\end{aligned} \tag{2.2}$$

where B is a lower triangular matrix with ones on the main diagonal and nonzero elements below it; $\lambda_{1,t}, \dots, \lambda_{n,t}$ are the diagonal elements of the matrix Λ_t , representing the time-varying variances of the shocks, and are assumed to evolve according to a geometric random walk; the variance-covariance matrix Φ of innovations e_t is assumed to be of full rank, i.e., correlations among innovations of different equations is permitted; and the Σ_t is the variance-covariance matrix of the reduced form residuals ε_t .

For complete estimation details, please refer to [Carriero et al. \(2016\)](#).

Time-varying parameters VAR with constant volatility

The specification follows the same setup as per [Cogley and Sargent \(2005\)](#), but allows only regression coefficients (including the intercepts) to be time-varying, and so does not allow stochastic volatility.

$$\begin{aligned}
Y_t &= X_t' A_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma) \\
A_t &= A_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, Q),
\end{aligned} \tag{2.3}$$

where $X_t = I_n \otimes [1, Y_{t-1}', \dots, Y_{t-p}']$, and $A_t = \text{vec}(A_0, A_1, \dots, A_p)$ is a $n(k) \times 1$ column vector. The VAR coefficients stacked in the vector A_t are assumed to evolve independently according to a random walk with shocks ϵ_t that are permitted to be correlated across equations (i.e., Q is of full rank).

For estimation details, please refer to [Koop and Korobilis \(2010\)](#).

Time-varying parameters VAR with stochastic volatility

The specification follows exactly the setup laid out by Primiceri (2005), but the estimation procedure follows Del Negro and Primiceri (2015).

$$\begin{aligned}
Y_t &= X_t' A_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \Sigma) \\
A_t &= A_{t-1} + \epsilon_t, & \epsilon_t &\sim N(0, Q) \\
\beta_t &= \beta_{t-1} + \nu_t, & \nu_t &\sim N(0, R) \\
\varepsilon_t &= B_t^{-1} \Lambda_t^{0.5} \mu_t, & \mu_t &\sim N(0, I_n), \quad \Lambda_t \equiv \text{diag}(\lambda_{1,t}, \dots, \lambda_{n,t}) \\
\log(\lambda_{j,t}) &= \log(\lambda_{j,t-1}) + e_{j,t}, & j &= 1, n \\
e_t &\equiv (e_{1,t}, \dots, e_{n,t})' \sim N(0, \Phi) \\
\Sigma_t &\equiv \text{Var}(\varepsilon_t) = B_t^{-1} \Lambda_t B_t'^{-1},
\end{aligned} \tag{2.4}$$

where B_t is a lower triangular matrix with ones on the main diagonal and non-zero elements below it; β_t is a column vector that stacks (by row) off-diagonal and non-zero elements of matrix B_t , and is assumed to evolve according to a random walk with innovations that are permitted to be correlated across equations; VAR coefficients stacked in the vector A_t are assumed to evolve independently according to a random walk with shocks that are permitted to be correlated across equations; $\lambda_{1,t}, \dots, \lambda_{n,t}$ are the diagonal elements of the matrix Λ_t that represent the time-varying variances of the shocks, and are assumed to evolve according to a geometric random walk; the variance-covariance matrix Φ of innovations; e_t is assumed to be of full rank, i.e., correlations among innovations of different equations are permitted; and Σ_t is the variance-covariance matrix of the reduced form residuals ε_t .

Model set

VAR models are commonly estimated using either a longer history (e.g. post World War II data) or a shorter history (e.g. post-1985). Accordingly, we estimate our set of VAR models using both a longer sample (1959.Q4 onwards) and a shorter sample (1985.Q1 onwards). As was discussed earlier, we consider both smaller (5 variables) and medium-sized (10 variables) specifications. Thus, we estimate a VAR with constant volatility and a VAR with stochastic volatility for both smaller and medium-sized versions. For time-varying VARs, only a small-scale three-variable version (containing real GDP growth, CPI inflation, and the unemployment rate) estimated over a longer sample is examined. All in all, our model space consists of 10 VAR models:⁷

- Small VAR est. 1960: fixed-parameter small VAR estimated with data from 1959.Q4

⁷We choose not to examine our proposal on nonlinear VAR models (e.g. threshold VARs and Markov-switching VARs) for three reasons: (1) several recent papers have shown empirically that, at best, the forecasting performances of these models are competitive to those of time-varying VARs (e.g. Barnett, Mumtaz, & Theodoridis, 2014; Alessandri & Mumtaz, 2017; Aastveit et al., 2017); (2) they are less popular for forecasting than the models we consider; and (3) to keep our results manageable.

onwards

- Small VAR SV est. 1960: fixed-parameter small VAR with SV estimated with data from 1959.Q4 onwards
- Small VAR est. 1985: fixed-parameter small VAR estimated with data from 1985.Q1 onwards
- Small VAR SV est. 1985: fixed-parameter small VAR with SV estimated with data from 1985.Q1 onwards
- Medium VAR est. 1960: fixed-parameter medium-sized VAR estimated with data from 1959.Q4 onwards
- Medium VAR SV est. 1960: fixed-parameter medium-sized VAR with SV estimated with data from 1959.Q4 onwards
- Medium VAR est. 1985: fixed-parameter medium-sized VAR estimated with data from 1985.Q1 onwards
- Medium VAR SV est. 1985: fixed-parameter medium-sized VAR with SV estimated with data from 1985.Q1 onwards
- TVP VAR: time-varying small VAR estimated with data from 1959.Q4 onwards
- TVP VAR SV: time-varying small VAR with SV estimated with data from 1959.Q4 onwards.

For TVP VAR and TVP VAR SV, the first 10 years of data (i.e., 1959.Q4 to 1969.Q3) are used as a training sample for prior elicitation.

Forecast evaluation metrics

The models defined above are estimated using standard Markov chain Monte Carlo (MCMC) methods. For details of the precise algorithms, we refer the reader to the papers cited under the description of each model.

At each forecast origin, the estimation of each model involves simulating the model with D draws.⁸ For each posterior draw, the simulated forecast path is constructed by iterating the model forward h quarters (where $h = 1, \dots, 40$). The D forecast paths constitute the multivariate predictive density. The point forecast is simply the (posterior) mean of the empirical predictive density.⁹

We perform a real-time out-of-sample forecasting evaluation using a recursively-expanding estimation window. The forecast evaluation sample spans the period 1994.Q1 to 2016.Q4. We

⁸ $D = 20,000$ for time-invariant VAR model specifications; $D = 40,000 + 4,000$ (burn-in) for all model specifications with stochastic volatility and or time-varying parameters; the results are very similar if we instead use 20,000 for burn-in.

⁹In the case of time-varying VARs, following the common practice, the point forecast is defined as the median of the predictive density.

evaluate point forecasts using the widely-used metric of the mean squared forecast error (MSE),

$$MSE_{i,h} = \frac{\sum_{t=T_0}^{T_1-h} (Y_{i,t+h} - \hat{Y}_{i,t+h})^2}{T_1 - h - T_0 + 1}, \quad (2.5)$$

where i corresponds to the macroeconomic variable of interest (e.g., real GDP growth), T_0 denotes one quarter prior to the start of the evaluation period (e.g., 1993.Q4), T_1 is the end of the evaluation period (2016.Q4), $Y_{i,t+h}$ is the actual realization, and $\hat{Y}_{i,t+h}$ is the forecast.

We evaluate the performances of density forecasts using the continuous ranked probability score (CRPS) metric, as proposed by Gneiting and Raftery (2007). CRPS measures how close the actual realization is to the predictive density: the closer the distribution is to the actual realization, the smaller the CRPS value and the more accurate the predictive density.¹⁰ Accordingly, it is defined as

$$CRPS_t^{i,h}(Y_{i,t+h}) = \int_{-\infty}^{\infty} (F(z) - 1 \{Y_{i,t+h} \leq z\})^2 dz = E_p |Y_{i,t+h}^* - Y_{i,t+h}| - 0.5 E_p |Y_{i,t+h}^* - Y_{i,t+h}^+|, \quad (2.6)$$

where $Y_{i,t+h}$ is the actual realization, F is the cumulative distribution function that corresponds to the predictive density f , $1 \{Y_{i,t+h} \leq z\}$ is an indicator function that takes a value of one if $Y_{i,t+h} \leq z$ and a value of zero otherwise, and $Y_{i,t+h}^*$ and $Y_{i,t+h}^+$ are independent random draws from $p(Y^{T+1,T+H}, \theta | Y^T)$.

The CRPS metric favors predictive densities that have higher probabilities near the actual realization. As defined above, a lower CRPS is preferable to a higher score. We report the average of the CRPS, computed over our forecast evaluation period.

We assess the statistical significance of the differences in forecast accuracy between the baseline and hybrid forecasts following KCR and Altavilla et al. (2017). We use the Diebold and Mariano (1995) and West (1996) test of equal predictive accuracy for pairwise comparisons of the RMSE using the two-sided tests of the standard normal. In computing the test, we use the HAC variance estimator (an input into the test statistic) with the lag $h - 1$ truncation parameter and adjust the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); see Clark and McCracken (2013). As was emphasized by KCR, the use of our test statistics based on standard normal critical values is likely to be on the conservative side, and should be treated as an approximation that, in our case, deals with issues such as the nesting of forecasts and conditional forecasting (see Clark & McCracken, 2017).

The density calibration (i.e., absolute accuracy) of the density forecasts is assessed using interval forecasts (i.e., 70% prediction intervals), and assessments of probability integral transforms (PITS) via a battery of statistical tests: those of Knüppel (2015), Berkowitz (2001), and Kolmogorov-Smirnov. For the latter, all three statistical tests generally point to similar infer-

¹⁰An assessment based on the log-score metric largely confirms the inference obtained from CRPS, and the results are reported in Section A3 of the supplementary appendix.

ences, and therefore we report only the results from the Knüppel test. Due to space limitations, the density calibration results are reported in Section A2 of the supplementary appendix.

2.3 Real-time Long-horizon Survey Forecasts versus BVAR Forecasts

In the United States, the SPF and the Blue Chip Economic Indicators are the two most widely known and easily available routinely-published forecast surveys. The SPF is a quarterly survey that is released in the middle of the second month of each quarter. Each SPF survey release contains forecasts of the macroeconomic and financial variables for the current quarter (i.e., nowcasts) and up to four quarters ahead. For the survey carried out in the third quarter of the year, the SPF asks respondents for their estimates of the natural rate of unemployment. Similarly, for the surveys carried out in the first quarter, the SPF asks respondents for their projections of long-run values (defined as the 10-year annual average) of real GDP growth, the short-term interest rate (i.e., yield on the 3-month Treasury bill), and a few other variables. The SPF contains the projections for all of the core set of variables of interest for this chapter; for our purposes, we treat the SPF's projections of the natural rate of unemployment as the long-run forecast of the unemployment rate. Following the forecasting literature, we use the median projection for the SPF projections and the mean projection for the Blue Chip. The evolution of the long-run forecasts across the two surveys is fairly similar, and therefore, for the sake of brevity, this section focuses only on the SPF, with estimates from the Blue Chip being relegated to the supplementary appendix for interested readers.¹¹

Our forecast evaluation exercises in this chapter use the nowcasts and long-run forecasts from the SPF.¹²

The four panels of Figure 1 plot the real-time evolution of the macroeconomic long-run forecasts for real GDP growth (upper left), the unemployment rate (upper right), CPI inflation (lower left), and the short-term interest rate (lower right). Each panel plots forecasts from the small VAR estimated using the longer history (1960+), the small VAR estimated over the shorter sample (1985+), a small time-varying VAR with SV, and the SPF. The sample period is from 1994.Q1 to 2016.Q4, matching our forecast evaluation sample.

The real-time data that we use to estimate our VAR models is a subset of the information set that would be available to professional forecasters. When coming up with their forecasts, the professional forecasters would probably rely on a larger information set that included judgmental opinions of their own and of the subject matter experts, along with their own econometric methodologies. The use of forward guidance by the central bank and, more generally, the era of

¹¹The supplementary appendix A11 plots the estimates of the long-run values from the FOMC's SEP alongside the Blue Chip estimates.

¹²Our choice of the SPF is due in part to the fact that it is available publicly. On a related note, Croushore (2010) documents the good inflation forecasting properties of long-run forecasts of CPI inflation from the SPF and suggests using them as a proxy for inflation expectations.

more predictable monetary policy (through central bank communications) since the beginning of the financial crisis are other examples of important information that will be at the disposal of the survey participants. As a result, survey participants (collectively) are likely to have more timely and better informed long-run projections. Figure 1 provides some evidence in support of this claim.

The figure indicates two notable observations. First, the long-run forecasts from the VAR estimated with a longer history adjust very sluggishly, whereas the survey projections fluctuate considerably more. Second, the time-varying VAR that is built to accommodate structural change explicitly (and which uses all of the available history for estimation) appears to adjust more rapidly than its time-invariant counterpart on average. However, it adjusts more slowly than SPF.

Starting with real GDP growth, the movements in forecasts from TVP-VAR-SV and Small-VAR(1985) are generally similar to those from the SPF, though the SPF projections are on the lower side. At the beginning of 1994, all three (SPF, TVP-VAR-SV and Small-VAR(1960)) were forecasting underlying growth in the range 2.7–2.9%, roughly four-tenths lower than that of Small-VAR(1985). As time rolled forward from 1994 to 1996, professional forecasters gradually lowered their estimates, while the forecasts from the VARs remained steady. In early 1997, professional forecasters revised their forecasts back up by a couple of tenths to 2.5%, and left it stable at that level through the end of 1999. Over this same period, the forecasts from all three VARs were also revised up. Moving into 2000, while the VARs' projections held steady in the range 3.1–3.4%, professional forecasters revised their projection up strongly by roughly six-tenths, to 3.1%. This upward revision was in response to stronger growth data the prior two years that averaged more than 4%. The upward revision continued through 2005, by which point the long-run forecasts by both survey forecasters and the VAR were roughly in agreement at 3.4%. Beginning in 2006 and onward, the survey forecasters gradually lowered their growth forecasts, reaching 2.3% by the end of 2016. This rate of growth roughly matches the US economy's average growth rate since the start of the post-crisis recovery. Over this same period, forecasts from all three VARs also edged lower, but by a smaller margin (3.3% to 3.0% for Small-VAR(1960), and 3.0% to 2.6% for both TVP-VAR-SV and Small-VAR(1985)) compared to professional forecasters (from 3.4% to 2.3%).

In the case of the unemployment rate, while the forecasts implied by the VAR models have fluctuated within a narrow range of around 6%, the professional forecasters' estimates of the natural rate have evolved in line with the movements in the business cycle. For example, the estimate trended lower between 1994 and the start of 2001, but then reversed at the onset of the 2001 recession and began to trend up until the beginning of the recovery. It then trended lower again until the onset of the Great Recession. In response to a large upward spike in the unemployment rate that reflected the severity of the recession and the associated disruptions to the labor market, the professional forecasters rapidly adjusted their projections of the natural rate of unemployment upward. By the end of 2011, the professional forecasters' estimate of the

natural rate of unemployment increased to 6.0%, close to that implied by the VAR. Thereafter, professional forecasters adjusted their estimates downward as the recovery picked up its pace, reaching 4.8% by the end of 2016.

We next examine forecasts for the nominal variables: CPI inflation and the short-term interest rate. First, there is a noticeable downward trend in the projections over our forecast evaluation sample. Second, there is a sizable gap between the forecasts from the professional forecasters and the VAR forecasts compared to the plot for real variables. For CPI inflation, the SPF projection is relatively stable from 1999 onwards, even though all three VAR projections continue gradually to trend lower. The projection from the TVP-VAR-SV is very similar to that from Small-VAR(1985), especially since the start of the recovery in late 2009. However, there is a sizable but declining gap between the SPF projection and the projections from each of the VARs. Since 2009.Q1, on average the SPF projection is 1.70 percentage points lower than that from Small-VAR(1960), 0.50 percentage points lower than that from Small-VAR(1985) and 0.70 percentage points lower than that from TVP-VAR-SV. In the case of the short-term interest rate, since 2009.Q1, the SPF projection has been 288 basis points lower than that from Small-VAR(1960) on average, and 160 basis points lower than that from Small-VAR(1985).

Overall, these charts suggest that professional forecasters are quick to adjust their inflation expectations and to recalibrate their estimates of the underlying trend growth. In reality, it is difficult to distinguish between transitory fluctuations and fluctuations associated with changes to the underlying trend. As a result, forecasters learn about shifts in the underlying trend gradually. Even so, though, their expectations adjust more rapidly than those implied by the statistical models.

2.4 Methodology for Tilting Forecasts

2.4.1 Relative entropy

The technique of relative entropy, applied to economic forecasting by Robertson et al. (2005), consists of modifying a given predictive distribution to form a new predictive distribution such that it satisfies a given set of moment conditions while minimizing the relative entropy between two predictive distributions.

Let us begin with an unrestricted predictive distribution, $p(Y^{T+1,T+H} | Y^T)$, that corresponds to an n -dimensional random variable Y obtained from a VAR model. We assume that this predictive density consists of D draws $\{Y_i, i = 1, \dots, D\}$, and the corresponding weights are $\{w_i = 1/D, i = 1, \dots, D\}$. If the modeler now wants to impose the moment conditions contained in the matrix \bar{g} on this predictive distribution $p(Y^{T+1,T+H} | Y^T)$ such that $\sum_{i=1}^D w_i p(Y_i^{T+1,T+H}) \neq \bar{g}$; i.e., the mean of the predictive distribution $p(\cdot)$ is not equal to the mean condition required by the modeler (denote it as “new” information). For the predictive distribution to satisfy the new information, the original weights $\{w_i, i = 1, \dots, D\}$ must be modified. The new weights $\{w_i^*, i = 1, \dots, D\}$ that satisfy this new information are equivalent to

finding a new predictive distribution that is as close as possible to the original predictive density in the information-criterion sense.

Specifically, the relative entropy or Kullback-Leibler information criterion (KLIC) of w^* to w is

$$K(w^* : w) = \sum_{i=1}^D w^* \log\left(\frac{w_i^*}{w_i}\right). \quad (2.7)$$

We solve for new weights that minimize $K(w^* : w)$ and satisfy the following constraints:

$$w_i^* \geq 0, \quad \sum_{i=1}^D w_i^* = 1, \quad \sum_{i=1}^D w_i^* p(Y_i^{T+1, T+H}) = \bar{g}. \quad (2.8)$$

The first and second terms reflect the fact that weights are probabilities and so should be non-negative and sum to one. The third term represents the new moment conditions.

The solution to the minimization problem above using the Lagrange method is

$$w_i^* = \frac{w_i \exp(\gamma' p(Y_i^{T+1, T+H}))}{\sum_{i=1}^D w_i \exp(\gamma' p(Y_i^{T+1, T+H}))}, \quad (2.9)$$

where γ is the vector of Lagrange multipliers associated with the constraints. According to this, the original weights w have been tilted “exponentially” to produce the new weights w^* .

The vector of Lagrange multipliers (i.e., tilting parameters) can be obtained as a solution to the following minimization problem:

$$\gamma = \arg \min_{\tilde{\gamma}} \sum_{i=1}^D w_i \exp(\tilde{\gamma}' [p(Y_i^{T+1, T+H}) - \bar{g}]). \quad (2.10)$$

Then, using the newly computed weights, the updated expectation of other functions of interest can be computed simply as

$$\sum_{i=1}^D w_i^* h(Y_i^{T+1, T+H}). \quad (2.11)$$

If the interest is in the modified probabilistic density $g(Y^{T+1, T+H})$, as will be the case in our density forecast evaluation exercises, then, as was discussed by Cogley, Morozov, and Sargent (2005), importance sampling techniques could be used to redraw $Y_i^{T+1, T+H}$ from the original density $p(Y^{T+1, T+H})$ using the newly-found weights w^* , which can be achieved using the multinomial resampling algorithm of Gordon, Salmond, and Smith (1993). The steps of the algorithm (taken from Cogley et al., 2005) are detailed in Section A13 of the supplementary appendix.

In our forecasting exercises, the current-quarter (median) forecast from the SPF is used as the mean condition on the one-step-ahead VAR predictive density. Similarly, the long-horizon

(median) forecast from the SPF acts as the mean condition on the VAR predictive density at the horizon determined for combining the VAR with the survey long-run projections (detailed in the next section).¹³ In addition, following KCR, we also restrict the variance of the one-step-ahead VAR predictive density (i.e., uncertainty around the current/nowcast quarter), with the variance condition being computed as the variance of the SPF forecast errors over a fixed-length rolling window that precedes the forecast origin. A variance condition that is constructed via this approach is defined as an ex-post forecast uncertainty measure (see Clements, 2014; KCR).

Specifically, if we treat $\hat{Y}_{t,h}^{SPF}$ as the SPF forecast for indicator Y_t , then the variance condition is formed as

$$\sum_{q=0}^{15} (Y_{t-Delay-q} - \hat{Y}_{t-Delay-q,h}^{SPF})^2, \quad (2.12)$$

where $h = 1Q$, q reflects the number of past forecasts that are used to compute the variance of errors, and *Delay* indicates the number of quarters it takes the forecaster to learn about the actual realization. To remain consistent with our measure of ‘truth’ defined earlier (see Section 2.2), *Delay* is set equal to 2 for macroeconomic variables (real GDP growth, CPI inflation, and the unemployment rate). For our financial variable, the federal funds rate, *Delay* is set equal to 1, reflecting the fact that the actual quarterly value is available immediately preceding the last day of the quarter. For example, the variance conditions for macroeconomic variables at the forecast origin 1997.Q1 are based on the variance of the SPF nowcast errors computed over the preceding period 1992.Q4 through 1996.Q3. Similarly, for the federal funds rate, it is the variance of the errors over the period 1993.Q1 to 1996.Q4.

In a VAR, conditioning or tilting on some future horizon will influence the forecast all the way from the jumping-off point to the conditioned forecast horizon. For example, if we tilt the real GDP growth at a forecast horizon of $h = 6Q$, then tilting it is likely to impact the forecast trajectory from $h = 1Q$ to $h = 5Q$ for all of the variables.¹⁴ Conditioning on multiple variables simultaneously (in a system such as VAR) would result in forecast trajectories that reflect the cumulative effect of those conditions.

We use relative entropy instead of other approaches to conditional forecasting because of its ease of use, computational simplicity, and flexibility. Relative entropy is an effective and flexible conditional forecasting methodology, because it allows us to combine effectively both the mean condition and the modeler’s confidence in that mean condition (i.e., variance). This is an important advantage if the interest is in density forecasts as well as point forecasts. Furthermore, specifying only the mean condition will not result in the automatic shrinkage of the variance around the mean condition to zero; relative entropy will assume that the variance

¹³SPF also reports forecasts for four subsequent quarters beyond the nowcast quarter (i.e., $h = 2Q$ to $h = 5Q$), and these additional survey forecasts could be used as the moment conditions for the respective VAR predictive densities in order to obtain more accurate hybrid forecasts for the remaining forecast horizons that are of interest to policy makers ($h = 6Q$ to $h = 12Q$). We explore the usefulness of these additional conditions in Section A6 of the supplementary appendix (“Are there benefits to utilizing survey information for additional horizons?”).

¹⁴We provide an explanation of this spillover feature in Section A12 of the supplementary appendix using an analytical Gaussian example.

around that mean condition is the same as the unconditional.¹⁵ In addition, as was discussed by Giacomini and Ragusa (2014), relative entropy does not require Gaussian assumptions for either the original densities or the modified tilted densities. This latter advantage makes it possible to apply our proposal to nonlinear VARs and to density forecasts generated from a combination of several component density forecasts, because both of these approaches generate non-normal densities.¹⁶

2.4.2 Determining the forecast horizon for tilting

In combining the survey long-run projections with the VAR forecast, the initial inclination would be to combine the survey projections with the VAR forecast at some very distant horizon. This assumption is valid, because, by definition, the terminology long-run projection suggests many years into the future. That said, several macroeconomic variables (transformed to growth rates) display little persistence and therefore tend to move back rapidly to their respective (unconditional) means — the unrestricted long-run model forecast. The real GDP growth fits into this category. The VAR model forecasts of real GDP growth typically tend to move back toward the estimated mean within a year.¹⁷ At the other extreme are series such as the unemployment rate, which are very persistent. Depending on the starting point, it may take such series several years to move back toward the model-implied long run.

Based on the work of Clements and Hendry (1999) and Kozicki and Tinsley (2001a,b), we also know that forecasts beyond four quarters in covariance-stationary VAR models are influenced heavily by the model’s implied equilibrium value (i.e., unconditional mean). Clements and Hendry (1999) illustrate that a poorly estimated mean of the variable is the dominant source of the forecast errors beyond four quarters (e.g. a higher estimate of the trend than is thought to be reasonable will result in forecasts that are persistently biased upwards).¹⁸ This suggests that influencing the trend estimate implied from the model with the one informed from the survey requires us to begin doing so as soon as the model’s implied trend is expected to dominate the forecast values. For real GDP growth, it suggests targeting the horizon somewhere at the forecast horizon $h = 4Q$ or $h = 5Q$. If the trend is imposed late in the forecast horizon, it will be influenced too far out to have a meaningful influence on the forecast horizons of interest. Hence, the forecast remains biased or corrupted from the influence of the model’s implied steady state.¹⁹

¹⁵Alternative approaches to the construction of conditional forecasts include Waggoner and Zha’s (1999) soft conditioning, which is an extension of the work of Doan, Litterman and Sims (1984); the approach of Andersson, Palmqvist, and Waggoner (2010); and the Kalman filter approach as per Banbura, Giannone, and Lenza (2015). All three of these approaches can also allow for both mean and variance conditions. Antolin-Diaz, Petrella, and Rubio-Ramirez (2019) formally prove the equivalence between VAR conditional forecasting and relative entropy for the Gaussian case. We get very similar results for the point forecast evaluation if we instead use the approach of Doan et al. (1984) for imposing our conditions.

¹⁶Both these avenues are left for future research.

¹⁷See Domit, Monti, and Sokol (2019) for UK GDP growth and KCR for US GDP growth.

¹⁸See technical appendix A0.2 for an illustration of the implied steady state on the forecast trajectory.

¹⁹We illustrate this with two empirical exercises, reported in Section A15 of the supplementary appendix. In

Accordingly, we propose that the horizon for tilting should be variable-specific, and we suggest the following.²⁰ At each forecast origin t , retrieve the persistence estimates (i.e., slope parameters) that correspond to variable i from equation i of the VAR system in Eq. (1):

$$\rho_{i,t}^{+,BVAR} = \sum_{l=1}^p \bar{A}_{i,l}^{(i,i)}, \quad (2.13)$$

where $\bar{A}_{i,l}^{(i,i)}$ represents the posterior mean estimate of the slope coefficient of variable i in equation i of the VAR system in Eq. (1). It reflects an estimate of variable i 's persistence conditional on the VAR system.²¹

The corresponding metric that roughly determines the number of quarters that it takes to revert back to the VAR's implied steady state is

$$h_{i,t}^{+,VAR} = \frac{1}{1 - \rho_{i,t}^{+,VAR}}. \quad (2.14)$$

The horizon $h_{i,t}^*$ at which the survey long-run forecast is combined with the VAR forecast for variable i is set as

$$h_{i,t}^* = \max \{P_t^Q, h_{i,t}^{+,VAR}\}, \quad (2.15)$$

where $P_t^Q - 1$ specifies the minimum number of quarters prior to which the long-horizon survey forecast takes over the VAR forecast. The max operator ensures that the hybrid forecast uses the VAR forecast at least for the $P_t^Q - 2$ quarters following the nowcast quarter. In our exercises we set $P_t^Q = 5$ to reflect our preference for having a dynamic and informative forecast in the short to medium term. We note that the choice of P_t^Q does not influence our results; setting $P_t^Q = 0$ gives us very similar forecasting results because this choice binds only on real GDP growth (for just one or two quarters), not other variables.²²

In our empirical forecasting exercises, over the forecast evaluation sample, the horizon at which the survey long-run projection takes over has ranged between $h = 5Q$ and $h = 17Q$ for CPI inflation, $h = 9Q$ and $h = 30Q$ for the federal funds rate, and $h = 10Q$ and $h = 27Q$ for the unemployment rate; for real GDP growth, on the other hand, the horizon for combination has remained steady at $h = 5Q$. In the interests of brevity, the figures that plot the evolution

the first exercise, the horizon for combination is set dogmatically at $h = 25Q$ (i.e., seven years out), and in the second exercise, it is set at $h = 40Q$ (10 years out). The results indicate reduced gains in forecast accuracy for the horizons of interest. The reduced effects are most notable in the specifications that include SV.

²⁰The suggestion is roughly equivalent to generating unconditional forecasts far into the future (e.g. 40 quarters out) from the VAR model and then determining, for each variable, the precise forecast horizon at which it converges to its equilibrium value (i.e., implied steady state).

²¹Our results are robust if the forecast horizon for tilting uses an estimate (of persistence) that is obtained by estimating a univariate regression AR(4) recursively (i.e., at each forecast origin) for the variable of interest. The results are reported in Tables A19 and A20 in Section A15 of the supplementary appendix.

²²Over our forecast evaluation sample, it takes roughly three quarters on average for the real GDP growth to move back to trend growth for most forecast origins. By setting $h = 5Q$, we delay the takeover of the model forecast by the survey long-run forecast by two quarters, and so permit the possibility of imposing conditions on additional VAR forecast quarters (i.e., $h = 2Q$ through $h = 4Q$) for which survey expectations are available.

of the tilting horizon by variable and for each VAR model are relegated to Section A14 of the supplementary appendix.²³

2.5 Results

2.5.1 Forecasting exercise

Our main question of interest pertains to whether we achieve meaningful improvements in the forecast accuracy of the variables of interest (to monetary policymakers) by tilting the model-based forecasts to match the modeler’s long-run value. In our examination, the modeler’s long-run value equates to the (median of the) long-horizon SPF projections.

We answer this by performing a real-time out-of-sample forecasting evaluation over the period 1994.Q1–2016.Q4. We begin by estimating our VAR models using real-time data from 1959.Q4 to 1993.Q4 and generating unconditional forecasts iteratively, one to forty quarters out. Next, we re-estimate the models using an additional data point and again generate forecasts up to 40 quarters out. We repeat this recursive exercise until 2016.Q3. That is, the last estimation sample uses data from 1959.Q4 to 2016.Q3, and the forecasts span the period 2016.Q4–2026.Q3; however, given that data for evaluation are available only until 2016.Q4, we can only evaluate the one-step-ahead forecast. The forecasts generated through this recursive exercise are denoted ‘raw’ VAR forecasts. Next, we use the technique of relative entropy to tilt the one-step-ahead ‘raw’ VAR forecasts (i.e., predictive densities) generated in the previous step to match the SPF nowcasts for real GDP growth, CPI inflation (and core CPI inflation²⁴ in the medium VAR), the unemployment rate, and the federal funds rate. We denote the resulting tilted forecasts (corresponding to all variables) ‘baseline’ forecasts. The ‘baseline’ forecasts tilt on the nowcasts only (i.e., both the nowcast mean and variance). Next, we generate another set of forecasts, but this time tilting the ‘raw’ VAR forecasts toward both the nowcasts and the survey long-run projections for the same set of variables (as in the ‘baseline’ forecasts). We denote these forecasts ‘hybrid’ forecasts. Finally, for each VAR model under consideration, we evaluate and compare the point and density forecast accuracies among the raw VAR, baseline, and hybrid forecasts in a pairwise fashion.

All of the tables (unless specified) are formatted so to facilitate quick comparisons across raw, baseline, and hybrid forecasts for each model, as well as comparisons of raw forecasts across models. The comparison of raw forecasts between small and medium VAR models helps

²³For all VAR models except for the TVP-VAR SV and TVP-VAR models, the forecast horizon for combination is determined based on the specification with a constant variance, and the same value of the combination horizon is used for the counterpart specification that allows for SV. This facilitates more direct comparisons between specifications for assessing the role of SV in the forecast accuracy. However, our results remain qualitatively similar if we relax this restriction.

²⁴For core CPI inflation, the SPF nowcasts are not available until 2007.Q1. The inflation nowcasting models of Knotek and Zaman (2017) or Modugno (2013) could be used to produce core CPI inflation nowcasts prior to 2007.Q1; however, for the sake of consistency and simplicity, we merely use core CPI nowcasts from the SPF starting in 2007.Q1.

in assessing the usefulness of additional variables for improving the accuracy of the core set of variables. In addition, the formatting of the results helps in assessing whether tilting helps more models that are inferior to begin with.²⁵ Tables 2.1–2.6 (and Tables A1–A4 in Section A1 of the supplementary appendix) report the real-time forecast accuracies of real GDP growth, CPI inflation, the unemployment rate, the federal funds rate, and credit spreads (Tables A1–A4 also report accuracy results for the additional five variables), and allow a quick assessment of the usefulness of the stochastic volatility for the forecast accuracy. Each table is split into two panels. The left panel reports accuracy results of the VAR model with constant variance, while the right panel reports results for the equivalent VAR model with stochastic volatility. For each variable, the first row reports the MSEs (CRPS values for density forecasts) for the raw VAR forecast, while the subsequent three rows report the relative MSEs (relative CRPS values for density forecasts): the MSE of the baseline forecast relative to the raw forecast, the MSE of the hybrid forecast relative to the raw forecast, and the MSE of the hybrid forecast relative to the baseline forecast. The accuracy results are reported for forecast horizons of one (i.e., nowcast quarter), four, eight, and 12 quarters out, respectively. The forecast evaluation is based on the full sample spanning the period 1994.Q1 to 2016.Q4.

Results: Small VAR estimated with the longer sample

Table 2.1 reports the real-time point forecast accuracy from the Small VAR estimated with data going back to 1959.Q4. A couple of things stand out immediately. First, allowing for SV helps to improve the point forecast accuracy; however, the magnitudes of the improvements and persistence in those gains vary across variables. The gains are strongest for CPI inflation and persist throughout the forecast horizon; for real GDP growth, on the other hand, improvements are achieved at least through four quarters out. The improvements for unemployment rate are marginal and persist throughout. These findings are in line with those of Clark (2011) and D’Agostino et al. (2013). For the federal funds rate and the credit spread, the improvements are short-lived, as allowing for SV appears to worsen the forecast accuracy by four quarters out and beyond. Second, across all variables, tilting towards the survey nowcasts improves the forecast accuracy significantly in the nowcast quarter (i.e., $h = 1Q$). The relative MSE is substantially below one for rows labelled ‘Baseline/Raw’ and ‘Hybrid/Raw’. These large improvements suggest that the SPF nowcasts are significantly more accurate than the VAR model’s one-step-ahead forecasts, which is consistent with the results of KCR and several other studies documenting the usefulness of external nowcasts for models estimated purely with quarterly data (e.g. Knotek & Zaman, 2019). Looking across all of the rows labeled ‘Baseline/Raw’, the spillover effects (on subsequent forecast horizons) from more accurate nowcasts last longer for persistent variables, CPI inflation, the unemployment rate, and the federal funds rate. For real GDP growth, the gains are relatively short-lived.

²⁵For a convenient assessment of the effectiveness of tilting, please refer to Section A4 of the supplementary appendix.

In all rows labeled ‘Hybrid/Raw’, the ratios are less than one and are generally smaller than those reported in the row immediately above (‘Baseline/Raw’), with few exceptions. This suggests that tilting the VAR forecasts to match the survey long-horizon forecasts in addition to the survey nowcasts leads to further improvements in accuracy. A notable difference across the two rows (i.e., Hybrid/Raw vs. Baseline/Raw) is the significantly improved accuracy at forecast horizons that are further out.

The rows labeled ‘Hybrid/Baseline’ facilitate our attempts to get a sense of the marginal gains in accuracy from tilting towards the survey long-horizon over and above that from tilting towards the survey nowcasts only. For example, eight quarters out, the hybrid forecast for real GDP growth (in both VAR specifications) is 15% more accurate than the baseline forecast on average, as is indicated by the ratio of 0.85. Digging deeper into error evaluation, the improvements in the average accuracy of the real GDP growth forecast come mainly from the Great Recession and the subsequent recovery. This is evident in Figure 2, which plots the timeline of the cumulative sum of squared forecast errors for the horizon $h = 8Q$: beginning around the height of the Great Recession and continuing until the end of evaluation sample, the baseline forecast consistently underperforms the hybrid forecast. This is evident from the fact that the plot that corresponds to the hybrid forecast lies below the baseline forecast beginning at the Great Recession, and the divergence between the two increases over the remaining evaluation sample. The pattern of an improved accuracy of hybrid forecasts around the crisis period and beyond fits with the formal statistical assessment of a structural change at that period (see Aastveit et al., 2017). This result suggests that the hybrid approach adjusts the forecasts to accommodate the possible structural change. This is confirmed further by an inspection of recursive forecast trajectories (not shown). The baseline forecast calls for stronger long-run projections, and as a result, the recursive baseline forecast trajectories continuously over-predict growth. However, the recursive trajectories from the hybrid forecast track the actual data relatively better because they rely on professional forecasters’ assessments of a lower growth potential of the economy, perhaps drawn from demographics or specific assessments of technological change.

For CPI inflation, the forecast accuracy gains are substantially higher, are statistically significant, and persist throughout; the hybrid forecast for CPI inflation is roughly 25% (11% for the specification with SV) more accurate than the baseline forecast two years out, and 30% (17% for the specification with SV) more accurate three years out. This is not surprising because it is well known that the inflation process has exhibited pronounced changes in the underlying trend since the 1950s, meaning that accounting for those changes turns out to be very important for achieving an improved accuracy. Just as in the case of real GDP growth, Figure 2 shows that hybrid forecasts have been substantially more accurate than baseline forecasts since the post-crisis period. Unlike in the case of real GDP growth, the superior accuracy of the hybrid CPI inflation forecasts is evident over the entire forecast evaluation sample preceding the Great Recession. This improved accuracy prior to the Great Recession has been possible because

tilting is helping to reduce the bias in the forecasts that is introduced by a failure to account for the structural break of the mid-80s (i.e., estimating with a longer history). The sizable divergence between the hybrid and baseline forecasts since the post-crisis period partly reflects the fact that tilting is helping the hybrid forecasts to mitigate the misspecification issues that result from a failure to account for structural breaks in both the mid-80s and the post-crisis period.

The improvements in accuracy for the effective federal funds rate are of similar magnitudes to those for CPI inflation. The combination of improved inflation forecasts and tilting on judgmental survey-based long-run values of the federal funds rate results in highly accurate forecasts of the federal funds rate. The average improvements in forecast accuracy over the baseline forecast by the end of the third year are 30% (40% in the VAR specification with SV). Figure 2 confirms the conjecture that tilting is helping hybrid forecasts to handle structural change better.

Credit spread is a variable that is not tilted directly, but its accuracy is affected indirectly through the spillover effects of variables that are tilted directly. Not surprisingly, the improved accuracies of the federal funds rate, inflation, and real GDP growth are associated with an improved accuracy of the credit spread. The magnitudes of the gains are similar to those seen for the federal funds rate.

In the case of the unemployment rate, the average gains from tilting towards survey long-horizon forecasts are small. As is evident from Figure 2, the marginal gains in accuracy reflect the slightly more accurate hybrid forecasts over the Great Recession and the post-recovery period.

Table 2.2 reports the corresponding accuracies of the density forecasts. A lower value of CRPS is preferred, so negative entries in the rows labeled ‘Baseline - Raw’ suggest that the baseline density forecast is more accurate than the raw density forecast on average. Similarly, negative entries for ‘Hybrid - Raw’ suggest that the hybrid forecast is more accurate than the raw forecast, and negative entries in the case of ‘Hybrid - Baseline’ suggest that the hybrid density forecast is more accurate than the baseline forecast on average. The results of the density forecast evaluation echo the results of the point forecast evaluation reported in Table 2.1. First, adding SV helps to improve the density forecasts, with the pattern of improvements being quite similar to that in the point forecast assessment. Second, the baseline forecasts are more accurate than the raw VAR forecasts, and the hybrid forecasts are more accurate than either the baseline or the raw VAR forecasts. Third, the results illustrate that the density forecast accuracy gains of the hybrid forecasts for CPI inflation, the federal funds rate, and credit spreads are substantial, are statistically significant, and persist far into the future. Table A5 in the supplementary appendix reports the density calibration diagnostics. For the VAR with constant volatility, we find that the density forecasts are badly calibrated (which is consistent with the findings of Rossi & Sekhposyan, 2014), with the exception of the nowcast quarter (and the subsequent quarter) for both hybrid and baseline forecasts. SV helps to improve the

calibration of the density forecasts via an improved coverage (i.e., predictions intervals that are close to the nominal coverage of 70%), which is consistent with the findings of Clark (2011) and D’Agostino et al. (2013). However, it generally remains the case (for forecast horizons beyond the short-term) that density forecasts with SV are unable to pass all of the necessary statistical tests (of PITS) to be categorized as calibrated correctly.

Results: Small VAR estimated with the post-1985 sample

Table 2.3 reports the real-time point forecast accuracy from the Small VAR estimated with data going back to 1985. The results indicate similar patterns of accuracy improvements to those seen in the case of the VAR estimated with the longer sample, with the gains being smaller in magnitude but still statistically significant in many instances. SV helps to improve the point forecast accuracy, but the gains are substantially smaller than in the case of the VAR estimated with the longer sample. Just like in the previous case for the VAR with the longer sample, the hybrid approach helps to improve the accuracy in both economically meaningful and statistically significant ways for the real GDP growth, CPI inflation, the federal funds rate, and the credit spread. One exception is the unemployment rate, in that the hybrid approach appears to worsen the forecast accuracy; however, those deteriorations are not flagged as statistically significant. In that regard, the results for the unemployment rate across the two VARs (longer and shorter samples) are similar, as the accuracy gains in the VAR with the longer sample were not statistically significant.

The plots in Figure 3 indicate that most of the accuracy improvements (or losses in the case of the unemployment rate) for the hybrid approach (compared to the baseline) come from the evaluation period beginning in 2010; i.e., coinciding with the sample period that is thought to have undergone structural change. This is exactly where we would expect to see gains for the hybrid forecasts from VARs estimated with data beginning after the well-documented structural breaks of the mid-80s, and supports our conjecture that the hybrid approach helps to accommodate structural breaks. Recall that the Federal Reserve adopted an inflation targeting framework in 2012, and it is since then that the CPI inflation forecasts derived from the hybrid approach have outperformed the baseline, with gains that are statistically significant though marginal. We also point out that these two VAR specifications outperform the VAR specifications estimated with a longer sample (see Table A12 in Section A4 of the supplementary appendix for the formal rankings based on forecast performances).

Table 2.4 reports the corresponding accuracies of the density forecasts. The results of the density forecast evaluation echo those of the point forecast evaluation reported in Table 2.3. One exception is the unemployment rate: the worsening in the density forecast accuracy of the hybrid forecasts is marginally significant, whereas in the case of the point forecast accuracy, the deteriorations in accuracy of the hybrid forecasts generally were not statistically significant. Table A6 in Section A2 of the supplementary appendix reports the density calibration diagnostics. Interestingly, the raw forecasts for real GDP growth from the VAR with constant variance

are calibrated correctly. SV helps to improve the calibration of the density forecasts significantly for CPI inflation; for the unemployment rate and the federal funds rate, SV improves the coverage rate but the density forecasts generally fail the statistical tests for categorization as calibrated correctly. SV helps to widen the prediction intervals (width of the 70% bands), which is consistent with the discussion by Clark (2011). Interestingly, we find in general that the hybrid forecast helps to improve the calibration via an improved coverage (i.e., prediction intervals that are close to the nominal coverage of 70%) only if the underlying (raw or baseline) forecast to be tilted is calibrated better. However, there are exceptions. For CPI inflation, all three forecasts (raw, baseline, and hybrid) from the VAR specification with SV appear to be well-calibrated (based on p -values from the Knüppel test being greater than 0.10). A closer inspection reveals that, of these three forecasts, the hybrid forecast could be characterized as slightly better calibrated, as its coverage rates are generally closer to the nominal rate of 70% than those of the other two.

The evidence of the hybrid approach helping to improve the calibration is more prominent for the VAR specification with a constant variance. As can be seen, the raw CPI inflation forecast is calibrated poorly, because the null hypothesis of a well-calibrated density is rejected at the 10% significance level for all horizons (p -values from the Knüppel test are below 0.10). This is supported further by the corresponding empirical coverage rates, which are well below the nominal value of 70%. Beyond the short-term, the baseline forecast also fails the tests of properly-calibrated densities. In contrast, the hybrid forecast appears to be well calibrated for all forecast horizons shown, as the p -values (from the Knüppel test) are all above 0.10 and the empirical coverage rates are much closer to the nominal value of 70%.

Results: Time-varying parameters VAR

Table 2.5 reports the real-time point forecast accuracies from the small-scale time-varying VAR estimated with data going back to 1959.Q4. The specification with SV helps to improve the point forecast accuracy for both real GDP growth and CPI inflation. For CPI inflation, the results point to economically meaningful and statistically significant accuracy improvements that are similar to those for time-invariant VARs (with and without SV). Specifically, for the time-varying VAR with a constant variance, the magnitudes of the accuracy gains for hybrid forecasts are similar to those seen earlier for the small VAR estimated with the longer sample (e.g., the three-year-out hybrid is 30% more accurate than the baseline). For the time-varying VAR specification with SV, the gains are relatively smaller and are similar to those seen for the small VAR estimated with post-1985 data (e.g., the three-year-out hybrid forecast is 18% more accurate than the baseline on average). We find that the time-varying VAR with SV is among the most accurate of the models considered in this chapter (see Table A12 in Section A4 of the supplementary appendix). In the cases of real GDP growth and the unemployment rate, the time-varying VAR with a constant variance is a close competitor. However, for CPI inflation, it is the combination of time-varying parameters and stochastic volatility that helps

to position it as a good forecasting model. We suspect that the superior accuracy of the SV specification over the constant-variance specification for CPI inflation is due in part to the much tighter priors that are required for the SV specification. Our results echo the findings of D’Agostino et al. (2013).²⁶ For real GDP growth, the results indicate a smaller benefit from the hybrid approach: even though they appear economically meaningful, they are not statistically significant. As in the case of the VAR models discussed earlier, Figure 4 illustrates that most of the gains from the hybrid approach come from the post-crisis period. In the case of the unemployment rate, the results are very similar to those of the small VAR estimated with the post-1985 sample. On average, the hybrid approach leads to inferior forecasts relative to the baseline for the unemployment rate, but the losses are not significant statistically. These losses come primarily from the post-crisis period.

Table 2.6 reports the corresponding accuracies of the density forecasts. The results of the density forecast evaluation echo those of the point forecast evaluation reported in Table 2.5. Table A7 in the supplementary appendix reports the density calibration diagnostics. Beyond the nowcast quarter, the density forecasts from the specification with a constant variance are calibrated badly. Allowing for SV helps to improve the calibration of the density forecasts for real GDP growth and CPI inflation. For real GDP growth, the density forecasts from the SV specification are calibrated correctly. For CPI inflation, the SV specification leads to improvements in the coverage rates but generally fails the statistical tests for classification as calibrated correctly.

Results: medium VAR

In the interests of brevity, the results for the medium VAR are relegated to Section A1 of the supplementary appendix. Here, we summarize our findings briefly. Firstly, the patterns of both point and density forecast accuracy improvements for hybrid forecasts are generally similar to those of the small VAR. Specifically, the results for the medium VAR estimated with the longer sample echo the results reported for the small VAR with the longer sample. One difference is that the magnitudes of the improvements for the hybrid forecasts are slightly smaller than those reported for the small VAR.

Secondly, the additional variables in the medium VAR help to improve the accuracy of the core variables of interest (relative to the small VAR), and therefore tilting helps slightly less with the more accurate raw forecasts. The finding that the medium VAR generates more accurate forecasts than the small VAR is in line with the results of Banbura et al. (2010) and Koop (2013). The results for the medium VAR estimated with the shorter sample echo the results reported for the small VAR with the shorter sample. Unlike in the case of estimation with the longer sample, where the medium-sized VAR was favored over the small VAR, we do not find

²⁶D’Agostino et al. do not report results for a time-varying VAR without SV, but one can perform a rough assessment by comparing the TVP-AR model with a time-varying VAR with SV (denoted TVP-VAR in their paper) for inflation. The TVP-AR performs significantly worse than the time-varying VAR with SV.

this pattern for VAR models estimated with the shorter sample.

Thirdly, the most useful aspect of the results for the medium VAR is the strong positive spillover effects that are seen on the accuracies of the variables that are not tilted directly. Impressive and statistically significant gains in the accuracy of forecasts derived from the hybrid approach are achieved for core CPI inflation, wage compensation, nonfarm payroll employment, and the credit spread. It is also worth highlighting the fact that adding core CPI inflation and SV to the medium VAR helps to improve the accuracy of the raw CPI forecasts greatly. This implies that tilting is less effective for CPI inflation for those specifications, but that it is very effective in all medium VAR specifications for core CPI inflation.

Results: Hybrid vs. the Federal Reserve’s Greenbook

The results shown thus far largely support our conjecture that the hybrid approach assists in mitigating misspecification issues by helping to accommodate possible structural change (due to changing trends) to the extent that it is detected in real-time by professional forecasters. As a final check in support of our conjecture, we performed an additional exercise comparing the accuracy of our hybrid forecasts to that of the Federal Reserve’s Greenbook (GB). The GB forecast can be thought of as a combination of model and judgement, and therefore will be expected to handle structural change better than standard VARs. We confirm that this is indeed the case, and, strikingly, that the hybrid forecasts for real GDP growth and CPI inflation from our simple VARs are competitive with the GB forecasts. In the case of the unemployment rate, the hybrid forecasts under-perform GB during the Great Recession but are competitive with GB on average. Due to space constraints, the results are presented in Section A7 of the supplementary appendix.

Overall, our point and density forecasting results using real-time data provide compelling evidence that tilting VAR forecasts to match the long-run forecasts from the Survey of Professional Forecasters systematically leads to improved forecast accuracies for most variables over the forecast horizons that are of interest to monetary policymakers. In general, our proposal is more helpful for models that perform worse in raw form (i.e., raw VAR forecasts). Interestingly, the rankings are generally maintained post-tilting; i.e., models that were ranked low prior to tilting continue to be ranked low post-tilting, though the differences in accuracy are much smaller (see Section A4 of the supplementary appendix).²⁷

²⁷As a further check of the robustness of this statement, we estimate the small VAR with loose priors and assess the extent to which our proposal helps to improve its accuracy. We find that this VAR specification is ranked the lowest (as would be expected) prior to tilting, gains the most from tilting in terms of accuracy improvements, and remains ranked at the bottom post-tilting, but with a significantly reduced margin relative to the next best model. The results of this exercise are reported in Section A5 of the supplementary appendix.

2.5.2 Inflation forecast accuracy of the tilted VAR compared to univariate benchmarks

Tilting the VAR forecasts to match the long-horizon survey forecasts leads to meaningful gains in forecast accuracy for most variables, and the gains in point forecast accuracy are substantial for nominal variables such as price inflation and wage inflation. Given these results, we investigate how the accuracy of inflation forecasts from the tilting approach compares to those from hard-to-beat univariate benchmark models. Accordingly, we next compare the inflation forecast accuracy from our medium VAR (with and without stochastic volatility) using the three best-known univariate benchmarks: a random walk model (Atkeson and Ohanian, 2001), the univariate unobserved component with stochastic volatility (UCSV) model of Stock and Watson (2007), and Faust and Wright’s (2013) inflation in gap.²⁸

Random walk model of Atkeson and Ohanian (2001). For our forecasting exercise, the forecasts of CPI inflation into the future are computed by averaging the previous four available quarterly annualized readings of the CPI.

To ensure a fair horse race, we set $\hat{\pi}_{t+1}$ equal to the *survey nowcast*

$$\hat{\pi}_{t+h} = 0.25(\hat{\pi}_{t+1} + \pi_t + \pi_{t-1} + \pi_{t-2}) \quad \text{for } h \geq 2. \quad (2.16)$$

Univariate unobserved component with stochastic volatility (UCSV) model of Stock and Watson (2007). The superior accuracy of this model for forecasting inflation is documented well in numerous studies. The model decomposes inflation into two components, a stochastic trend component and a transitory component, and assumes time-varying variances of the respective shocks to these two components. The specification of this model is as follows (for ease of exposition, we retain the notation used by Stock & Watson, 2007):

$$\pi_t = \tau_t + \eta_t, \quad \text{where } \eta_t = \sigma_{\eta,t} \zeta_{n,t} \text{ and } \zeta_{n,t} \text{ is i.i.d. } N(0, I_1) \quad (2.17)$$

$$\tau_t = \tau_{t-1} + \varepsilon_t, \quad \text{where } \varepsilon_t = \sigma_{\varepsilon,t} \zeta_{\varepsilon,t} \text{ and } \zeta_{\varepsilon,t} \text{ is i.i.d. } N(0, I_1) \quad (2.18)$$

$$\ln(\sigma_{\eta,t}^2) = \ln(\sigma_{\eta,t-1}^2) + \nu_{\eta,t}, \quad \text{where } \nu_{\eta,t} \text{ is i.i.d. } N(0, \gamma_1) \quad (2.19)$$

$$\ln(\sigma_{\varepsilon,t}^2) = \ln(\sigma_{\varepsilon,t-1}^2) + \nu_{\varepsilon,t}, \quad \text{where } \nu_{\varepsilon,t} \text{ is i.i.d. } N(0, \gamma_2). \quad (2.20)$$

The model forecast for inflation an infinite number of quarters into the future is simply the model’s current estimated trend inflation rate.²⁹

To ensure a fair horserace, we set $\hat{\pi}_{t+1}$ equal to the *survey nowcast*. For forecasts $h \geq 2$, we estimate the model through $t + 1$, treating the survey nowcast as data and computing the

²⁸See Tallman and Zaman (2017) for a broader examination of the forecasting properties of other models, as well as of these two.

²⁹The scalar parameters γ_1 and γ_2 determine the smoothness of the stochastic volatility process. Following Stock and Watson (2007), we fix both at 0.2.

updated trend estimate:

$$\hat{\pi}_{t+h} = \hat{\tau}_{t:t+1} \quad \text{for } h \geq 2. \quad (2.21)$$

Inflation in gap model of Faust and Wright (2013). This model is patterned along the lines of Faust and Wright’s univariate AR1 inflation in gap model. According to this model, CPI inflation (quarterly annualized rate) is modeled as a gap, where ‘gap’ is defined as a deviation from a trend, which for CPI inflation is taken to be the SPF long-horizon forecast:

$$\pi_t^{Gap} = \alpha_0 + \alpha_1 \pi_{t-1}^{Gap} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \quad (2.22)$$

where $\pi_t^{Gap} = \pi_t - \pi_t^{SPF \text{ LR}}$.

The estimated model is used iteratively to produce forecasts h periods ahead. The forecasts are of the gap (i.e., the CPI inflation gap), and to it we add the latest available estimate of the SPF trend (as of forecast origin t) to get the implied forecast of the CPI inflation.

Table 2.7 reports the results from a comparison of the out-of-sample CPI inflation forecasting performances (both point and density) across five models: Medium VAR est. 1960, Medium VAR SV est. 1960, random walk model, UCSV model, and Faust and Wright’s inflation in gaps model. Also included is the forecasting performance of the SPF survey itself. Since the SPF covers only forecast horizons $h = 1Q$ to $h = 5Q$, its entries in the table for $h = 8Q$ and $h = 12Q$ are blank. The top portion of the table reports the results that correspond to the full evaluation sample, while the lower portion corresponds to the pre-crisis sample (i.e., 1994.Q1 to 2006.Q4). The left panel reports an accuracy comparison of the Medium VAR (with constant variance) with the univariate forecasts and the right panel reports an accuracy comparison of the Medium VAR with SV to the same univariate forecasts.

As is shown by the numbers reported in the table, the hybrid CPI point forecasts are at least competitive with univariate forecasts and more accurate in many cases, with several of those gains being flagged as statistically significant. This result holds for each of the forecast evaluation samples.

In regard to the density forecast accuracy, the UCSV forecast is substantially superior to the baseline forecast obtained from the Medium VAR with constant volatility (as evidenced by positive values), and the gains are statistically significant throughout. However, relative to the hybrid forecast, the UCSV forecast is competitive through most of the forecast horizons but slightly more accurate at the longer horizons (with the gain being marginally significant). In the case of the Medium VAR specification with SV, both the baseline and hybrid density forecasts are competitive to the UCSV forecasts. These results suggest that combining the long-horizon survey forecasts with the VAR forecasts improves the density forecasts notably. Adding SV helps to improve the density forecast accuracy of the hybrid forecasts further.

The evidence that the multivariate model (VAR) estimated using a sample that goes back to 1960, when tilted towards survey long-horizon forecasts, generates inflation forecasts that

rival both the point and density forecast accuracies of forecasts from tough benchmarks is a worthwhile result, because policymakers at central banks who are faced with inflation targeting may benefit from using models that allow for feedback between policy, the real economy and inflation. Such models have been known to underperform simple univariate approaches in terms of forecasting inflation, at times significantly so. An examination of our exercises suggests that the modeling technique that we propose and illustrate may have some useful payoffs for policymakers.

2.6 Conclusions

This chapter proposes a technique for adjusting the medium- and long-horizon forecasts from a VAR toward plausible values that are proposed by judgmental forecasters. We construct hybrid forecasts that consist of survey nowcasts, VAR forecasts, and long-horizon survey projections. Specifically, by applying the flexible and powerful technique of relative entropy, we tilt the VAR forecast both in the near term with the survey nowcast and in the long run with the survey long-run projection. The horizon at which the long-run survey projection is combined with the VAR forecast is variable-specific and is determined by the variable's estimated persistence at the forecast origin.

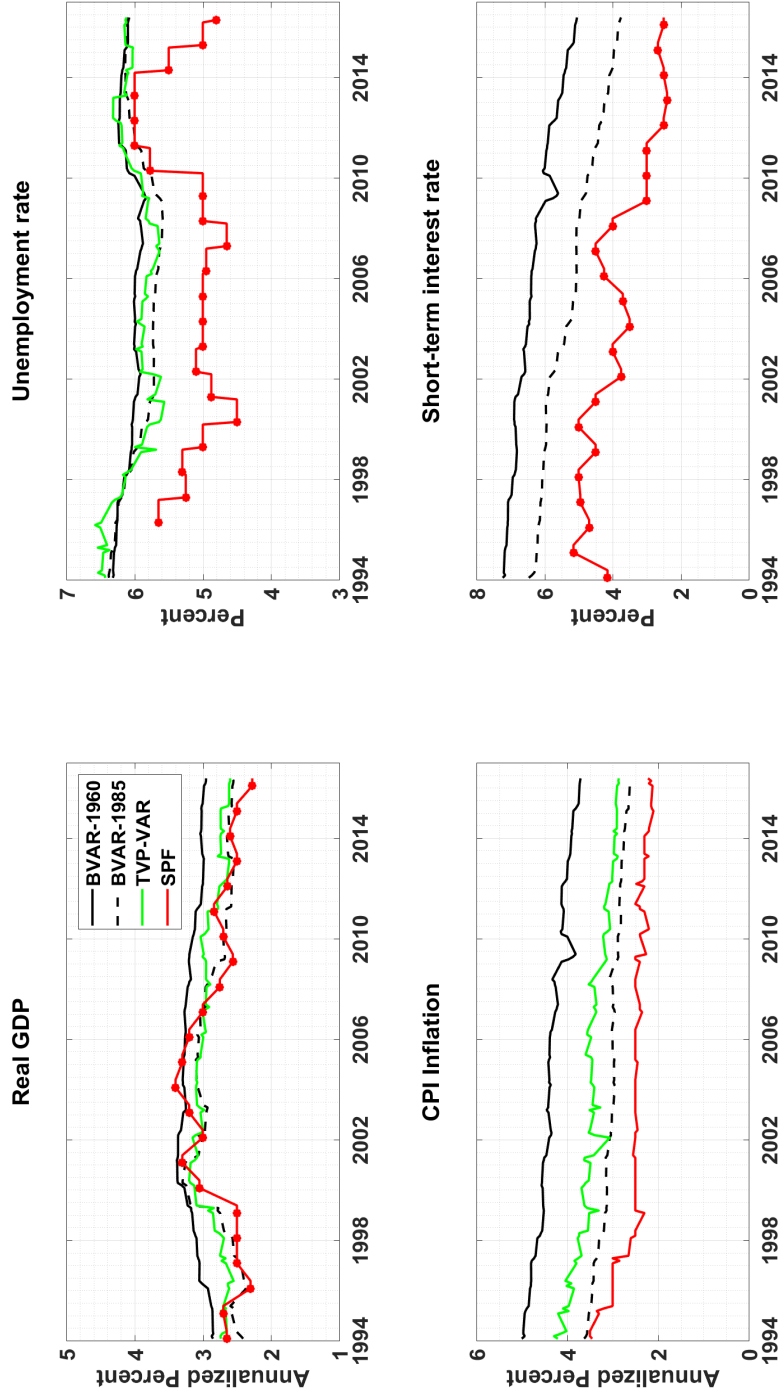
We consider the efficacy of our proposal on a variety of VAR models estimated using Bayesian methods. We find that all of the models examined benefit through an improved accuracy, with the largest gains in forecast accuracy being seen in models that are estimated with longer histories and the smallest gains for models that attempt to accommodate structural changes. We find that tilting VAR forecasts to match the long-run forecasts from the Survey of Professional Forecasters systematically leads to improvements in forecast accuracy, as is indicated by the lower MSE and lower CRPS values for most variables over the forecast horizons that are of interest to monetary policymakers (i.e., one to three years out). The greatest gains in accuracy are achieved for variables that are believed to have undergone marked shifts in their permanent components (e.g. inflation and the federal funds rate). We also show substantial forecast accuracy improvements for a host of variables (such as compensation, payroll employment, credit spread) that are not tilted directly but are affected through the spillover effects of the tilted variables.

We also show that hybrid inflation forecasts from simple VAR models rival those of relatively accurate univariate benchmark models. We view this as a useful and practical contribution because the many frustrations of monetary policymakers include the inability of multivariate models, which allow for feedback effects from policy to the real economy and inflation, to match the forecasting performances of univariate forecasting models.

Finally, we show that all of the models considered display demonstrable improvements in the forecast accuracies of hybrid forecasts for real GDP growth and CPI inflation over the last decade (i.e., from the Great Recession to the present), which coincides with the time period

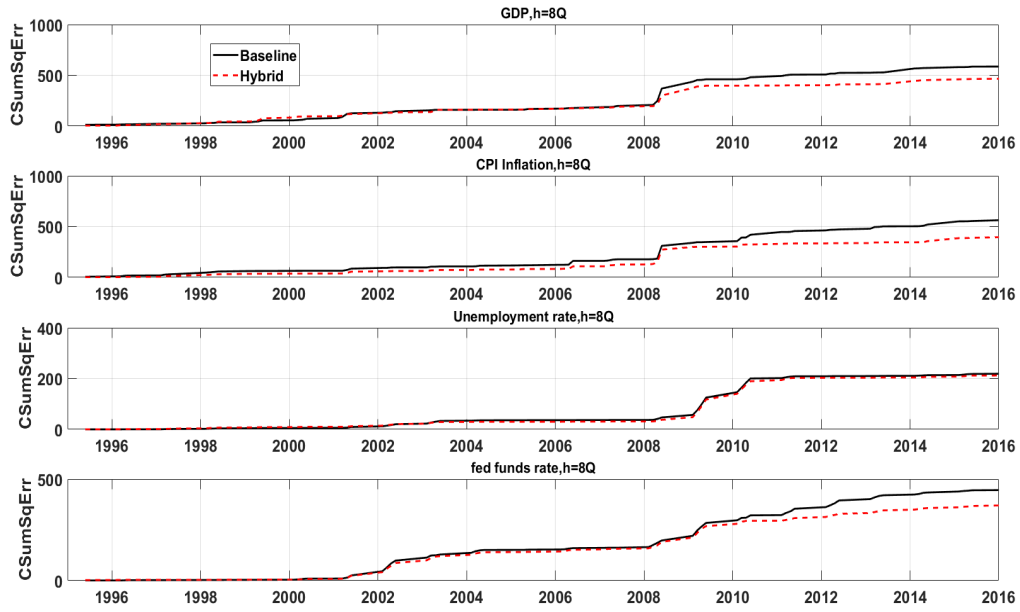
that is associated with possible structural change. These results lead us to view our proposal as a low-cost method for mitigating model instability issues that may arise from structural shifts caused by moving end points.

Figure 2.1: Real-time long-run forecasts.



Notes: The figure plots the real-time long-run forecasts from the Survey of Professional Forecasters (SPF), namely the 10-year-ahead average forecast. To facilitate comparison, the figure also plots the real-time estimates of the long-run forecasts implied by a fixed-parameter small BVAR estimated recursively with data beginning in 1959:Q4, a fixed-parameter small BVAR estimated with data beginning in 1985, and a time-varying parameter small BVAR. The short-term interest rate refers to the 3-month T-bill rate in the case of SPF, but the federal funds rate in the case of the BVARs.

(a) Cumulative sum of squared errors



(b) Cumulative sum of CRPS

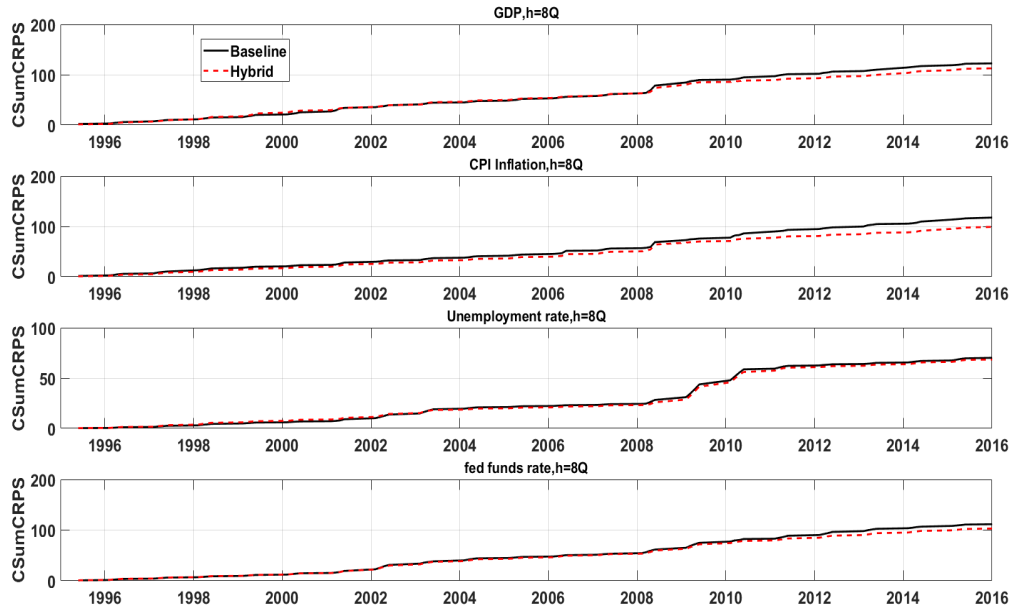
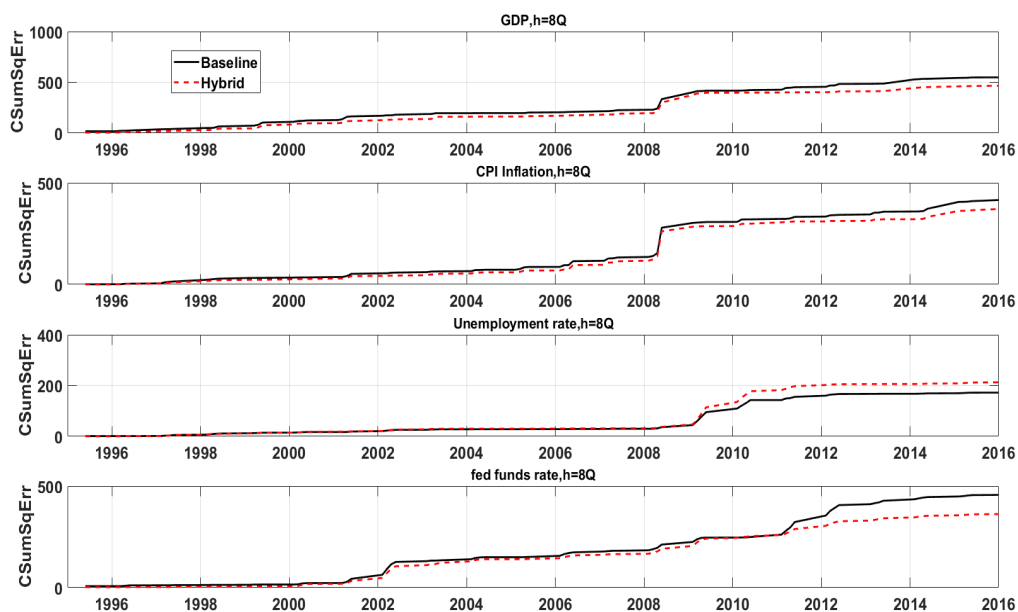


Figure 2.2: Results of the small VAR est. 1960: cumulative squared errors (top) and cumulative CRPS (bottom).

(a) Cumulative sum of squared errors



(b) Cumulative sum of CRPS

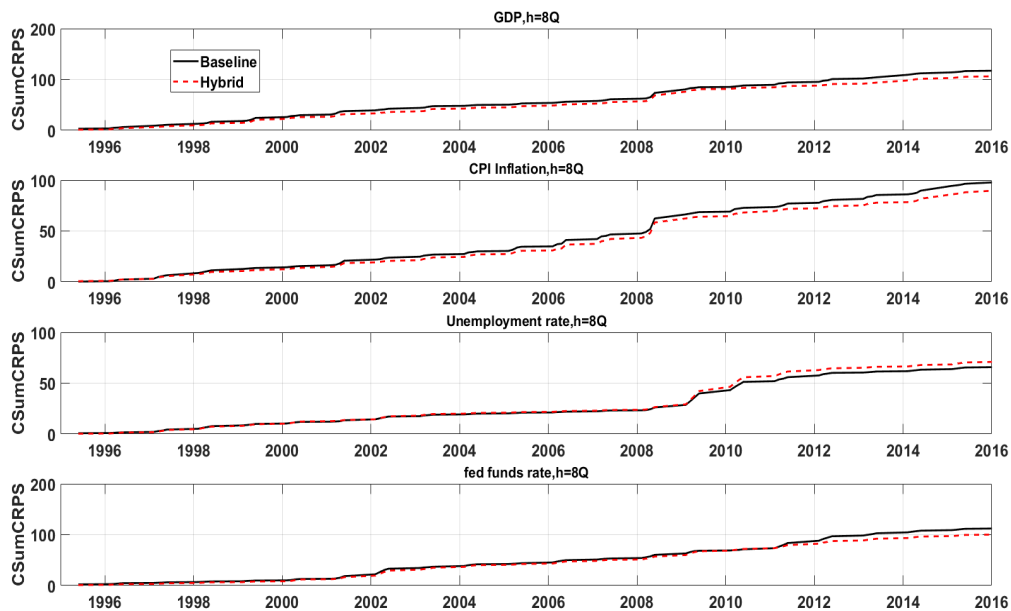
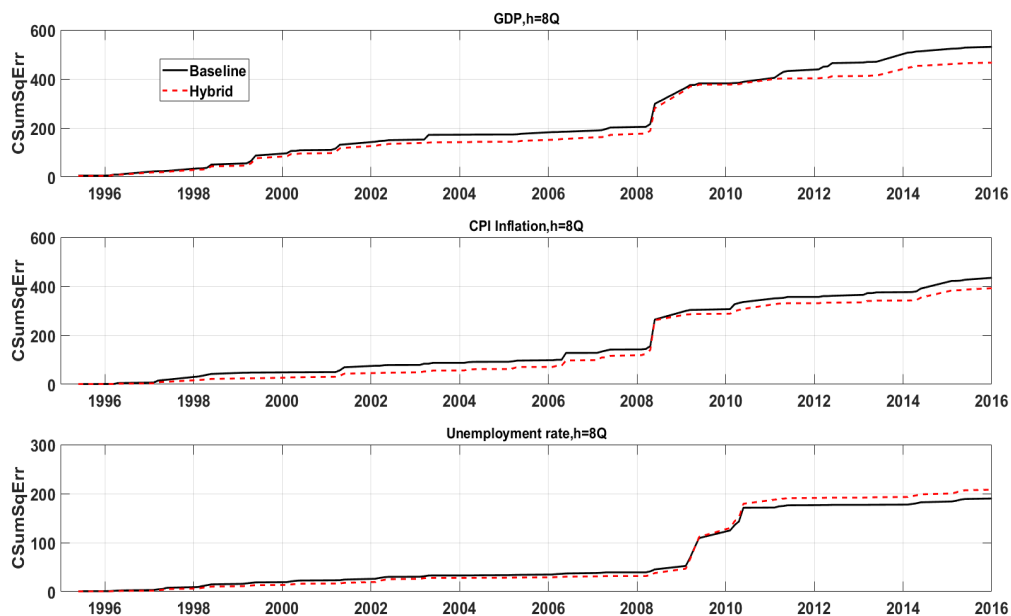


Figure 2.3: Results of the small VAR est. 1985: cumulative squared errors (top) and cumulative CRPS (bottom).

(a) Cumulative sum of squared errors



(b) Cumulative sum of CRPS

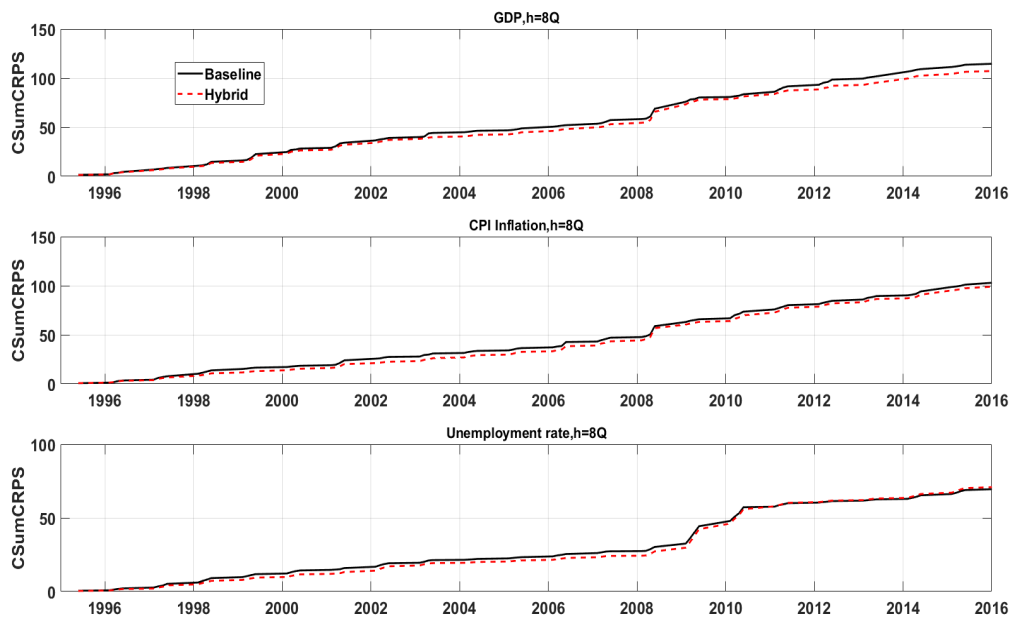


Figure 2.4: Results of the small TVP-VAR SV: cumulative squared errors (top) and cumulative CRPS (bottom).

Table 2.1: Out-of-sample point forecasting performance: small BVAR est. 1960.

Full sample (recursive evaluation: 1994.Q1–2016.Q4)								
	Small VAR				Small VAR with SV			
	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
GDP								
Raw	4.35	7.19	6.55	6.20	3.67	6.68	6.61	6.26
Relative MSE								
Baseline/Raw	0.62***	1.01	1.05**	1.02	0.74**	1.00	1.01	1.00
Hybrid/Raw	0.62***	0.94	0.91	1.00	0.74**	0.90**	0.86*	0.98
Hybrid/Baseline	1.00	0.93*	0.86*	0.98*	1.00	0.90**	0.85**	0.98
CPI								
Raw	3.11	5.61	6.70	8.52	2.94	5.26	5.50	6.50
Relative MSE								
Baseline/Raw	0.33***	0.90**	0.95	0.92*	0.35***	0.86***	0.99	0.94***
Hybrid/Raw	0.33***	0.83***	0.71***	0.61***	0.35***	0.85***	0.88*	0.79***
Hybrid/Baseline	1.00	0.92*	0.75***	0.67***	1.00	1.00	0.89**	0.83***
UR								
Raw	0.05	0.64	2.41	3.78	0.05	0.60	2.39	3.75
Relative MSE								
Baseline/Raw	0.29***	0.73*	0.92	0.95	0.32***	0.73**	0.90*	0.96
Hybrid/Raw	0.29***	0.75**	0.90	0.95	0.32***	0.74**	0.87	0.97
Hybrid/Baseline	1.00	1.03	0.98	0.99	1.00	1.01	0.97	1.00
FFR								
Raw	0.19	2.11	6.12	10.16	0.10	2.12	7.17	11.66
Relative MSE								
Baseline/Raw	0.03***	0.68***	0.87**	0.93***	0.05***	0.63***	0.85***	0.92**
Hybrid/Raw	0.03***	0.77**	0.77*	0.65***	0.05***	0.64***	0.68***	0.57***
Hybrid/Baseline	1.00	1.13*	0.88*	0.70***	1.00	1.02	0.80**	0.62***
Credit spread								
Raw	0.09	0.67	1.14	1.28	0.08	0.73	1.30	1.53
Relative MSE								
Baseline/Raw	0.77*	0.95***	0.97***	0.99	0.89**	0.99	0.98**	0.99*
Hybrid/Raw	0.80	0.94*	0.85***	0.85***	0.89**	0.93***	0.80***	0.72***
Hybrid/Baseline	1.04	1.00	0.87***	0.86***	0.99	0.94**	0.81***	0.73***

Notes: GDP: quarterly annualized real GDP growth rate; CPI: quarterly annualized inflation rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; credit spread: in levels. The raw forecast is defined as the unconditional forecast from the VAR. The baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). The hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with a constant variance and the right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the rows labeled ‘Raw’ are the mean squared error (MSE), while the three rows below report relative MSEs: baseline relative to raw, hybrid relative to raw, and hybrid relative to baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator, with the test statistic adjusted for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. The test statistics use two-sided standard normal critical values.

Table 2.2: Out-of-sample density forecasting performance: small BVAR est. 1960.

Full sample (recursive evaluation: 1994.Q1-2016.Q4)								
	Small VAR				Small VAR with SV			
	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
GDP								
Raw	1.21	1.48	1.43	1.41	1.08	1.39	1.42	1.42
Relative CRPS								
Baseline - Raw	-0.28***	0.01	0.05**	0.01	-0.15***	0.00	0.01	0.00
Hybrid - Raw	-0.29***	-0.04	-0.02	-0.01	-0.15***	-0.06**	-0.05	0.00
Hybrid - Baseline	0.00	-0.05*	-0.07	-0.02	0.00	-0.06**	-0.07	0.01
CPI								
Raw	0.95	1.22	1.42	1.57	0.95	1.20	1.34	1.46
Relative CRPS								
Baseline - Raw	-0.38***	-0.07***	-0.04	-0.02	-0.37***	-0.11***	-0.03	-0.05**
Hybrid - Raw	-0.38***	-0.12***	-0.22***	-0.26***	-0.37***	-0.11***	-0.09	-0.12***
Hybrid - Baseline	0.00	-0.05***	-0.18***	-0.24***	0.00	0.00	-0.06*	-0.07*
UR								
Raw	0.12	0.40	0.80	1.07	0.12	0.37	0.79	1.05
Relative CRPS								
Baseline - Raw	-0.05***	-0.06**	-0.04	-0.02	-0.05***	-0.06**	-0.05*	-0.02
Hybrid - Raw	-0.05***	-0.06**	-0.05	-0.02	-0.05***	-0.05**	-0.06	0.00
Hybrid - Baseline	0.00	0.00	-0.01	0.01	0.00	0.00	-0.01	0.02
FFR								
Raw	0.28	0.84	1.43	1.87	0.17	0.78	1.54	2.04
Relative CRPS								
Baseline - Raw	-0.24***	-0.13***	-0.08*	-0.06**	-0.13***	-0.16***	-0.13***	-0.06**
Hybrid - Raw	-0.24***	-0.10**	-0.15*	-0.33***	-0.13***	-0.15***	-0.27***	-0.43***
Hybrid - Baseline	0.00	0.03*	-0.07	-0.28***	0.00	0.01*	-0.14**	-0.37***
Credit spread								
Raw	0.15	0.43	0.59	0.65	0.14	0.45	0.66	0.75
Relative CRPS								
Baseline - Raw	-0.01**	-0.01***	-0.01*	0.00	-0.01**	-0.01***	-0.01	0.00
Hybrid - Raw	-0.01*	-0.01	-0.06***	-0.07***	-0.01**	-0.02***	-0.08***	-0.11***
Hybrid - Baseline	0.00	0.00	-0.05***	-0.07***	0.00	-0.02***	-0.08***	-0.11***

Notes: GDP: quarterly annualized real GDP growth rate; CPI: quarterly annualized inflation rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; credit spread: in levels. The raw forecast is defined as the unconditional forecast from the VAR. The baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). The hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with a constant variance and the right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the rows labeled 'Raw' are the mean cumulative ranked probability score (CRPS), while the three rows below report relative CRPS values: baseline relative to raw, hybrid relative to raw, and hybrid relative to baseline. A lower value for CRPS is preferable, meaning that a negative value in the row labeled 'Hybrid - Baseline' suggests that, on average, the hybrid forecast is more accurate than the baseline forecast. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator, with the test statistic adjusted for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. The test statistics use two-sided standard normal critical values.

Table 2.3: Out-of-sample point forecasting performance: small BVAR est. 1985.

Full sample (recursive evaluation: 1994.Q1-2016.Q4)								
	Small VAR				Small VAR with SV			
	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
GDP								
Raw	3.43	6.23	6.65	6.46	3.42	6.07	6.51	6.62
Relative MSE								
Baseline/Raw	0.79**	0.93	1.00	1.03***	0.79**	0.96	1.01	1.00
Hybrid/Raw	0.79**	0.96	0.89*	0.96	0.79**	0.98	0.85***	0.92**
Hybrid/Baseline	1.00	1.02	0.89**	0.93*	1.00	1.02	0.84***	0.92*
CPI								
Raw	2.95	4.71	5.11	6.12	2.77	4.79	4.67	5.50
Relative MSE								
Baseline/Raw	0.35***	1.00	1.00	1.00	0.37***	0.95***	1.05	1.02
Hybrid/Raw	0.35***	0.95	0.88***	0.87***	0.37***	0.91***	0.97	0.89***
Hybrid/Baseline	1.00	0.95***	0.88***	0.87**	1.00	0.96*	0.93**	0.88***
UR								
Raw	0.04	0.48	2.22	3.94	0.04	0.46	2.09	3.84
Relative MSE								
Baseline/Raw	0.35***	0.75*	0.84	0.93	0.35***	0.76**	0.86	0.92*
Hybrid/Raw	0.35***	0.86	0.96	0.99	0.35***	0.87***	0.97	0.96
Hybrid/Baseline	1.00	1.14	1.14	1.06	1.00	1.15*	1.13	1.03
FFR								
Raw	0.08	1.95	6.79	10.58	0.08	1.95	6.85	11.33
Relative MSE								
Baseline/Raw	0.06***	0.57***	0.81**	0.94	0.06***	0.57***	0.81***	0.93*
Hybrid/Raw	0.06***	0.57***	0.71**	0.75	0.06***	0.57***	0.72***	0.76**
Hybrid/Baseline	1.00	1.00	0.87*	0.80*	1.00	1.00	0.89	0.82**
Credit spread								
Raw	0.09	0.63	1.06	1.12	0.09	0.66	1.09	1.19
Relative MSE								
Baseline/Raw	0.72**	0.91**	1.00	1.07	0.84**	0.91**	0.99	1.00
Hybrid/Raw	0.73**	0.93	0.91	0.97	0.82*	0.90**	0.94	0.92*
Hybrid/Baseline	1.02*	1.02	0.91	0.91*	0.98	0.99	0.95	0.92*

Notes: GDP: quarterly annualized real GDP growth rate; CPI: quarterly annualized inflation rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; credit spread: in levels. The raw forecast is defined as the unconditional forecast from the VAR. The baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). The hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with a constant variance and the right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the rows labeled 'Raw' are the mean squared error (MSE), while the three rows below report relative MSEs: baseline relative to raw, hybrid relative to raw, and hybrid relative to baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator, with the test statistic adjusted for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. The test statistics use two-sided standard normal critical values.

Table 2.4: Out-of-sample density forecasting performance: small BVAR est. 1985.

Full sample (recursive evaluation: 1994.Q1–2016.Q4)								
	Small VAR				Small VAR with SV			
	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
GDP								
Raw	1.04	1.37	1.41	1.37	1.05	1.36	1.43	1.43
Relative CRPS								
Baseline - Raw	-0.11**	-0.03	0.01	0.04***	-0.13***	-0.03*	-0.01	0.00
Hybrid - Raw	-0.11**	-0.03	-0.10*	-0.02	-0.13***	-0.03	-0.10***	-0.03*
Hybrid - Baseline	0.00	0.00	-0.11**	-0.06**	0.00	0.01	-0.09***	0.03*
CPI								
Raw	0.94	1.14	1.20	1.33	0.92	1.14	1.16	1.26
Relative CRPS								
Baseline - Raw	-0.36***	-0.02	0.00	0.01	-0.35***	-0.04***	0.00	-0.01
Hybrid - Raw	-0.36***	-0.06**	-0.10***	-0.11***	-0.35***	-0.08***	-0.05**	-0.06***
Hybrid - Baseline	0.00	-0.04**	-0.11***	-0.13***	0.00	-0.04**	-0.04**	-0.05***
UR								
Raw	0.12	0.36	0.80	1.13	0.11	0.35	0.76	1.11
Relative CRPS								
Baseline - Raw	-0.05***	-0.05*	-0.04	-0.02	-0.04***	-0.05***	-0.05*	-0.04*
Hybrid - Raw	-0.05***	-0.03	0.00	0.00	-0.04***	-0.03**	-0.02	-0.04
Hybrid - Baseline	0.00	0.02*	0.04*	0.02	0.00	0.02**	0.04	0.00
FFR								
Raw	0.16	0.78	1.48	1.86	0.15	0.75	1.47	1.92
Relative CRPS								
Baseline - Raw	-0.12***	-0.19***	-0.12***	-0.05	-0.12***	-0.19***	-0.15***	-0.08*
Hybrid - Raw	-0.12***	-0.20***	-0.21**	-0.23	-0.12***	-0.19***	-0.22***	-0.27**
Hybrid - Baseline	0.00	0.00	-0.09	-0.19*	0.00	0.00	-0.07	-0.20**
Credit spread								
Raw	0.15	0.43	0.57	0.62	0.15	0.43	0.57	0.62
Relative CRPS								
Baseline - Raw	-0.02**	-0.02*	0.00	0.02	-0.01**	-0.02***	-0.01**	-0.01***
Hybrid - Raw	-0.02**	-0.02	-0.03*	-0.03	-0.01**	-0.02**	-0.03**	-0.02*
Hybrid - Baseline	0.00	0.00	-0.03*	-0.04*	0.00	0.00	-0.02**	-0.02*

Notes: GDP: quarterly annualized real GDP growth rate; CPI: quarterly annualized inflation rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; credit spread: in levels. The raw forecast is defined as the unconditional forecast from the VAR. The baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). The hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with a constant variance and the right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the rows labeled ‘Raw’ are the mean cumulative ranked probability score (CRPS), while the three rows below report relative CRPS values: baseline relative to raw, hybrid relative to raw, and hybrid relative to baseline. A lower value for CRPS is preferable, meaning that a negative value in the row labeled ‘Hybrid - Baseline’ suggests that, on average, the hybrid forecast is more accurate than the baseline forecast. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator, with the test statistic adjusted for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. The test statistics use two-sided standard normal critical values.

Table 2.5: Out-of-sample point forecasting performance: small TVP-VAR.

Full sample (recursive evaluation: 1994.Q1–2016.Q4)								
	Small TVP VAR				Small TVP VAR with SV			
	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
GDP								
Raw	4.65	6.91	6.33	6.00	3.99	6.55	6.24	6.08
Relative MSE								
Baseline/Raw	0.58***	0.95**	1.03	1.00	0.68***	0.92**	0.98	1.01*
Hybrid/Raw	0.58***	0.98	0.93	1.04*	0.68***	0.86**	0.90	1.01
Hybrid/Baseline	1.00	1.03	0.90	1.05	1.00	0.94	0.92	0.99
CPI								
Raw	3.41	6.01	6.65	7.82	3.10	5.21	5.45	6.25
Relative MSE								
Baseline/Raw	0.30***	0.91**	0.95	0.95***	0.33***	0.93**	0.94***	0.96**
Hybrid/Raw	0.30***	0.82***	0.69***	0.67***	0.33***	0.88***	0.82***	0.79***
Hybrid/Baseline	1.00	0.91***	0.73***	0.70**	1.00	0.94*	0.88**	0.82***
UR								
Raw	0.05	0.49	1.98	3.16	0.05	0.50	2.04	3.36
Relative MSE								
Baseline/Raw	0.32***	0.82***	0.93*	0.98	0.34***	0.82**	0.93	0.98
Hybrid/Raw	0.32***	0.91	1.04	1.13	0.34***	0.84***	1.00	1.09
Hybrid/Baseline	1.00	1.11	1.12	1.16	1.00	1.03	1.08	1.11

Notes: GDP: quarterly annualized real GDP growth rate; CPI: quarterly annualized inflation rate; UR: unemployment rate in levels. The raw forecast is defined as the unconditional forecast from the VAR. The baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). The hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with a constant variance and the right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the rows labeled ‘Raw’ are the mean squared error (MSE), while the three rows below report relative MSEs: baseline relative to raw, hybrid relative to raw, and hybrid relative to baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator, with the test statistic adjusted for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. The test statistics use two-sided standard normal critical values.

Table 2.6: Out-of-sample density forecasting performance: small TVP-VAR.

Full sample (recursive evaluation: 1994.Q1-2016.Q4)								
	Small TVP VAR				Small TVP VAR with SV			
	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$	$h = 1Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
GDP								
Raw	1.27	1.49	1.44	1.42	1.14	1.42	1.35	1.33
Relative CRPS								
Baseline - Raw	-0.34***	-0.04**	0.01	-0.01	-0.21***	-0.07**	0.00	0.01
Hybrid - Raw	-0.34***	-0.01	-0.04	0.02**	-0.21***	-0.12**	-0.05	0.02
Hybrid - Baseline	0.00	0.02	-0.06	0.03*	0.00	-0.05*	-0.05	0.01
CPI								
Raw	1.03	1.30	1.45	1.54	1.01	1.18	1.27	1.33
Relative CRPS								
Baseline - Raw	-0.45***	-0.08***	-0.04**	-0.03***	-0.42***	-0.07***	-0.05***	-0.02***
Hybrid - Raw	-0.45***	-0.11**	-0.18**	-0.19**	-0.42***	-0.11***	-0.12**	-0.10**
Hybrid - Baseline	0.00	-0.03	-0.14**	-0.15*	0.00	-0.04***	-0.06*	-0.08**
UR								
Raw	0.12	0.38	0.76	0.99	0.12	0.38	0.80	1.08
Relative CRPS								
Baseline - Raw	-0.05***	-0.04***	-0.03**	-0.01	-0.05***	-0.04***	-0.02	0.00
Hybrid - Raw	-0.05***	-0.04**	-0.02	0.04	-0.05***	-0.05***	-0.02	0.01
Hybrid - Baseline	0.00	0.01	0.02	0.05	0.00	0.00	0.00	0.01

Notes: GDP: quarterly annualized real GDP growth rate; CPI: quarterly annualized inflation rate; UR: unemployment rate in levels. The raw forecast is defined as the unconditional forecast from the VAR. The baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). The hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with a constant variance and the right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the rows labeled 'Raw' are the mean cumulative ranked probability score (CRPS), while the three rows below report relative CRPS values: baseline relative to raw, hybrid relative to raw, and hybrid relative to baseline. A lower value for CRPS is preferable, meaning that a negative value in the row labeled 'Hybrid - Baseline' suggests that, on average, the hybrid forecast is more accurate than the baseline forecast. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator, with the test statistic adjusted for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. The test statistics use two-sided standard normal critical values.

Table 2.7: CPI inflation real-time out-of-sample point forecasting performance.

Panel A: Full sample (recursive evaluation: 1994.Q1–2016.Q4)								
	Medium VAR est. 1960				Medium VAR with SV est. 1960			
CPI inflation	$h = 2Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
MSE								
Hybrid	4.36	4.58	4.55	4.91	4.50	4.47	4.57	4.95
Relative MSE								
Hybrid/RW (AO)	0.86*	0.90*	0.84*	0.93**	0.88*	0.89**	0.84*	0.93**
Hybrid/UCSV (SW)	0.98	1.00	0.94	1.00	1.00	0.99	0.94	1.00
Hybrid/FW	0.98	1.00	0.93**	0.95**	1.00	0.99	0.93**	0.95**
Hybrid/SPF	1.08	1.05	N/A	N/A	1.10	1.04	N/A	N/A
Relative CRPS								
Baseline - UCSV	0.02	0.08***	0.12***	0.21**	0.04	0.00	0.01	0.06
Hybrid - UCSV	-0.01	0.03	0.02	0.07*	0.01	0.02	0.00	0.08

Panel B: Pre-crisis sample (recursive evaluation: 1994.Q1–2006.Q4)								
	Medium VAR est. 1960				Medium VAR with SV est. 1960			
CPI inflation	$h = 2Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$	$h = 12Q$
MSE								
Hybrid	1.47	1.57	1.79	1.86	1.35	1.45	1.71	1.89
Relative MSE								
Hybrid/RW (AO)	0.96	0.91	0.79	0.88	0.89	0.84*	0.75*	0.89
Hybrid/UCSV (SW)	1.01	1.00	0.91	0.93	0.93	0.92	0.87	0.94
Hybrid/FW	1.01	0.98	0.94	0.87	0.92	0.90	0.89	0.88
Hybrid/SPF	0.99	0.94	N/A	N/A	0.91	0.87	N/A	N/A
Relative CRPS								
Baseline - UCSV	0.01	0.06	0.11	0.32***	0.00	-0.02	-0.01	0.08**
Hybrid - UCSV	0.01	0.04	0.06	0.11	-0.02	-0.03	-0.05	0.02

Notes: The first row in each panel reports the mean squared error (MSE) that corresponds to the hybrid forecast obtained from a medium BVAR. Rows 2 to 5 report relative MSEs, meaning that a ratio of less than 1 indicates that, on average, the hybrid forecast is more accurate than the respective univariate forecast. Row 6 reports the relative CRPS values of the baseline density forecasts relative to UCSV, where a negative value suggests that the baseline density forecast is more accurate on average. Row 7 reports the relative CRPS of the hybrid density forecast relative to UCSV, with a negative value suggesting that the hybrid density forecast is more accurate on average. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator, with the test statistic adjusted for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. The test statistics use two-sided standard normal critical values. All models use the SPF nowcast for the one-step-ahead forecast; as a result, the relative MSE is equal to one and is not reported.

Chapter 3

Real-Time Density Nowcasts of US Inflation: A Model-Combination Approach

Based on the paper: Knotek, E.S., II and Zaman, S. (2020). Real-Time density nowcasts of US inflation: A model-combination approach. Federal Reserve Bank of Cleveland, Working Paper No. 20-31.

3.1 Introduction

Inflation developments are of interest to policymakers, forecasters, financial market participants, and the general public. This interest includes not only the point forecast but also the range of potential inflation outcomes and their probability of occurring—i.e., the density forecast. Building on the literature that finds that the accuracy of multistep point forecasts can be improved by conditioning on high-quality point nowcasts, Krüger, Clark, and Ravazzolo (2017) and Tallman and Zaman (2020) document that conditioning quarterly macroeconomic models with both nowcast means and nowcast densities leads to improvements in the accuracy of multistep point and density forecasts, especially for inflation.¹ Realizing these gains in practice requires relatively accurate nowcast means and nowcast densities for inflation. Previous research by Modugno (2013), Monteforte and Moretti (2013), Breitung and Røling (2015), Knotek and Zaman (2017), and Clement (2017) has developed mixed-frequency approaches to nowcast U.S. inflation, with an exclusive focus on point nowcast accuracy. In this chapter, we develop a flexible framework that uses model-combination strategies with three classes of mixed-frequency models to generate highly accurate point and density nowcasts for U.S. inflation.

The past two decades have seen considerable growth in the density forecasting and density

¹E.g., Faust and Wright (2013) and Knotek and Zaman (2019) find that conditioning quarterly macro models with more accurate jumping-off points from external nowcasts improves multistep forecast accuracy.

nowcasting literature.² Density forecasts help to illuminate the balance of risks around point forecasts (e.g., Rossi, 2014; Mazzi, Mitchell, and Montana, 2014) and strengthen the “credibility” of forecasts (Wright, 2019). However, recent work on density nowcasting has focused on real GDP growth and other indicators of real economic activity. The paper that comes closest to the notion of inflation density nowcasting is Garratt, Mitchell, and Vahey (2014), but there the nowcasts are one-step-ahead forecasts from models estimated with quarterly data; intra-quarterly daily, weekly, and monthly data are not used, and hence the density nowcast estimates are largely unchanged during the quarter.

We contribute to the density nowcasting literature by proposing a flexible framework to produce both point and density nowcast estimates for U.S. headline and core inflation measures: CPI inflation, core CPI inflation, PCE inflation, and core PCE inflation. Our flexible framework is based on model combinations across three classes of mixed-frequency models that previous research has shown produce high-quality point nowcasts for inflation; for one of the model classes, we develop a procedure to generate the density nowcasts. By using mixed-frequency models, we can produce inflation density nowcasts at a trading-day frequency that take advantage of high-frequency data and update as information accumulates over the course of a month or a quarter.

We use model combinations because the characterization of uncertainty from a single (possibly misspecified) model could be too restrictive. In addition, combining density estimates across a range of models provides a flexible density that can potentially accommodate non-Gaussian features such as skewness and kurtosis, which may more closely approximate the true density.³ Inspired by previous research on density forecast combinations (e.g., Bache et al., 2011; Aastveit et al., 2014; Garratt, Mitchell, and Vahey, 2014), we combine the density nowcasts in a two-stage procedure.⁴ In stage 1, density nowcasts coming from different model specifications within each of the three model classes are combined. In stage 2, we combine across the three stage 1 combinations to form a “grand” combination. As illustrated in Aastveit et al. (2014), an advantage of the two-stage procedure is that it directly accommodates instabilities and uncertainty about model specification within each model class.

Combining densities requires a functional form for the aggregation and weights to apply to the different densities. Previous research has used either the linear opinion pool or the logarithmic opinion pool as the functional form for aggregation, with some researchers using both methods and presenting results for the approach that is more accurate over some evaluation period. Instead of enforcing a particular functional form for aggregation at the outset, we devise

²See Tay and Wallis (2000) for a survey on density forecasting, including its application in macroeconomics and finance, and Aastveit et al. (2018) for a more recent survey on density forecasting and density combinations.

³Alternatively, one could flexibly characterize uncertainty using a single mixed-frequency model featuring stochastic volatility and estimate the model with Bayesian methods, as in Carriero, Clark, and Marcellino (2015b) for real GDP growth or Koop, McIntyre, and Mitchell (2020) for U.K. regional indicators.

⁴Our general strategy for combining densities from many models is similar in spirit to the approach in Aastveit et al. (2014) for density nowcasting GDP, but with many differences in implementation. Chernis and Sekkel (2018) employ a similar two-stage procedure to produce point nowcast combinations for real economic indicators.

and implement a novel flexible aggregation strategy that lets the data dynamically determine which of the two functional forms it prefers. A potential advantage of this flexibility is that it allows for the possibility of switching between the two functional forms, at different points within a month or a quarter based on the nowcast origin or at different points in time during the sample as more observations become available. This novel flexible aggregation strategy can be used broadly in multistep forecasting applications when combining point or density forecasts.

The literature has considered a variety of weighting schemes to use when combining density forecasts, in part because no single scheme has been shown to work “best” under all circumstances. The bulk of the density combination literature has considered a limited number of weighting schemes—although Krüger (2015) and Ganics (2017) are exceptions—with the scheme based on recursive updating of past performance using the log score metric being the most popular. In contrast, we consider a relatively large number of weighting schemes, ranging from equal weights to schemes based on past predictive performance to “optimal” schemes that optimize some loss function over a historical sample. In most cases, these weights dynamically update over time to learn from past performance.

Using high-frequency real-time data over the evaluation period 2000-2015, we conduct a comprehensive set of out-of-sample inflation density nowcasting exercises to assess our flexible framework using a variety of inflation measures, inflation rates, weighting strategies, and mixed-frequency model classes. This examination reveals that combining individual densities generally helps improve density nowcast accuracy, and as information accumulates over the course of a month or a quarter, the accuracy of the combined density nowcasts and the associated point nowcasts steadily improves. We also document evidence of dynamic model switching, which highlights the importance of combining estimates from a range of models to circumvent the instability issues from a single model. But it matters how the densities are combined: not all combination methods improve accuracy compared with the best-performing individual densities. The grand combinations based on our flexible aggregation strategy and the log score weighting scheme, which relies on past predictive performance, or the “optimal” weighting scheme of Conflitti, De Mol, and Giannone (2015) are among the best performing in terms of relative accuracy for headline inflation and are well calibrated. The Ganics (2017) weighting scheme, which optimizes the calibration fit, produces the best calibrated densities for headline inflation, but its relative accuracy is inferior to the log score or Conflitti, De Mol, and Giannone (2015) weighting schemes. In the case of core inflation, all weighting schemes generate comparable accuracy of point and density nowcasts. Overall, the accuracy of the implied point nowcasts from the grand combination matches the accuracy of the best performing mixed-frequency model of Knotek and Zaman (2017) and is more accurate in the case of core PCE inflation.

Our empirical results indicate evidence of both time-varying skewness and kurtosis in the predictive densities of inflation measures, suggesting that asymmetries and fat tails are an empirical feature of inflation data. Combination methods that produce density estimates derived from a richer set of models tend to display a higher degree of skewness and kurtosis in the

predictive densities. We also find evidence of time-varying variances (i.e., uncertainty in the nowcast estimates) that echo the broader patterns reported elsewhere in the inflation uncertainty literature using stochastic volatility models (e.g., Carriero, Clark, and Marcellino, 2019; Knotek, Zaman, and Clark, 2015).

Finally, we conduct a horse race with the Survey of Professional Forecasters (SPF) for point and density nowcast accuracy. Our grand combination’s density nowcasts provide superior point and density nowcasts for CPI inflation and PCE inflation. For core CPI inflation and core PCE inflation, our grand combination’s nowcasting performance is competitive with the SPF. The ability of our proposed framework to generate highly accurate point and density nowcasts of inflation is a useful outcome for practitioners.

The chapter proceeds as follows. Section 3.2 describes the mixed-frequency models. Section 3.3 discusses the combination methods to combine individual densities. Section 3.4 describes the real-time data. Section 3.5 discusses the nowcast evaluation strategy. Section 3.6 presents empirical results. Section 3.7 compares the accuracy of the combined density nowcasts to SPF. Section 3.8 concludes.

3.2 Mixed-Frequency Models

Building on the literature that has shown that relatively parsimonious approaches dominate more sophisticated approaches for nowcasting and near-term forecasting of inflation (e.g., Koop and Korobilis, 2012; Knotek and Zaman, 2017), we consider three classes of mixed-frequency models that relate aggregate inflation to its components and to a limited number of other indicators.⁵ To economize on space, we briefly discuss the models and procedures for constructing their density nowcasts here. The supplementary appendix B provides detailed explanations of the models and the bootstrap algorithms for constructing the density nowcasts.

3.2.1 Deterministic Model Switching (DMS)

Knotek and Zaman (2017) construct a mixed-frequency model for inflation point nowcasting that relies on a small number of variables and combines univariate and multivariate regressions. The model uses disaggregate and aggregate variables to construct nowcasts for the aggregate, but the disaggregate information is used only if it is available and deemed useful.⁶ The latter aspect gives rise to time-varying coefficients that change in a deterministic fashion based on the available information set, which we label as deterministic model switching (DMS).

⁵These model classes have also been applied in other contexts, such as nowcasting GDP growth. We restrict attention to models that are well understood by econometricians and economic forecasters, but one could extend the model pool to include a larger and broader set of model classes, including machine learning models.

⁶Hendry and Hubrich (2011) and Ravazzolo and Vahey (2014) focus on multistep inflation forecasting using disaggregates.

Monthly inflation rates π_t for month t are modeled via a general representation:

$$A_{s(\tau)}Z_t = B_{s(\tau)} + C_{s(\tau)}X_t + \sum_{j=1}^J D_{j,s(\tau)}Z_{t-j} + \varepsilon_{s(\tau),t} \quad (3.1)$$

For headline inflation, Z_t is a vector of aggregates—CPI inflation, π_t^{CPI} , and PCE inflation, π_t^{PCE} —and X_t is a vector of disaggregate components comprising gasoline inflation ($\pi_t^{Gasoline}$), food inflation (π_t^{Food}), and core inflation rates ($\pi_t^{CoreCPI}$ and $\pi_t^{CorePCE}$).

The coefficient matrices A , B , C , and D_j can potentially vary over time with the information set $s(\tau)$, where τ is the point in time at which the nowcast is being made. As is common in nowcasting applications, we produce and evaluate nowcasts at different points in time for a given month; we use the notation τ to capture the former and t to capture the latter.⁷ The information set $s(\tau)$ captures the data flow from statistical agencies that is available at that point in time τ ; in Section 3.4, we describe the data flow at the points τ for which we evaluate the nowcasts. We explicitly include the information set, $s(\tau)$ in equation (3.1) to emphasize the dependence of the coefficients on the available information. As some of these coefficients can take on values of zero, the model can switch over time between univariate and multivariate forms depending on the available information set. To illustrate the type of deterministic model switching that occurs, based on the available information set $s(\tau)$, for example, the nowcast for π_t^{PCE} in month t is a function of actual π_t^{CPI} in month t via $A_{s(\tau)}$, if π_t^{CPI} is available from statistical agencies and hence part of the information set; if actual π_t^{CPI} is not available, the nowcast for π_t^{PCE} in month t is a function of disaggregates' nowcasts included in X_t via $C_{s(\tau)}$. If the vector of disaggregates is incomplete for month t , then $C_{s(\tau)} = 0$. High-frequency data on gasoline and oil prices are used to construct a gasoline inflation nowcast via an auxiliary model, while food inflation nowcasts are derived using a univariate AR specification. Core inflation rates are modeled using equation (3.1), where no disaggregates are used and nowcasts are formed either via univariate AR models or bridge regressions if $\pi_t^{CoreCPI}$ is available for month t while $\pi_t^{CorePCE}$ is not. This mixed-frequency model switches between univariate and multivariate regressions depending on the available information within a month or a quarter.

We innovate on the DMS approach for constructing inflation point nowcasts in two ways. First and foremost, we devise and implement parametric block wild bootstrap algorithms to produce density nowcasts for the DMS framework. When the DMS selects the multivariate regression model that uses disaggregates, the density nowcasts are constructed by: (1) constructing density estimates for each of the three disaggregates (core inflation, food inflation, and gasoline inflation); and (2) combining the density estimates using the weights in $C_{s(\tau)}$ to construct the density nowcast for aggregate inflation, similar to the combination approaches in Ravazzolo and Vahey (2014) and Tallman and Zaman (2017) but in this case with an application to nowcasting. While Knotek and Zaman (2017) estimate the model using short rolling

⁷Note that the timing of the nowcasts τ can occur before, during, and after the target month t —the backcasts made after the month can still precede the official release of data by the statistical agencies.

windows, which leads to very flexible parameters and incorporates changing volatility in a parsimonious way, we consider density combinations that allow for a variety of rolling or expanding estimation windows, as discussed below. The supplementary appendix B contains more details on the model and the construction of the density nowcasts.

3.2.2 Mixed Data Sampling (MIDAS)

Monteforte and Moretti (2013) generate inflation point nowcasts using a MIDAS model with leads, which is a reduced-form regression relating a low-frequency variable to high-frequency variable(s).⁸ In our application, oil and gasoline prices act as the high-frequency leads, while monthly inflation is at a lower frequency. To prevent parameter proliferation, MIDAS works with distributed lag polynomial operators that reduce the estimation to a smaller number of parameters. The model is estimated using nonlinear least squares.

The MIDAS model with leads for inflation at time τ for month $t + h$, $\pi_{t+h,\tau}$, takes the form

$$\pi_{t+h,\tau} = \alpha_{(h),\tau} + \sum_{j=0}^{P(M)-1} \chi_{j+1,(h)} \pi_{t-j} + \sum_{j=0}^{P(M)-1} \gamma_{j+1,(h)} Z_{t-j} + \beta_{(h),\tau} \sum_{j=0}^{P(HF)-1} \omega_{P(HF)-j} (\theta_{(h)}^{HF}) X_{P(HF)-j,t+1}^{HF} + e_{t+h,\tau} \quad (3.2)$$

where Z includes other monthly variables; $P(M)$ is the number of lags of the monthly regressors (we use 1); and $P(HF)$ is the number of high-frequency observations, $X_{1,t+1}^{HF}, \dots, X_{P(HF),t+1}^{HF}$ in month $t + 1$ (i.e., the target nowcast month). The coefficients are independently estimated for each forecast horizon (h).⁹ The assumption $\sum_{j=0}^{P(HF)-1} \omega_{P(HF)-j} (\theta_h^{HF}) = 1$ helps identify $\beta_{(h),\tau}$.

We extend Monteforte and Moretti's (2013) work on inflation point nowcasts by generating inflation density nowcasts from MIDAS models; see the supplementary appendix for details. These density nowcasts are constructed by drawing errors from a normal distribution with a standard deviation coming from past residuals after rescaling to correct the variance. Aastveit, Foroni, and Ravazzolo (2017) show that this approach is slightly inferior to the block wild bootstrap in their application when nowcasting GDP, but it provides a substantial computational advantage in our exercises.

⁸Ghysels, Santa-Clara, and Valkanov (2005, 2006) popularized these models; Clements and Galvão (2008) is an influential paper on the application of MIDAS to macroeconomic forecasting. These models are increasingly used for nowcasting macroeconomic indicators across the globe (e.g., see Allan et al., 2014, for the economy of Scotland).

⁹In our exercises, h ranges from 1 to 2 for nowcasting monthly inflation and from 1 to 4 for nowcasting quarterly inflation.

3.2.3 Dynamic Factor Model (DFM)

Building on Giannone, Reichlin, and Small (2008), mixed-frequency dynamic factor models (DFMs) are widely used for nowcasting. Modugno (2013) uses a DFM to generate inflation point nowcasts from a data set comprising monthly, weekly, and daily data by extracting a common factor at a daily frequency via the estimation method of Bańbura and Modugno (2014). At the trading-day τ frequency, the DFM takes the form

$$y_\tau = Cf_\tau + \varepsilon_\tau, \quad \varepsilon_\tau \sim N(0, \Sigma) \quad (3.3)$$

where y_τ is a vector of observations of mixed frequencies, C is a matrix of loadings, ε_τ is a vector of idiosyncratic components, and f_τ is a vector of unobserved common components that follows

$$Bf_\tau = A(L)f_{\tau-1} + u_\tau, \quad u_\tau \sim N(0, Q) \quad (3.4)$$

where B and $A(L)$ are coefficient matrices that capture factor dynamics, some of which may be time-varying, and u_τ is a vector of residuals. The estimated latent daily factor(s) aggregate to weekly and monthly factors, which are used to construct nowcast estimates for monthly variables, including inflation. Please refer to the supplementary appendix B.1.2. (DFM Model) for details about coefficient matrices B and $A(L)$.

We extend the previous work using DFMs for inflation point nowcasts in order to generate density nowcasts for U.S. inflation. Our density nowcasts are constructed using a standard parametric bootstrapping procedure for factor models, similar to Aastveit et al. (2014), which we detail in the supplementary appendix. This bootstrapping procedure accounts for factor, parameter, and shock uncertainty.

3.2.4 Mixed-Frequency Model Space

Each of the three mixed-frequency model classes requires assumptions about certain elements, such as the size of the rolling windows in the DMS model, the polynomial specification in the MIDAS model, or the number of lags of the factors in the DFM. To account for this uncertainty within each model class, we consider many different specifications, listed in Table 3.1. We employ 132 model variants distributed unequally across our three model classes, with 108 in the DMS class, 12 in the MIDAS class, and 12 in the DFM class.¹⁰ We combine density nowcasts within each model class in “stage 1” combinations, and we then combine the stage 1 combinations into a “grand” combination in stage 2.

¹⁰In preliminary results, we considered 36 specifications for the MIDAS model class (12 each for Beta, BetaNN, and Almon polynomials). This combination performed similarly to the combination in the chapter, but the latter approach greatly reduced computing time. For the DFM class we initially used combinations with one or two factors, giving us 24 specifications. This combination slightly underperformed the combination reported in the chapter using only one factor, consistent with DFM inflation point nowcasting accuracy findings in Modugno (2013).

3.3 Combination Methods

Simple combinations of point forecasts have a long history (see Bates and Granger, 1969) and often perform well, even when compared with weighting schemes that minimize some loss function (e.g., Clark and McCracken, 2010). In contrast, the density forecast combination literature has generally documented performance gains from optimal weighting schemes and schemes based on past predictive performance over the use of equal weights.¹¹ Combining candidate density estimates requires both a functional form to use in combining the densities and a mechanism for deriving the weights to place on each density. We present a novel functional form to aggregate our candidate density nowcasts and consider a variety of weighting schemes in this context. To be clear about our approach, in stage 1 we use a particular functional form and weighting scheme to combine the individual densities within a model class, for each of the three model classes; and then in stage 2, we use the same functional form and weighting scheme to combine the densities from the three model classes into the grand combination.

3.3.1 Functional Forms for Aggregation

We consider three functional forms or aggregation methods: the linear opinion pool, the logarithmic opinion pool, and a novel flexible method that combines the previous two functional forms in a data-dependent way.

The linear opinion pool is the weighted linear combination of individual component densities and is widely used (e.g., Bache et al., 2011; Aastveit et al., 2014; Mazzi, Mitchell, and Montana, 2014; Ravazzolo and Vahey, 2014). If there are M models, then

$$p_{\tau,t,h}^{LIN}(y_t) = \sum_{i=1}^M w_{\tau,t,i,h} f_{\tau,t,i,h}(y_t|I_{\tau,t,i}) \quad (3.5)$$

where $p_{\tau,t,h}^{LIN}(y_t)$ is the combined linear pool predictive density for variable y at different points in time τ for month t for forecast horizon h . The density forecast from model i , $f_{\tau,t,i,h}(y_t|I_{\tau,t,i})$, is conditional on information set $I_{\tau,t,i}$ which can differ across models. The potentially time-varying, nonnegative weights $w_{\tau,t,i,h}$ are recursively updated at each forecast origin based on some criteria that we discuss below and sum to 1.

This linear form of aggregation implies that if all the individual component densities are distributed normally with different mean and variance, then the combined density will be non-normal (or a mixture-normal). The nonnormal characteristic for the combined density via the linear pool is desirable if the unknown true density is nonnormal. An advantage of a mixture-normal distribution is that it permits skewness and kurtosis.

The logarithmic opinion pool is the geometric weighted average of the individual component

¹¹For some examples, see Hall and Mitchell (2007), Jore, Mitchell, and Vahey (2010), Bache et al. (2011), Bjornland et al. (2011), and Aastveit et al. (2014). Kascha and Ravazzolo (2010) and Garratt, Mitchell, and Vahey (2011) provide counterexamples.

densities:

$$p_{\tau,t,h}^{LOG}(y_t) = \frac{\prod_{i=1}^M f_{\tau,t,i,h}(y_t|I_{\tau,t,i})^{w_{\tau,t,i,h}}}{\int \prod_{i=1}^M f_{\tau,t,i,h}(y_t|I_{\tau,t,i})^{w_{\tau,t,i,h}} dy_t} \quad (3.6)$$

where $p_{\tau,t,h}^{LOG}(y_t)$ is the combined log pool predictive density for variable y at the point in time τ within month t for forecast horizon h . (To economize on notation, we omit h subscripts going forward.) Importantly, the combination based on this geometric functional form assigns a zero probability to a region if any single individual density assigns a zero probability to that region. This may be an undesirable feature because a single incorrectly specified density can significantly influence the specification of the combined density (see Bjornland et al., 2011).¹²

In this chapter, we propose a novel flexible aggregation method that is in the spirit of “letting the data speak” about which of the two functional forms—the linear pool or the logarithmic pool—is preferred.¹³ Specifically, instead of taking a stand on a particular functional form at the outset, we allow flexibility in letting the data determine which of the two functional forms is preferred at every point in time. A potential advantage of this flexibility is that it allows for the possibility of dynamically switching between the two functional forms, at different points in time in the sample (t in our notation above) and at different points in time within a month or a quarter (τ in our notation above), to take advantage of the differing information sets and the high-frequency data flow for nowcasting applications. For example, at the beginning of a month, when uncertainty around the point nowcast for that month would be expected to be higher, since much of the underlying source information is not yet available, the data may prefer the linear opinion pool (i.e., over the historical sample the combined density constructed using the linear opinion pool is more accurate on average than combined density based on logarithmic opinion pool), while later in the month the data may prefer the log opinion pool. This flexible aggregation method could be applied in more general multistep forecasting applications where the functional form is allowed to vary based on the forecasting horizon.

We implement this flexible aggregation method by determining, for each point in time τ for each target month to be nowcasted t , which of the two functional forms has historically produced the more accurate densities. We denote the start and end of the sample by T_0 and T , respectively, and we let D_τ denote the normal data release lags (in number of months) at the point in time τ , which captures the delay in calculating the historical density accuracies.¹⁴ We initialize the flexible functional form by using the linear pool, i.e., if $t \leq T_0 + D_\tau - 1$,

$$\text{functional form}_{\tau,t} = \text{Linear Pool} \quad (3.7)$$

¹²Figure B1 and Figure B2 in the appendix visually contrast the properties of the linear and log opinion pools.

¹³Knüppel and Krüger (2019) and Garratt, Henckel, and Vahey (2019) propose approaches to modify the linear opinion pool but do not consider the hybrid approach that we pursue.

¹⁴In our empirical application, the delay parameter D is dependent on τ , the position within the month when the nowcast is being made. Because we use the third monthly inflation release as the “true” inflation reading for a particular month, the delay $D = 4$ early in the month when nowcasting monthly inflation, because the most recent “true” inflation reading comes from four months prior to the current month being nowcasted. The delay $D = 3$ late in the month, however, as the statistical agencies have released prior months’ inflation numbers by that point.

Subsequently,

$$functional\ form_{\tau,t} = \begin{cases} Linear\ Pool & \text{if } LIN_{\tau,t} \geq LOG_{\tau,t} \\ Logarithm\ Pool & \text{if } LIN_{\tau,t} < LOG_{\tau,t} \end{cases} \quad (3.8)$$

for $t = T_0 + D_\tau, \dots, T$, with

$$LIN_{\tau,t} = \frac{\sum_{s=T_0}^{t-D_\tau} \log(p_{\tau,s}^{LIN}(y_s = y_s^o))}{t - D_\tau - T_0 + 1} \quad (3.9)$$

$$LOG_{\tau,t} = \frac{\sum_{s=T_0}^{t-D_\tau} \log(p_{\tau,s}^{LOG}(y_s = y_s^o))}{t - D_\tau - T_0 + 1} \quad (3.10)$$

where the values for $p^{LIN}(\cdot)$ and p^{LOG} are computed from equation (3.5) and equation (3.6), respectively, and y_t^o is the observed value in month t , which is independent of the point in time when the nowcast was made (τ) and the nowcast or forecast horizon (h).

3.3.2 Weighting Schemes

We consider five weighting schemes, ranging from equal weights to schemes based on average past predictive performance to schemes based on optimization of a specific loss function.

1. Equal weights. Simply, each density i gets a weight

$$w_{\tau,t,i} = \frac{1}{M} \quad (3.11)$$

in the construction of the combination density nowcast. Aastveit et al. (2018) characterize such a combination as a “restrictive finite mixture.”

2. Log-score weights. The logarithmic score (log-score) is the logarithm of the density forecast evaluated at the observation and is widely used to assess the accuracy of density forecasts. Accordingly, it makes sense to derive weights based on the past nowcast performance of the candidate densities using the log-score metric, with more accurate candidate densities receiving larger weights.¹⁵ The weights $w_{\tau,t,i}$ are computed by averaging the past predictive performance using an expanding window for nowcast evaluation. We initialize the weights by

¹⁵This approach amounts to “learning from past mistakes” and is widely used in the density combination literature due to its simplicity (e.g., Gerard and Nimark, 2008; Jore, Mitchell, and Vahey, 2010; Kascha and Ravazzolo, 2010; Bjornland et al., 2011; Garratt et al., 2011; Aastveit et al., 2014; Beckmann et al., 2020). While we use the entire expanding history, we also explored the strategy of computing the weights over rolling 12-month periods, to “learn from recent mistakes” in computing the average score. The density nowcasting accuracy results were similar. We thank Gary Koop for suggesting this exercise.

setting them to equal weights (equation 3.11) if $t \leq T_0 + D_\tau - 1$. For $t = T_0 + D_\tau, \dots, T$:

$$w_{\tau,t,i} = \frac{\exp[\sum_{s=T_0}^{t-D_\tau} \log f_{\tau,s,i}(y_s = y_s^o)]}{\sum_{j=1}^M \exp[\sum_{s=T_0}^{t-D_\tau} \log f_{\tau,s,j}(y_s = y_s^o)]} \quad (3.12)$$

3. Continuous ranked probability score (CRPS). The CRPS is a widely used alternative metric for assessing density forecasts that is also based on past predictive performance. It is popular because it is more robust to outliers compared with the log-score metric and it rewards densities that have probability mass closer to the actual observation. The CRPS score for density nowcast i at the point in time τ within month t , $CRPS_{\tau,t,i}$, is the squared difference between the CDF of the density forecast and the CDF of the actual realizations:

$$CRPS_{\tau,t,i} = \int_{-\infty}^{+\infty} (F_{\tau,t,i}(y_t) - 1\{y_t^o \leq y_t\})^2 \quad (3.13)$$

where $F_{\tau,t,i}$ is the CDF of density nowcast i , $f_{\tau,s,i}(y_t)$, and $1\{y_t^o \leq y_t\}$ is the indicator function equal to 1 if $y_t^o \leq y_t$ and 0 otherwise.¹⁶ The smaller the difference between the two cumulative distributions, the more accurate is the density nowcast, and the more weight that density nowcast i should receive, such that weights depend on the inverse CRPS. We initialize the weights by setting them to equal weights $w_{\tau,t,i} = 1/M$ if $t \leq T_0 + D_\tau - 1$. For $t = T_0 + D_\tau, \dots, T$, the weights are computed based on an expanding window of historical predictive performance:¹⁷

$$w_{\tau,t,i} = \frac{[\sum_{s=T_0}^{t-D_\tau} CRPS_{\tau,t,i}^{-1}]}{\sum_{j=1}^M [\sum_{s=T_0}^{t-D_\tau} CRPS_{\tau,t,i}^{-1}]} \quad (3.14)$$

4. Conflitti, De Mol, and Giannone (2015) iterative algorithm. Following Hall and Mitchell (2007), the optimal vector of weights $W_{\tau,t}^* = (w_{\tau,t,1}^*, \dots, w_{\tau,t,M}^*)$ minimizes the Kullback-Leibler information criterion (KLIC) divergence, $K\hat{L}IC_{\tau,t} = (1/t) \sum_{s=1}^t [\log g_s(y_s = y_s^o) - \log f_{\tau,s}(y_s = y_s^o, W_{\tau,t})]$, where $f_{\tau,s}(\cdot, W_{\tau,t})$ is the combined density nowcast across the M individual densities at time τ which is a function of weights $W_{\tau,t}$ and $g_t(\cdot)$ is the true but unknown density corresponding to the actual realizations y_t^o . The solution to the optimization problem, subject to constraints $w_{\tau,t,i} \geq 0 \forall i$ and $\sum_{i=1}^M w_{\tau,t,i} = 1$ in each period, is

$$W_{\tau,t}^* = \arg \max_{W_{\tau,t}} \frac{1}{t - T_0 + 1} \sum_{s=T_0}^t [\log f_{\tau,s}(y_s = y_s^o, W_{\tau,s})] \quad (3.15)$$

¹⁶The CDF of the observation is a Heaviside step function, which takes a value of 0 for all values of the density that are less than the actual realization and 1 for all values of the density that are greater than or equal to the realization.

¹⁷We also explored the strategy of computing weights using a 12-month rolling window. The results for density nowcasting accuracy were similar.

Hall and Mitchell (2007) and Amisano and Geweke (2011) note that this optimization can be solved using numerical search algorithms.¹⁸ Conflitti, De Mol, and Giannone (2015) propose an iterative solution to the above optimization problem that is computationally feasible for combining density estimates for large M . Specifically, they break the objective function (equation 3.15) into a set of auxiliary functions that can be easily maximized in an iterative fashion to solve for the optimal weights each period subject to the constraints. The maximization of auxiliary functions is convenient because it is simply a sum of M terms, and when each term is a function of a single weight, there are a total of M weights. The algorithm is initialized with equal weights, $w_{\tau,t,i}^{(0)} = 1/M$. If $t \leq T_0 + D_\tau - 1$, then $w_{\tau,t,i}^* = w_{\tau,t,i}^{(0)}$. For $t = T_0 + D_\tau, \dots, T$, then

$$w_{\tau,t,i}^{(k+1)} = w_{\tau,t,i}^{(k)} \frac{1}{t - D_\tau - T_0 + 1} \sum_{s=T_0}^{t-D_\tau} \frac{f_{\tau,s,i}(y_s = y_s^o)}{\sum_{j=1}^M w_{\tau,s,j}^{(k)} f_{\tau,s,j}(y_s = y_s^o)} \quad (3.16)$$

If $(w_{\tau,t,i}^{(k+1)} - w_{\tau,t,i}^{(k)}) \leq \varepsilon$, then $w_{\tau,t,i}^* = w_{\tau,t,i}^{(k+1)}$. The constraints on the weights are satisfied at every iteration so long as the weights are initially equal. We denote this combination as CMG.

5. Ganics (2017) optimization based on calibration fit. Ganics (2017) proposes an approach to derive optimal weights based on the calibration fit of the model using the probability integral transform (PIT). If the preference is for well-calibrated densities irrespective of the user’s loss function, then intuitively it makes sense to devise a weighting strategy that directly accounts for the calibration fit of each individual (model) density forecast. Ganics (2017) illustrates via Monte Carlo applications and an empirical application that a combination approach relying on the calibration fit to derive the weights not only yields well-calibrated densities but also leads to superior density forecasts in terms of log-score. The paper examines the efficacy of three popular metrics to evaluate the calibration fit—the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling (AD) statistics—and finds that the three metrics perform comparably with a slight advantage for the scheme based on the AD statistic in terms of improved calibration of the combination. Accordingly, we consider the weighting strategy that optimizes the calibration fit based on the AD metric.

The Ganics optimization-based weights are calculated as follows.¹⁹ For a combined density nowcast $f_{\tau,t}(y_t, W_{\tau,t})$ that is a function of a vector of weights $W_{\tau,t}$, compute the spread between the CDF of the perfect uniform distribution and the empirical CDF of the PIT $z_{\tau,t}$ corresponding to the combined density nowcast, $\Omega_{\tau,t}(r, W_{\tau,t}) = 1[z_{\tau,t} \leq r] - r$, where $r \in [0, 1]$ denotes the quantile of the combined density nowcast. Next, compute an average of the spread over the

¹⁸Pauwels and Vasnev (2016) perform an in-depth analysis of this optimization procedure and document a number of practical recommendations.

¹⁹We are grateful to Greg Ganics for sharing computer code.

forecast evaluation sample,

$$\Upsilon_{\tau,t}(r, W_{\tau,t}) = \frac{1}{t - T_0} \sum_{s=T_0}^{t-1} \Omega_{\tau,s}(r, W_{\tau,s}) \quad (3.17)$$

The AD statistic, which is used as the objective function for the combined density, is defined as

$$AD_{\tau,t}(W_{\tau,t}) = \int_{\rho} \frac{\Upsilon_{\tau,t}(r, W_{\tau,t})}{r(1-r)} dr \quad (3.18)$$

for $\rho \in [0, 1]$. Since a lower value for the AD statistic is preferred to a higher value, we again initialize the Ganics weights as equal weights $w_{\tau,t,i}^* = 1/M$ if $t \leq T_0 + D_{\tau} - 1$; otherwise, for $t = T_0 + D_{\tau}, \dots, T$, the Ganics weights solve the minimization problem

$$W_{\tau,t}^* = \arg \min_{W_{\tau,t}} AD_{\tau,t}(W_{\tau,t}) \quad (3.19)$$

3.4 Real-Time Data

Under the assumption that density nowcasts are most informative when they are made in real time, our analysis relies on the real-time data that would have been available to forecasters in the past. Our mixed-frequency model-combination framework can generate nowcasts on a daily basis. To keep the results manageable, we assess nowcasting performance for each month's inflation reading at six representative dates. Table 3.2 provides details about the information flow corresponding to these representative dates.²⁰

We use the following data transformations and data set in our nowcasting exercises. The monthly inflation rate is defined as $\pi_t = 100(P_t/P_{t-1} - 1)$, where P_t is the price index in month t . The 12-month trailing (year-over-year) inflation rate is $\pi_{t,t-12} = 100(P_t/P_{t-12} - 1)$. Quarterly annualized inflation rates are $\pi_T^Q = 100[(P_T^Q/P_{T-1}^Q)^4 - 1]$, with $P_T^Q = 1/3(P_{T,t=1} + P_{T,t=2} + P_{T,t=3})$ the price index for quarter T and $P_{T,t=k}$ is the price level in the k -th month of quarter T . The mixed-frequency models forecast monthly inflation rates, which are used to back out the corresponding price indices to calculate the 12-month trailing inflation rate and quarterly annualized inflation rate.

We nowcast U.S. headline and core inflation rates in both the consumer price index (CPI) and the personal consumption expenditures price index (PCE), which are monthly series. All three model classes use higher-frequency data on gasoline and oil prices. The DMS and DFM models also use data on monthly food and gasoline inflation. To evaluate our models, we use the real-time data set from Knotek and Zaman (2017). The real-time vintages for the monthly PCE price index and core PCE price index begin in June 2000 and come from the Federal Reserve Bank of Saint Louis' Archival Federal Reserve Economic Data (ALFRED). The

²⁰We also produce results for quarterly inflation rates, shown in the appendix. Table B1 shows the seven representative dates that we use for the quarterly nowcasting exercise and the available information.

real-time vintages for CPI inflation, core CPI inflation, and food CPI inflation going back to November 1996 are also from ALFRED. The real-time data for gasoline CPI inflation beginning in January 1999 come from Haver Analytics and ALFRED. Weekly retail gasoline prices (for all grades) going back to the start of 1993 are obtained from the Energy Information Administration (EIA).²¹ Daily Brent crude oil spot prices going back to 1987 are obtained from the Financial Times via Haver Analytics. The real-time CPI gasoline series, which is seasonally adjusted by the Bureau of Labor Statistics, is used to compute real-time seasonal factors that are then applied to retail gasoline prices to adjust them for seasonality.

Additional data are included in the estimation of the DFM. The additional weekly data include the prices of diesel fuel, regular-grade retail gasoline, mid-grade retail gasoline, and premium-grade retail gasoline (from the EIA). The additional daily variables include the food-stuffs price index from the Commodity Research Bureau (CRB), the grains price index from Standard & Poor’s (S&P), the fats and oils price index from CRB, the raw sugar price from the International Sugar Organization, the raw industrials price index from CRB, the agricultural commodities price index from S&P, the textiles and fibers price index from CRB, the industrial metals price index from S&P, steel scrap prices from the Foundation for International Business & Economic Research, the 10-year Treasury note constant maturity yield and the 3-month Treasury bill rate from the Federal Reserve Board, the S&P 500 stock price index as reported in the Wall Street Journal, and the nominal trade-weighted exchange value of the dollar against major currencies from the Federal Reserve Board. These high-frequency data are all downloaded from Haver Analytics and are assumed to be unrevised and hence real-time in nature. Hence, at any point in time τ , we can generate the density nowcasts that would have been available as of that particular day.

To facilitate a horse race between nowcasts produced from our modeling framework and those reported in the Survey of Professional Forecasters (SPF), we also download the historical SPF nowcast estimates from the Federal Reserve Bank of Philadelphia’s real-time database.

Our real-time evaluation period runs from September 2000 through June 2015, to compare the implied point nowcast accuracy of our results to those in Knotek and Zaman (2017). Following Tulip (2009), we treat the third inflation reading for each month or quarter as our measure of “truth.” The third estimate has the advantage that it incorporates more complete source data than earlier estimates, but it usually abstracts from methodological revisions, which would have been difficult to predict in real time.

3.5 Nowcast Evaluation

We use a range of metrics to evaluate our inflation density nowcasts by examining both the absolute accuracy and the relative accuracy of the density nowcasts along with the accuracy of the implied point nowcasts. Absolute accuracy tests for the calibration fit of the density

²¹The EIA publishes weekly gasoline prices every Monday.

estimate. Density forecasts are considered well calibrated or correctly specified when the density forecasts match the distribution of the observations over a large sample. The preference is for densities that are well calibrated. We assess the calibration properties of the density nowcasts using probability integral transforms (PITs), originally proposed by Diebold, Gunther, and Tay (1998), and interval forecasts (i.e., 70% prediction intervals).

The PIT $z_{\tau,t}$ is the CDF of the predictive density nowcast $p_{\tau,t}(y_t)$ from the point in time τ when the nowcast is made for month t evaluated at the actual data realization y_t^o :

$$z_{\tau,t} = \int_{-\infty}^{y_t^o} p_{\tau,t}(u) du, \quad t = T_0, \dots, T \quad (3.20)$$

Intuitively, the PIT indicates in which region (i.e., percentile) of the density nowcast the actual realizations fall (see Gerard and Nimark, 2008). A realization that falls at the middle of the density nowcast would be assigned a PIT value of 0.5, while a realization at the 10th percentile would be assigned a value of 0.1. Density nowcasts that are well calibrated have PITs that are uniformly distributed across observations t ; in a large sample, the actual realizations would be expected to span the entire region of the density nowcast with a probability matching the probability implied by the density nowcast. Therefore, a visual assessment for calibration can be performed by plotting PITs in the form of a histogram along with the uniform $U(0, 1)$ distribution. Correctly specified density estimates resemble rectangles (i.e., flat histograms), while severe departures from uniformity suggest calibration failure. In addition to this visual assessment, we conduct a battery of formal statistical assessments using Pearson’s Chi-squared test (of uniformity and independence), the Berkowitz (2001) test (of normality of the inverse normal of the PITs), the Kolmogorov-Smirnov test (of uniformity), the Anderson-Darling test (of uniformity that specifically puts more weight on deviations between the empirical CDF of the PITs and the CDF of the uniform distribution in the tails), and the Knüppel (2015) test.²² The Knüppel test allows for simultaneous testing for both uniformity and independence of the PITs.

Interval forecasts are another popular metric to gauge the calibration of the density forecasts (e.g., Clark, 2011; Carriero, Clark, and Marcellino, 2015b; Tallman and Zaman, 2020). Accordingly, we compute the empirical 70% prediction intervals (i.e., the coverage rates) of the density nowcasts, which are defined as the difference between the 85th and 15th percentiles of the density nowcasts. We compare the empirical 70% coverage rates with a nominal value of 70% to assess the extent to which the density nowcast estimates are correctly calibrated.

The PITs from two or more competing density nowcasts can all be uniformly distributed and hence appear correctly specified, but we would not be able to distinguish whether one density is of higher quality (i.e., more accurate) than the others based on PITs alone. Relative accuracy involves comparing competing density estimates for their quality based on numerical

²²See Hall and Mitchell (2007) and Rossi and Sekhposyan (2014) for details about the exact implementation of these tests. In implementing the calibration metrics, we benefitted from Matlab code made available on the websites of Barbara Rossi, Tatevik Sekhposyan, and Malte Knüppel.

scores. The density forecast that assigns a higher density (i.e., a higher probability mass) at the actual realization gets a higher numerical score and is considered more accurate. Conditional on obtaining well-calibrated density nowcasts, the density nowcast with the highest numerical score is preferred. Scoring metrics such as log score and CRPS are two widely used scoring rules that allow for the ranking of rival density nowcasts. As discussed earlier, the log score is the logarithm of the probability density function (corresponding to the density nowcast) evaluated at the actual realization. The higher the log score, the more accurate the density nowcast. The CRPS is the difference between the predicted and the realized cumulative distributions. Smaller CRPS values imply more accurate density nowcasts.

We also examine point nowcasts based on the mean of the (combined) density nowcasts. We assess point nowcast accuracy via the standard metric of root mean squared error (RMSE):

$$RMSE_{E_{\tau,t}} = \sqrt{\frac{\sum_{t=T_0}^T (y_t^o - E(p_{\tau,t}(y_t)))^2}{T - T_0 + 1}} \quad (3.21)$$

where $E(p_{\tau,t}(y_t))$ refers to the mean of the density nowcast, $p_{\tau,t}(y_t)$, and $T - T_0 + 1$ is the size of the forecast evaluation sample.

3.6 Empirical Results Using Real-Time Data

We perform a comprehensive investigation of nowcast combination methods with multiple mixed-frequency inflation nowcasting approaches to produce density and point nowcasts of U.S. inflation. Overall, we examine three combination methods; five weighting schemes; four inflation measures (headline and core inflation, CPI and PCE inflation); three inflation rates (12-month trailing inflation, month-over-month inflation, and quarterly inflation); and multiple intraperiod points at which we generate our nowcasts (six distinct points for each monthly inflation reading, and seven distinct points for each quarterly inflation reading). In short, we end up with a massive number of results.

To keep the discussion of the results and the length of the chapter manageable, we focus on a subset of results. Specifically, our discussion focuses entirely on results corresponding to combinations derived from our novel flexible aggregation strategy.²³ The main text also considers inflation nowcast accuracy for 12-month trailing inflation rates.²⁴ We first briefly examine the accuracy of the individual density nowcasts from one model in each of the three mixed-frequency model classes and then the accuracy of the stage 1 combinations. We then compare results for grand combinations across all five weighting schemes, followed by a comparison of

²³The results based on this flexible strategy are equivalent to the ones based on the linear opinion pool, because in our empirical exercises the linear opinion pool always performs better than the log opinion pool and hence the flexible strategy always selects the linear opinion pool over the log opinion pool as the aggregation function. However, we see value in the flexible aggregation strategy because this result need not always hold in all applications.

²⁴See the appendix for results for month-over-month inflation rates and quarterly annualized inflation rates.

the accuracy of the grand combination constructed using the log-score weighting scheme with its three component densities (DMS, MIDAS, and DFM). Finally, we examine the time-varying properties of the grand combination.

3.6.1 Density Nowcasts from Mixed-Frequency Model Classes

We first establish that density nowcasts obtained from a single specification within each of the three model classes are incorrectly calibrated with relatively few exceptions. The exact specifications we consider for the single DMS model, the single MIDAS model, and the single DFM model follow the baseline models in each class in Knotek and Zaman (2017).

Figure 3.1 plots the density nowcast PITs from these three mixed-frequency models for the four inflation measures, based on nowcasts made at two distinct points in time: using the available data through the end of the month preceding the target nowcast month (case 1), and using the available data as of day 22 of the target nowcast month (case 4). We find these points to be broadly representative of our results without showing every case. In general, the arrival of additional data during the target month very marginally improves the calibration of the density nowcasts, as is evident by comparing the proximity to uniformity in the top and bottom panels.

Each of these three individual mixed-frequency models fails one or more of the necessary statistical tests of calibration fit for at least one of the inflation measures of interest. For example, the DFM generates well-calibrated densities for CPI inflation (in cases 1 and 4) and PCE inflation (in case 4) but fails to do so for both core CPI and core PCE, as evidenced by notable departures from uniformity. The DMS model has difficulties producing well-calibrated densities overall, although as more information accumulates for the target month, the calibration of the CPI density nowcasts produced from the DMS model improves significantly (e.g., case 4). The calibration of the MIDAS model's CPI density nowcasts also improves considerably with the arrival of additional information; by contrast, the core PCE inflation density nowcasts are fairly well calibrated both early and late in the targeted month.

The requirement of correctly specified density nowcasts is a necessary condition. So with these individual specifications generally failing this important requirement, we omit a discussion of relative accuracy performance. In summary, the density nowcasts produced from a single model specification have difficulties capturing the true degree of uncertainty around the point nowcasts.

We next consider the stage 1 combinations that produce density nowcasts within each class of mixed-frequency models. Figure 3.2 plots the PITs from the combined density nowcasts from each model class. The plots indicate some improvements in the calibration fit of the combinations compared with the respective individual specifications, especially for the DFM model class. However, in some instances there is evidence of a slight deterioration in the calibration fit. For example, the calibration fit of the combined density nowcasts from the MIDAS class for core PCE inflation in case 4 worsens compared with the single specification as shown in Figure 3.1. This latter result highlights an important aspect of density combination, which is that if

the candidate densities are all individually well calibrated, then their combination via the linear opinion pool may suffer calibration failure due to an increased variance (i.e., the combination overestimates uncertainty). Fortunately, in our application, the density nowcasts for core PCE inflation from the 12 different MIDAS specifications are similar. Therefore, the variance of the resulting combination does not increase enough to cause issues with the calibration fit; i.e., the disagreement component, disagreement about the mean, is small.

In addition to some improvements in calibration fit, the relative accuracy of the stage 1 combinations in terms of log scores is at least as accurate as, and often more accurate than, the respective individual specifications (see Figure B3 in the appendix). Taken together, these results suggest that there are gains from combining models within a model class, but there remain deficiencies in the calibration fit of these within-model-class combinations.

3.6.2 Comparison across Grand Combinations

We next explore whether the calibration fit of the density nowcasts can be improved further by combining the stage 1 combination density nowcasts from each of the three mixed-frequency model classes into a stage 2 “grand” combination.²⁵ Figure 3.3 plots the PITs across various grand combinations based on the five weighting schemes listed above. Immediately, the figure demonstrates that there are gains from combining densities across mixed-frequency model classes as evidenced by PITs that are now closer to the uniform distribution. More precisely, combining densities via our two-stage procedure fixes the defective stage 1 combination density nowcasts. A close inspection of the PIT histograms across all four inflation measures, and in both cases 1 and 4, reveals that all five grand combination density estimates are better calibrated compared with the density estimates from either the individual specifications or those formed by combining specifications only within a model class.

Tables 3.3 and 3.4 report the formal statistical assessment of the calibration fit. Among the five combinations we consider, we find that the combination based on the weighting scheme proposed by Ganics generates the best-calibrated densities, with the smallest number of rejections of the null hypothesis of correct calibration.²⁶ The density estimates based on the log score and CMG weighting schemes are the next best combinations, with somewhat higher numbers of rejections of the null. The ability of the Ganics weighting scheme to produce the best-calibrated densities makes intuitive sense, as his approach is based on direct optimization of the calibration fit of the candidate densities.

Beyond calibration fit, we are also interested in relative accuracy. Figure 3.4 assesses relative accuracy via the log score (panel a) and CRPS (panel b) metrics. In general, the relative accuracy of the grand combinations improves as additional information arrives over the course

²⁵As a reminder, the same weighting scheme is used to form both the stage 1 and stage 2 combinations. Figure B4 in the appendix plots the nowcasts coming from the grand combination using real-time data at two representative dates (case 1 and case 4) for each month along with the actual outcomes.

²⁶We omit the results for the equal weights scheme and the CRPS weighting scheme to economize on space in the table; in general, these weighting schemes were inferior to those shown, which is visible in Figure 3.3.

of the target nowcast month, as is evident by steadily increasing average log scores and declining CRPS values. The figure shows that the grand combination based on the Ganics weighting scheme generates inferior density nowcasts for CPI inflation and PCE inflation compared with both the log score and CMG weighting schemes. In the case of core CPI inflation, all weighting scheme combinations perform comparably in terms of relative accuracies. For core PCE inflation, the log score and Ganics weighting schemes are a touch worse than the other schemes when using log score as the relative accuracy metric, but the log score weights are just as accurate as other weighting schemes when using CRPS to judge relative accuracy. Given that the density nowcast estimates based on log score and CMG weighting schemes generate superior relative density nowcast accuracy and satisfactory calibration fit across inflation measures, we generally favor these schemes over the other weighting schemes.

The patterns observed for the relative accuracy scores of the density nowcasts echo the point nowcast accuracy results reported in Knotek and Zaman (2017). Knotek and Zaman (2017) document that a single version of the DMS model was substantially more accurate than competing MIDAS or DFM models in nowcasting CPI inflation and PCE inflation, while all three models were competitive in nowcasting core CPI inflation and core PCE inflation. In terms of the relative accuracy of the density nowcasts, combination schemes such as log score that put more weight on the DMS model class generate more accurate density nowcasts for CPI inflation and PCE inflation, even though the calibration fit of the density nowcasts from the DMS model class is inferior to both the DFM and MIDAS model classes for these inflation measures. However, the mean of the density nowcasts—i.e., the point nowcasts—from the DMS model class is substantially more accurate than the means coming from either the DFM or the MIDAS model classes. When evaluated using the log score metric, which is considered a broader measure of density accuracy (e.g., Clark, 2011), the DMS model class’s more accurate density nowcast mean more than offsets its slightly poorer calibration fit, resulting in a higher log score.

In contrast, the Ganics weighting scheme, which focuses on calibration fit, assigns large weights to the DFM and MIDAS model classes for nowcasting CPI inflation and PCE inflation (see Figure B5 in the appendix for an example), because these two model classes produce better-calibrated densities than the DMS model class. The equal weights scheme and the CRPS-based weighting scheme also tend to put more weight on the DFM and MIDAS model classes. Because these two model classes produce substantially inferior point nowcasts compared with the DMS model class, the additional weight assigned to them results in lower log scores and higher CRPS values.

In addition to absolute and relative accuracy in a density sense, Figure 3.5 plots the RMSE of the implied point nowcasts corresponding to the grand combinations. The results for point nowcast accuracy echo the results for density nowcast accuracy shown in Figure 3.4. As information accumulates over the course of the month and we move from case 1 to case 6, the accuracy of the point nowcasts steadily improves and the RMSEs steadily decline. In the case

of CPI inflation and PCE inflation, the combinations using log score weights produce lower RMSEs than the other weighting schemes. For core CPI inflation and core PCE inflation, the differences in RMSEs across weighting schemes are very small and all combinations generally perform comparably.

3.6.3 Comparing the Grand Combination with Its Underlying Component Densities

We noted above that the grand combination helps improve the calibration of the density nowcasts, which is an important objective. We also compare the accuracy of the grand combination to its three component densities—DMS, MIDAS, and DFM—when both the stage 1 and stage 2 combinations are made using the log score weighting scheme to assess the extent to which relative accuracy and point accuracy gains are coming from combining densities across mixed-frequency model classes.

Figure 3.6 plots the density nowcast relative accuracy comparison using the log score and CRPS metrics. These figures indicate that the grand combination’s nowcasts are generally among the best performing, especially when assessed using the CRPS metric. However, when evaluated using the log score metric, there are some instances for CPI inflation and PCE inflation where the DMS combination is significantly more accurate than the grand combination. Indeed, the DMS combination alone is often quite competitive with the grand combination. Only in the case of core PCE inflation with the log score metric does the MIDAS combination outperform the DMS combination.

Figure 3.7 compares the point nowcast accuracy of the grand combination with the DMS combination and the single DMS specification from Knotek and Zaman (2017).²⁷ The point nowcast accuracy of the grand combination is similar to the single DMS specification for CPI inflation, PCE inflation, and core CPI inflation. In the case of core PCE inflation, the grand combination is more accurate than the single DMS specification, with notably lower RMSEs for cases 1 through 4; for cases 3 and 4, the accuracy gains are statistically significant at the 10% significance level.²⁸ We view these results as desirable because it satisfies an important objective of our model-combination approach, which is that we want the point nowcasting accuracy of the combined density to be at least as good as the single DMS specification from Knotek and Zaman (2017).

Given that the point accuracy of the DMS combination is similar to the accuracy of the grand combination for all four inflation measures, one could focus only on combinations from the DMS model class. However, we view combining density nowcasts across the three mixed-frequency model classes as more desirable for at least two reasons. First, given the preference

²⁷We omit the MIDAS and DFM combinations because the DMS combination is substantially more accurate for CPI inflation and PCE inflation, and all combinations are competitive for core PCE inflation and core CPI inflation. The grand combination and DMS combination use the log score weighting scheme.

²⁸Statistical significance is based on the Diebold-Mariano and West test (with the truncation lag parameter $h-1$ for the HAC variance estimator), and two-sided standard normal critical values.

for well-calibrated densities, the grand combination comes closer to being correctly specified than the DMS combinations. Second, while the DMS model class performed well over our evaluation sample, there is no guarantee that it will continue to do so going forward. The grand combination based on a richer set of models provides better insurance in the face of future uncertainties about model specifications and will be more robust to structural instabilities.²⁹

3.6.4 Time-Varying Properties of the Grand Combination: Weights, Uncertainty, Skewness, and Kurtosis

The blend of adaptive, possibly time-varying weights and the use of a flexible aggregation strategy could yield time-varying estimates of the variance (uncertainty), skewness (asymmetry), and kurtosis in the inflation density nowcasts. Figures 3.8 and 3.9 plot the properties of the density nowcasts for the four inflation measures for cases 1 and 4 during the target nowcast month: the evolution of the weights applied to the stage 1 combinations in making the stage 2 grand combination (top row); the evolution of the uncertainty (i.e., the volatility) around the point nowcasts, measured as the width of the 70% prediction intervals of the density nowcasts (second row); estimates of skewness (third row); and estimates of kurtosis (fourth row).³⁰

We note the following items from the figures. First, in the case of CPI inflation and PCE inflation, the DMS model class quickly dominates the DFM and MIDAS model classes, receiving much or nearly all of the weight for most of our evaluation sample. However, it is worth pointing out that even if a single model class receives a nearly 100% weight in the stage 2 combination, that model class is nevertheless a stage 1 combination of many individual model specifications. These stage 1 combinations almost always include more than one model specification (see Figures B10, B11, and B12 in the appendix). In the case of core CPI inflation and core PCE inflation, there is considerably more variation regarding which model class receives the most weight, although the MIDAS model class tended to be the best performing late in the sample and so received the most weight.

Second, we highlight the fast model-switching behavior of the weighting scheme based on the log-score. When the differences in the density accuracy among the candidate densities are in the moderate to large range, then the log score metric discriminates among densities rather sharply by heavily penalizing the poor performers. This sharp distinction implies that the density nowcast assessed as the most accurate gets a very high score, which translates into a substantially higher average score and, in turn, significantly higher weight in the combination, possibly resulting in a fast-switching pattern.³¹ Recent research generally views this feature of fast model-switching as desirable, and so a framework that allows for it is viewed favorably (see

²⁹Figures B6, B7, B8, and B9 in the appendix report similar results for month-over-month inflation and quarterly annualized inflation rates.

³⁰Skewness=0 and kurtosis=3 for a variable that is normally distributed, so departures from these values suggest evidence for asymmetric distributions if skewness is different from zero or fat tails if kurtosis is greater than 3.

³¹Aastveit et al. (2014) document a similar characteristic for the log score weighting scheme in their real GDP nowcasting application.

Beckmann et al., 2020).

Third, as information accumulates over the course of the target nowcast month, the precision of the density nowcasts improves, and there is a shift lower in the uncertainty estimates in case 4 compared with case 1.³²

Fourth, there are visible movements in the uncertainty estimates, more so in the case of headline inflation than core inflation. It is also evident that the profiles of the uncertainty estimates differ across inflation measures and across monthly cases, reflecting different nowcast origins within a month for a given inflation measure. Interestingly, the patterns seen in the case of quarterly inflation provide stronger evidence of shifts in uncertainty over time (see Figure B13 in the appendix). Moreover, the broader movements in the uncertainty estimates implied by our density nowcasts for quarterly inflation are similar to uncertainty estimates reported elsewhere in the literature using stochastic volatility models (e.g., Knotek, Zaman, and Clark, 2015).

Fifth, there is evidence of both time-varying skewness and kurtosis in the density nowcasts. There have been extended stretches in which skewness differed from zero and kurtosis differed from three, consistent with departures from Gaussianity, along with occasional spikes in both measures.³³ Overall, the density nowcast estimates generated from our two-stage combination process featuring a flexible aggregation strategy with the log score weighting scheme can adapt in a time-varying manner to accommodate non-Gaussian features such as asymmetry and/or heavy tails and are doing an adequate job of capturing uncertainty around future inflation outcomes.³⁴

3.7 Comparison with the Survey of Professional Forecasters

Given the ability of their respondents to use a variety of high-frequency data sources in a flexible fashion, nowcasts coming from surveys of professional forecasters are a difficult benchmark to beat. We test the nowcasting performance of our grand combination against the inflation nowcasts provided by the Survey of Professional Forecasters (SPF), in terms of both point nowcasting performance and density nowcasting performance. In doing so, we match our model's real-time information set to the survey dates from the SPF each quarter.

For point nowcasting, we compare the median SPF response to the mean of the combined density nowcast coming from the grand combination using the log score weighting scheme and the flexible aggregation strategy. For density nowcasting, we compare our grand combination with *estimated* survey density nowcasts formed using a normal distribution, whose mean is set equal to the median SPF point nowcast and whose variance is set to match the variance of the

³²See the narrowing of the prediction intervals in case 4 compared with case 1 in Figure B4 in the appendix.

³³Figure B14 in the appendix illustrates the stage 2 grand combination as of case 1 for nowcasting the target month of January 2001; the resulting CPI inflation and PCE inflation grand combination densities are noticeably fat tailed and asymmetric compared with the densities for core CPI inflation and core PCE inflation.

³⁴Figures B15, B16, and B17 in the appendix plot the weights and higher-order moments when using the CMG, Ganics, and CRPS weighting schemes, respectively.

past historical errors of the SPF point nowcasts over a short rolling window.

Estimates of survey density nowcasts based on historical errors have been shown to be a good benchmark, especially for inflation.³⁵ The Federal Open Market Committee uses historical forecast errors to provide an estimate of the uncertainty surrounding the outlook in the Summary of Economic Projections (see Reifschneider and Tulip, 2019). While the SPF does provide some density forecasts by combining individual respondents' density forecasts, we favor the historical errors approach because Clements (2018) shows that the survey projection's second moments are inferior to simple statistical models. In addition, the SPF only reports fixed-event density forecasts for core PCE inflation and core CPI inflation, which limits their comparability to our results to the fourth quarter of each year.

Table 3.5 reports the results from the out-of-sample nowcasting horse race between our grand combination using real-time data and the SPF for point and density nowcasts. The evaluation period runs from 2000Q4 through 2015Q2 for CPI inflation, and 2007Q1 through 2015Q2 for core CPI inflation, PCE inflation, and core PCE inflation.

The point nowcasts implied by the grand combination are substantially more accurate (i.e., have lower RMSEs) than the SPF for both CPI inflation and PCE inflation, and the gains are statistically significant. For these two inflation measures, the density nowcasts from the grand combination are substantially more accurate than the simple SPF-based benchmark density nowcasts, as indicated by significantly higher log scores.

For core CPI inflation and core PCE inflation, both the point accuracy and the density accuracy of the grand combination are competitive with the SPF. As noted above, there is only limited evidence of skewness and kurtosis in the predictive distributions for core inflation, suggesting that flexible density estimates are not far from the normality assumption embedded in the estimated SPF density. In the case of core inflation, we see our framework's ability to adapt in a dynamic fashion to produce an approximately normal distribution as a testament to the benefits of our flexible approach.

3.8 Conclusion

We develop a flexible framework based on model combinations to produce density nowcasts for U.S. inflation. By combining individual density nowcasts from three classes of parsimonious mixed-frequency models, this framework generates nowcasts at a trading-day frequency and updates as information accumulates over the course of a month or a quarter. We propose a novel flexible aggregation strategy to combine the density nowcasts both within and across model classes. We complement this flexible aggregation strategy with an examination of a variety

³⁵Krüger, Clark, and Ravazzolo (2017) and Tallman and Zaman (2020) document competitive nowcasting performance, including the good calibration fit of the density nowcasts of inflation constructed through this simple procedure. The procedure's use of a short rolling window in computing the variance of the past historical errors is a simple and convenient way to incorporate the changing variance of the density estimates instead of explicitly modeling stochastic volatility, a point also emphasized by Ganics, Rossi, and Sekhposyan (2019)

of dynamic model averaging approaches, where the weights used to combine the nowcasts can be updated based on learning from past performance. An important feature of this proposed framework is its ability to accommodate non-Gaussian and time-varying properties of variance, skewness, and kurtosis in the density nowcast estimates. These dynamic features are essential in improving the accuracy of density nowcasts for headline inflation.

Our flexible framework allows us to incorporate a range of recent density combination methods proposed in the literature in a comprehensive empirical examination. Overall, using high-frequency, real-time data over the period 2000-2015, we show that the grand combination from our approach can generate highly accurate density nowcasts, but the combination method matters for the accuracy of the combined density nowcasts. Density combinations based on our novel flexible aggregation strategy using the log score weighting scheme, which relies on past predictive performance, or the “optimal” weighting scheme proposed by Conflitti, De Mol, and Giannone (2015) are among the best performing in terms of relative accuracy and are well calibrated. The Ganics (2017) weighting scheme produces the best calibrated densities for headline inflation, but its relative accuracy is inferior to the log score and the Conflitti, De Mol, and Giannone (2015) weighting schemes. In the case of core inflation, all combination methods perform comparably.

In a horse race with the Survey of Professional Forecasters, the grand combination’s density nowcasts provide superior point and density nowcasts for CPI inflation and PCE inflation. For core CPI inflation and core PCE inflation, our grand combination’s nowcasting performance is competitive with the SPF. The ability of our proposed framework to generate highly accurate point and density nowcasts of inflation is a useful outcome for practitioners.

Our empirical findings should serve as a guide to practitioners about the combination methods that may or may not work for nowcasting U.S. inflation. Our study provides further evidence that, when it comes to density combinations, there is no single best procedure; rather, it is important to examine combination methods on a case-by-case basis.

Table 3.1: Model Space: Mixed-Frequency Model Classes and Specifications

Model Class	Modeling Options	Number of Models
DMS	Autoregressive (AR) Lags = [1, 2] Estimation window core inflation = [24, 36] months Estimation window headline inflation = [24, 36, 84] months Estimation window oil-gasoline error correction = [60, 72, 84] months Number of years to use for computing seasonal factors = [3, 5, 7] years	108
MIDAS	Estimation window = [5-year, 7-year, 10-year rolling, expanding] Polynomial option = [Beta] High-frequency data = [Daily only, weekly only, both daily and weekly]	12
DFM	Number of factors = [1] Number of lags = [1, 2, 3, 4, 5, 6] Estimation window = [Expanding, 5-year rolling]	12
Total		132

Notes: DMS is deterministic model switching; MIDAS is mixed data sampling; and DFM is dynamic factor model. See the text and the appendix for details.

Table 3.2: Representative Dates for Monthly Nowcasting Performance

Case	Date	Information Set $s(\tau)$ (Example: Nowcasting target month is January)	Target Month Horizon (t+h)
1	Last day of the previous month	December 31: Have CPI and PCE through November; high-frequency information through December 31	CPI: h=2 PCE: h=2
2	Day 8 of the target month	January 8: Have CPI and PCE through November; high-frequency information through the end of the 1st week of Jan., which includes weekly retail gasoline reading for the 1st week of Jan.	CPI: h=2 PCE: h=2
3	Day 15 of the target month	January 15: Receive CPI for December and have PCE through November; high-frequency information through end of 2nd week of Jan., which includes two weekly retail gasoline readings for Jan.	CPI: h=1 PCE: h=2
4	Day 22 of the target month	January 22: Have CPI for December and PCE through November; high-frequency information through end of 3rd week of Jan., which includes three weekly retail gasoline readings for Jan.	CPI: h=1 PCE: h=2
5	Last day of the target month	January 31: Have CPI for December and receive PCE for December; high-frequency information for all of January, which includes all four weekly retail gasoline readings for Jan.	CPI: h=1 PCE: h=1
6	Day 15 of the following month	February 15: Receive CPI for January and have PCE through December; high-frequency information for all of Jan.	CPI: h=n/a PCE: h=1

Table 3.3: Calibration Diagnostics

Nowcasting Case	Berk.	Chi-Sq.	AD	KS	KL	70% Cov Rate.
CPI: Grand Combination Based on Log Score						
Case 1	0.80	0.05	0.02	0.35	0.26	60.67
Case 2	0.23	0.98	0.42	0.91	0.74	72.47
Case 3	0.50	0.38	0.74	0.81	0.79	69.66
Case 4	0.00	0.33	0.36	0.50	0.38	77.53
Case 5	0.01	0.05	0.44	0.18	0.09	76.97
CPI: Grand Combination Based on CMG						
Case 1	0.14	0.92	0.63	0.95	0.99	70.22
Case 2	0.00	0.28	0.36	0.20	0.50	74.16
Case 3	0.05	0.28	0.37	0.52	0.56	76.40
Case 4	0.01	0.14	0.37	0.17	0.37	77.53
Case 5	0.00	0.03	0.07	0.13	0.07	80.90
CPI: Grand Combination Based on Ganics						
Case 1	0.10	0.55	0.37	0.76	0.22	71.35
Case 2	0.01	0.75	0.39	0.56	0.44	71.35
Case 3	0.07	0.02	0.35	0.38	0.29	77.53
Case 4	0.11	0.56	0.36	0.48	0.43	76.40
Case 5	0.04	0.23	0.39	0.17	0.21	75.84
Core CPI: Grand Combination Based on Log Score						
Case 1	0.32	0.02	0.42	0.30	0.20	65.73
Case 2	0.17	0.22	0.43	0.22	0.30	61.80
Case 3	0.52	0.41	0.36	0.46	0.43	64.61
Case 4	0.39	0.38	0.35	0.43	0.46	65.17
Case 5	0.33	0.28	0.37	0.37	0.47	65.73
Core CPI: Grand Combination Based on CMG						
Case 1	0.02	0.70	0.42	0.24	0.53	67.42
Case 2	0.01	0.46	0.09	0.16	0.42	67.98
Case 3	0.41	0.72	0.36	0.44	0.52	65.17
Case 4	0.38	0.43	0.35	0.60	0.53	66.29
Case 5	0.35	0.60	0.35	0.58	0.53	66.85
Core CPI: Grand Combination Based on Ganics						
Case 1	0.08	0.16	0.06	0.10	0.35	65.73
Case 2	0.17	0.61	0.08	0.19	0.36	65.73
Case 3	0.44	0.05	0.36	0.26	0.47	63.48
Case 4	0.25	0.23	0.42	0.14	0.39	63.48
Case 5	0.37	0.14	0.39	0.30	0.28	65.17

Notes: Entries except for those in the final column are p-values. “Berk” is the Berkowitz test. “Chi-Sq” is the Pearson Chi-Squared test. “AD” is the Anderson-Darling test. “KS” is the Kolmogorov-Smirnov test. “KL” is the Knüppel test. Entries in bold indicate rejection of the null hypothesis of correctly calibrated density nowcasts at a 5% significance level. “70% Cov. Rate” shows the 70% empirical coverage rates. The evaluation sample: September 2000 through June 2015;

excludes September 2001 and October 2001 for PCE and core PCE inflation calculations.

Table 3.4: Calibration Diagnostics (continued)

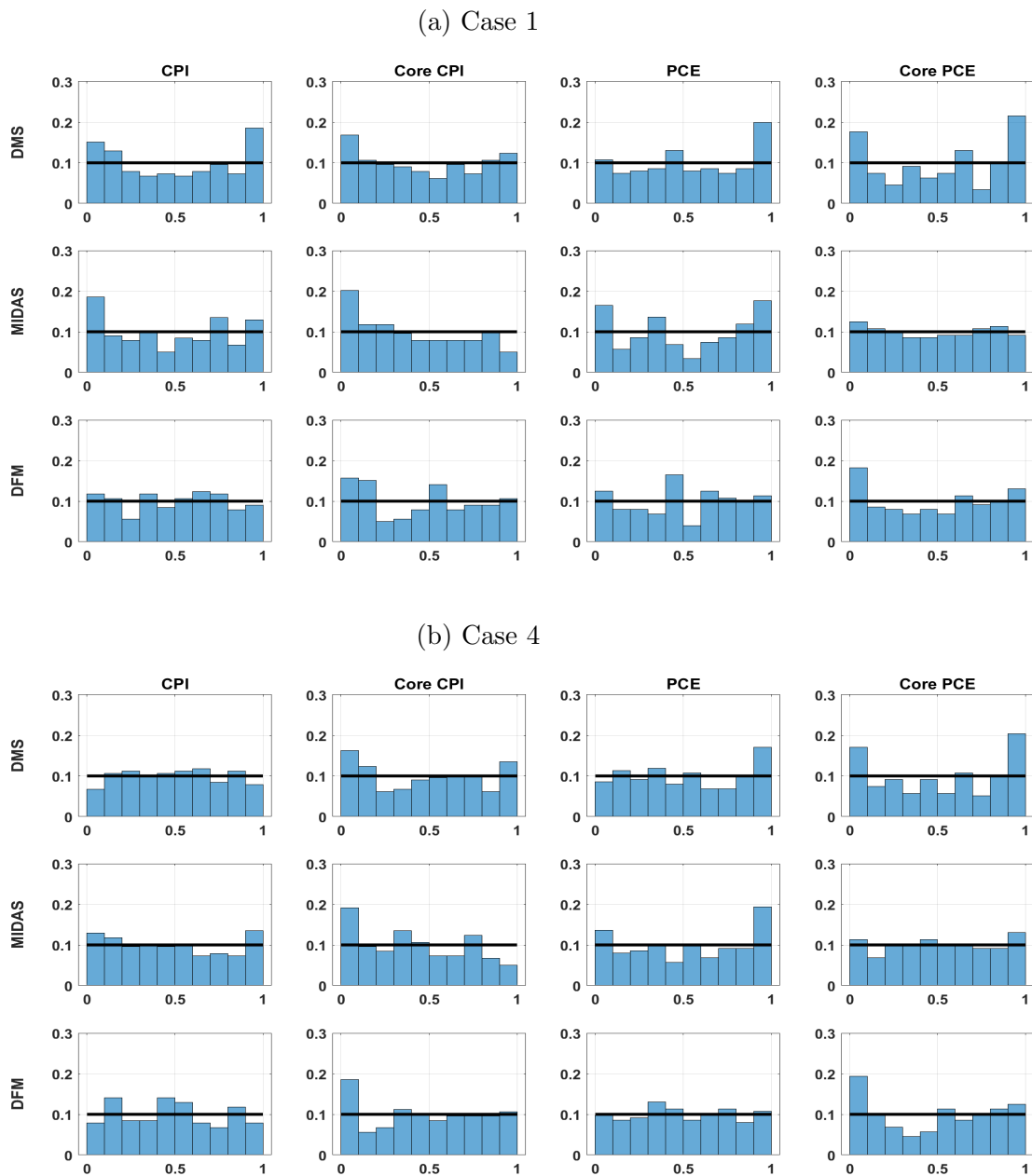
Nowcasting Case	Berk.	Chi-Sq.	AD	KS	KL	70% Cov Rate.
PCE: Grand Combination Based on Log Score						
Case 1	0.53	0.02	0.01	0.06	0.14	59.66
Case 2	0.05	0.30	0.01	0.01	0.34	65.91
Case 3	0.20	0.46	0.37	0.24	0.37	68.75
Case 4	0.09	0.11	0.39	0.45	0.43	72.73
Case 5	0.05	0.18	0.03	0.11	0.51	69.89
Case 6	0.49	0.19	0.01	0.14	0.13	63.07
PCE: Grand Combination Based on CMG						
Case 1	0.23	0.51	0.40	0.31	0.73	67.61
Case 2	0.02	0.48	0.07	0.03	0.57	69.89
Case 3	0.07	0.01	0.36	0.26	0.37	71.02
Case 4	0.03	0.07	0.37	0.37	0.28	71.02
Case 5	0.00	0.11	0.07	0.09	0.51	70.45
Case 6	0.04	0.82	0.39	0.27	0.72	65.34
PCE: Grand Combination Based on Ganics						
Case 1	0.39	0.56	0.45	0.54	0.94	68.18
Case 2	0.05	0.74	0.42	0.21	0.68	69.89
Case 3	0.93	0.33	0.50	0.67	0.54	67.05
Case 4	0.91	0.76	0.57	0.47	0.50	67.05
Case 5	0.13	0.47	0.41	0.52	0.29	73.86
Case 6	0.07	0.63	0.36	0.68	0.71	65.34
Core PCE: Grand Combination Based on Log Score						
Case 1	0.26	0.01	0.36	0.28	0.44	63.07
Case 2	0.43	0.78	0.36	0.38	0.61	64.20
Case 3	0.17	0.06	0.01	0.12	0.12	59.09
Case 4	0.06	0.01	0.00	0.25	0.16	60.80
Case 5	0.42	0.49	0.38	0.67	0.55	66.48
Case 6	0.77	0.05	0.04	0.28	0.26	63.07
Core PCE: Grand Combination Based on CMG						
Case 1	0.00	0.70	0.63	0.90	0.76	67.61
Case 2	0.00	0.94	0.55	0.87	0.67	67.61
Case 3	0.04	0.40	0.67	0.90	0.89	69.32
Case 4	0.06	0.25	0.64	0.82	0.90	68.18
Case 5	0.17	0.70	0.36	0.21	0.67	66.48
Case 6	0.21	0.94	0.40	0.83	0.75	67.05
Core PCE: Grand Combination Based on Ganics						
Case 1	0.00	0.64	0.44	0.83	0.57	67.05
Case 2	0.00	0.68	0.51	0.79	0.79	67.05
Case 3	0.52	0.70	0.35	0.71	0.63	65.34
Case 4	0.46	0.88	0.36	0.71	0.49	64.77
Case 5	0.97	0.88	0.37	0.80	0.73	69.32
Case 6	0.54	0.51	0.36	0.83	0.70	64.20

Table 3.5: Nowcasting Comparison with the Survey of Professional Forecasters

	CPI	Core CPI	PCE	Core PCE
Point Nowcast Comparison				
Grand combination RMSE	1.040	0.569	0.793	0.525
SPF (median) RMSE	1.429	0.577	1.089	0.504
Ratio, avg. SPF MSE/Grand MSE	1.888	1.025	1.883	0.922
GW <i>p-values</i>	0.010	0.881	0.002	0.617
Density Nowcast Comparison				
Grand combination log score (Grand LS)	-1.370	-0.811	-1.064	-0.735
SPF log score (SPF LS)	-1.913	-1.519	-1.572	-0.780
Relative, SPF LS – Grand LS	-0.543	-0.709	-0.508	-0.045
DM type test <i>p-values</i>	0.000	0.185	0.000	0.525

Notes: The grand combination uses real-time data available through the SPF survey date for each quarter. The SPF density nowcasts are based on historical forecast errors; see the text for details. The CPI exercise uses real-time data from 2000Q4 through 2015Q2. The core CPI, PCE, and core PCE exercises use real-time data from 2007Q1 (the first available SPF estimate) through 2015Q2. The DM type test reports the results of a test for equal predictive accuracy based on testing whether the constant term in the regression of the differences in the log score on the constant is statistically different from zero.

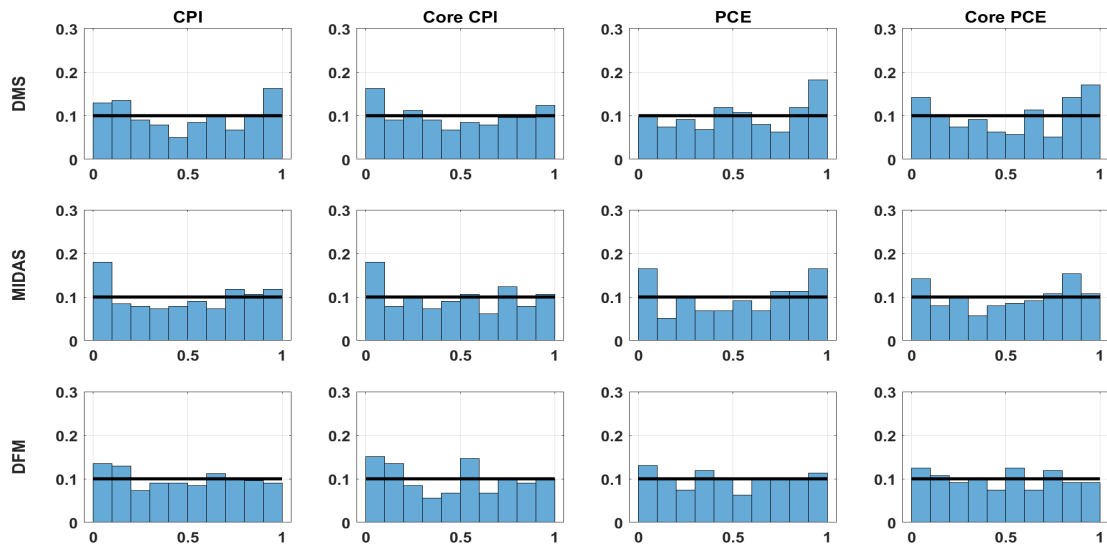
Figure 3.1: Comparison of PITs across Single Specifications of Mixed-Frequency Model Classes



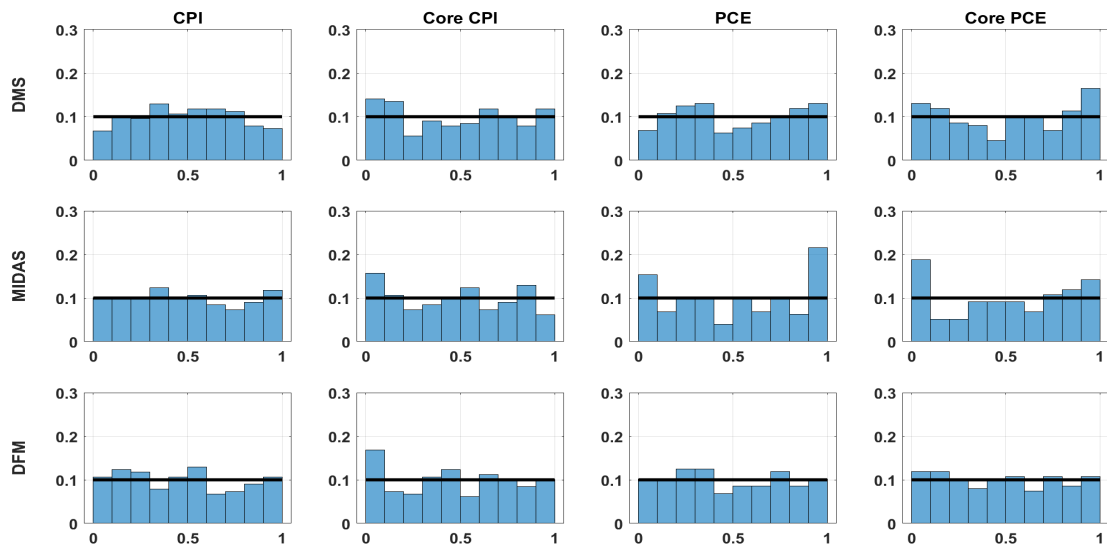
Notes: The figure plots histograms of the empirical distribution of the PITs for single specifications of the DMS, MIDAS, and DFM model classes (blue bars) and the uniform $U(0,1)$ distribution (black lines), generated at either the last day of the month preceding the target nowcast month (case 1) or day 22 of the target nowcast month (case 4). The x-axis shows the decile bins and the y-axis shows the percentage of observations falling within each decile bin. The nowcast evaluation sample spans September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure 3.2: Comparison of PITs across Stage 1 Combinations within Model Classes

(a) Case 1



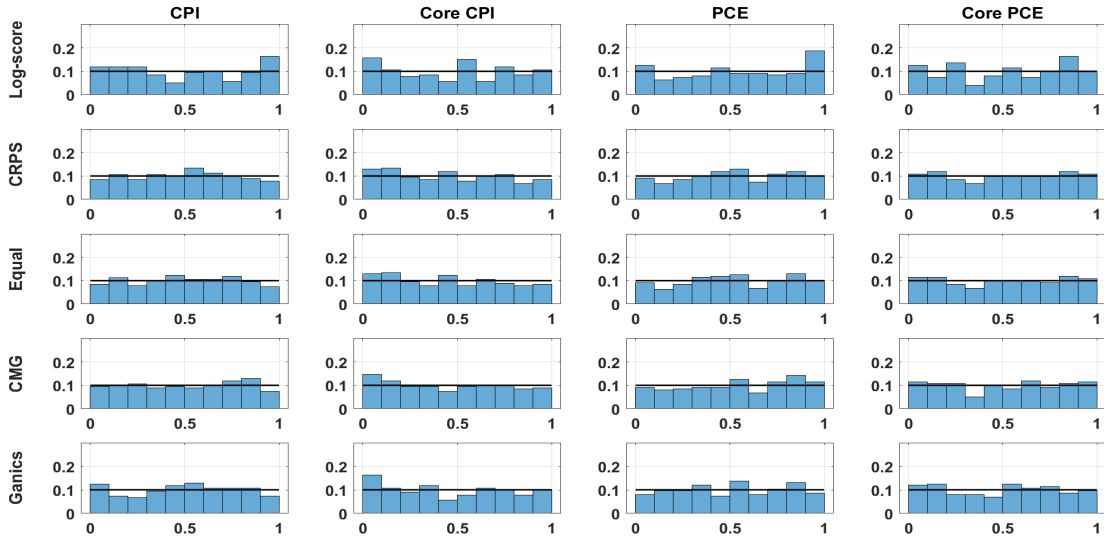
(b) Case 4



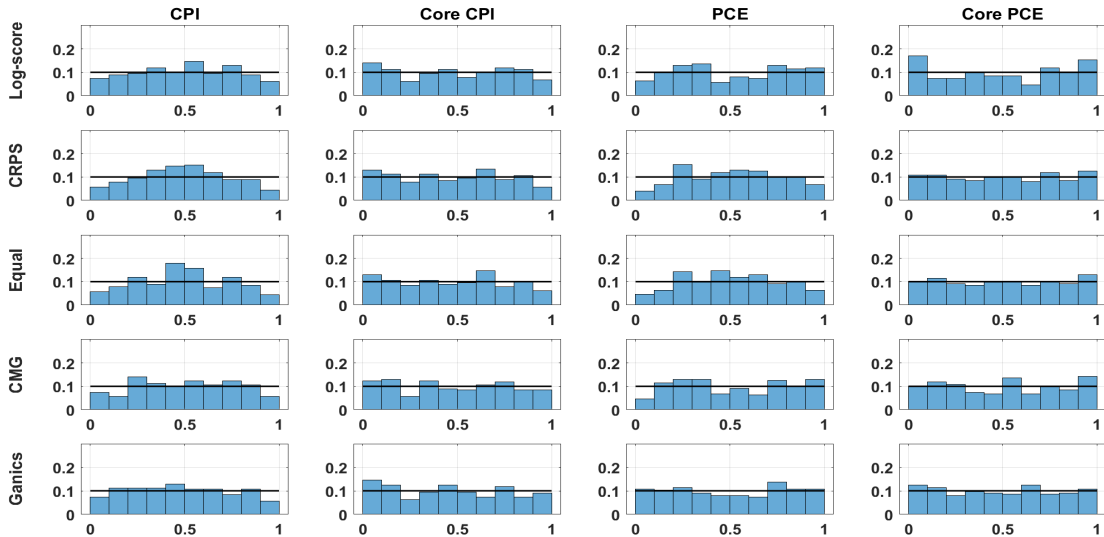
Notes: The figure plots histograms of the empirical distribution of the PITs for stage 1 combinations within the DMS, MIDAS, and DFM model classes (blue bars) and the uniform (0,1) distribution (black lines), generated at either the last day of the month preceding the target nowcast month (case 1) or day 22 of the target nowcast month (case 4). The x-axis shows the decile bins and the y-axis shows the percentage of observations falling within each decile bin. The nowcast evaluation sample spans September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure 3.3: Comparison of PITs across Grand Combinations

(a) Case 1

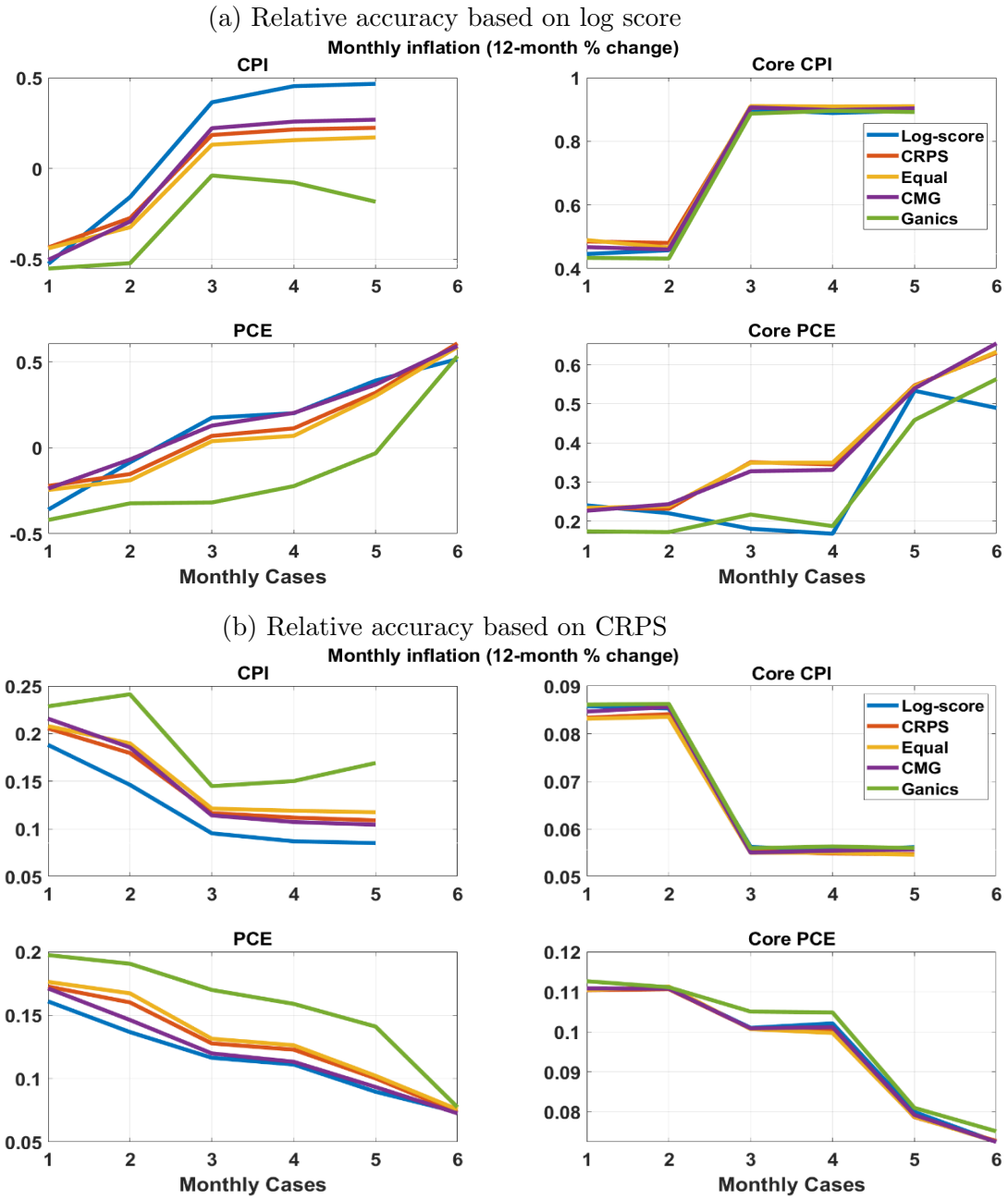


(b) Case 4



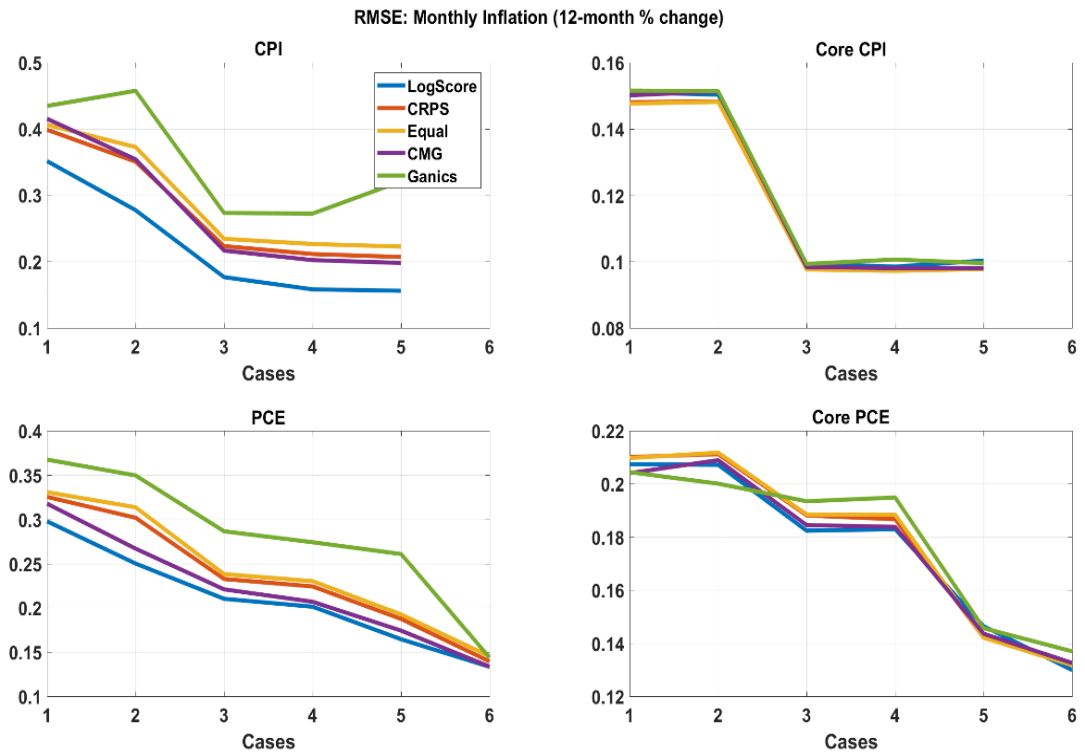
Notes: The figure plots histograms of the empirical distribution of the PITs for stage 2 combinations across the DMS, MIDAS, and DFM model classes (blue bars) and the uniform (0,1) distribution (black lines), generated at either the last day of the month preceding the target nowcast month (case 1) or day 22 of the target nowcast month (case 4). The x-axis shows the decile bins and the y-axis shows the percentage of observations falling within each decile bin. The nowcast evaluation sample spans September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure 3.4: Density Performance Comparisons across Grand Combinations



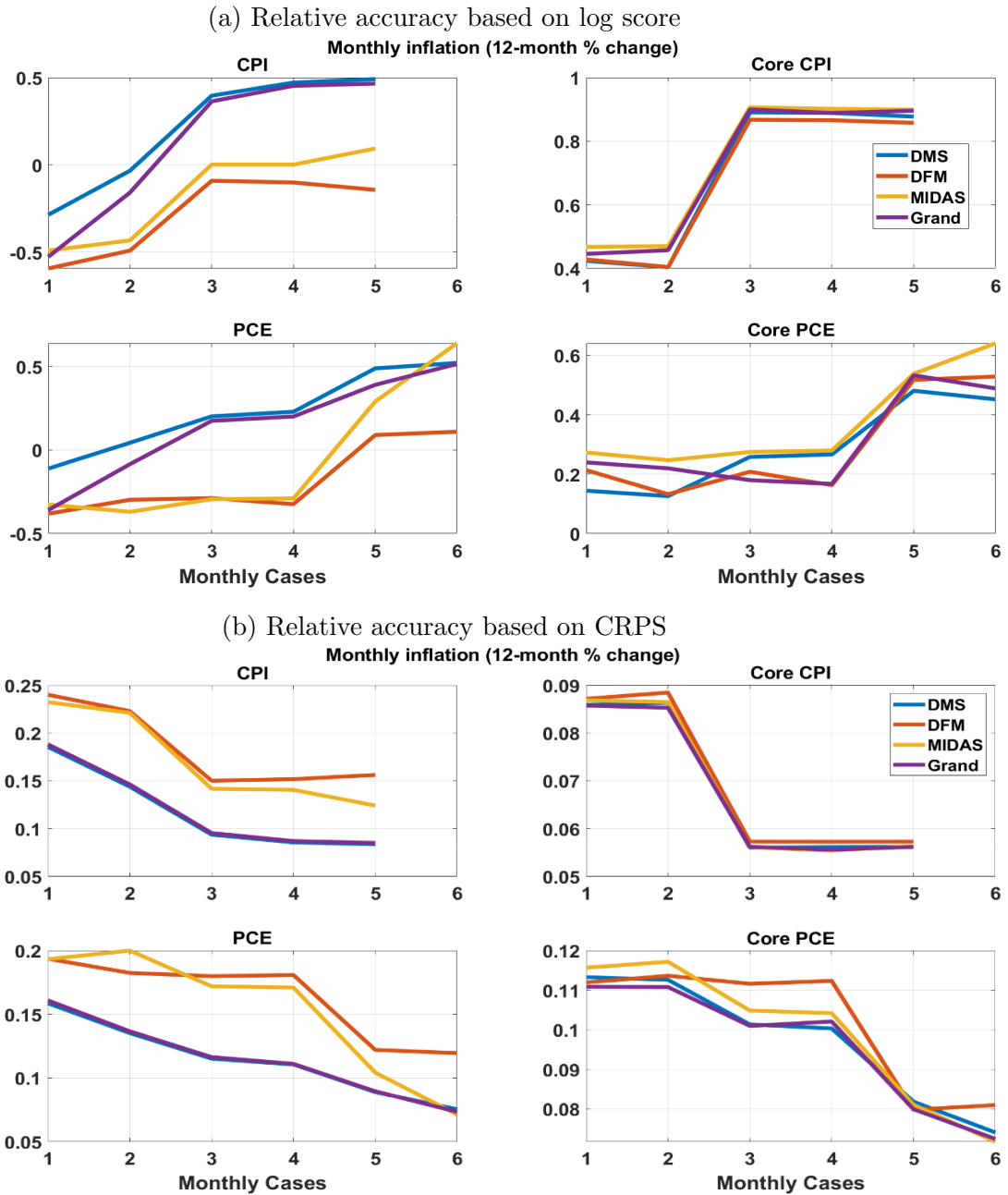
Notes: The top panel plots the average log score and the bottom panel plots the average CRPS for grand combinations based on log score, CRPS, equal, CMG, and Ganics weighting schemes. The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure 3.5: Point Nowcasting Performance across Grand Combinations



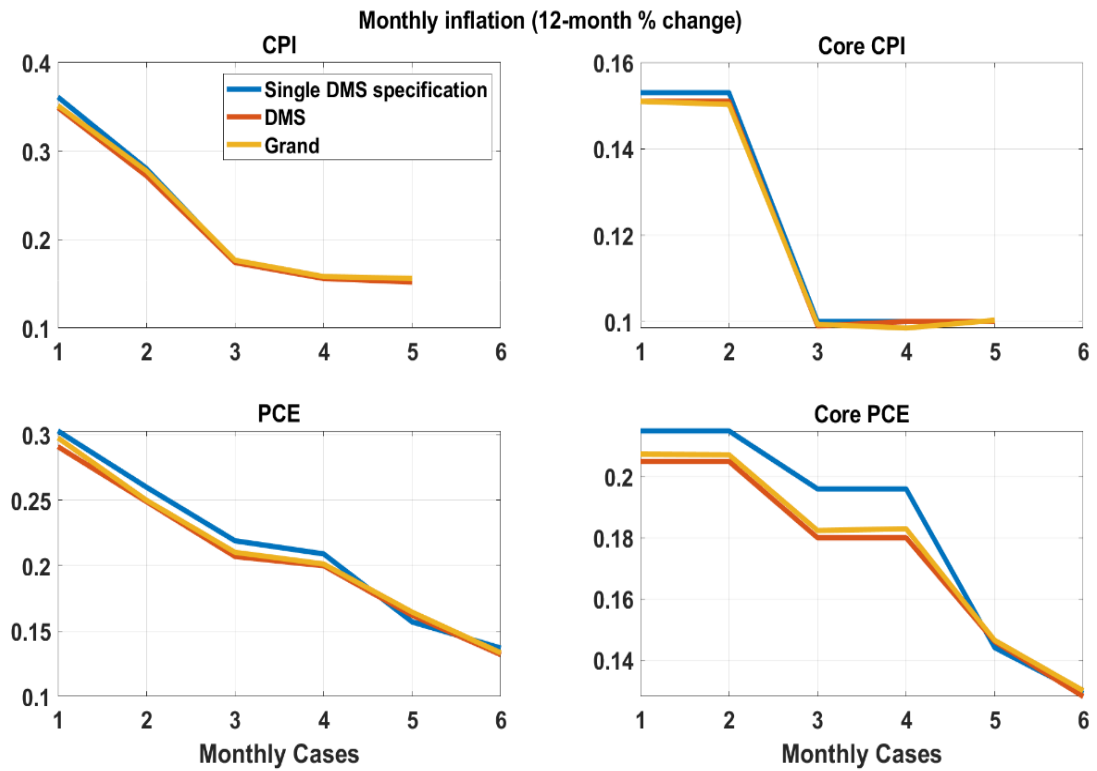
Notes: The figure plots the RMSEs for grand combinations based on log score, CRPS, equal, CMG, and Ganics weighting schemes and using the flexible aggregation strategy. The cases reflect the point in time when each nowcast was made relative to the target nowcast month; see Table 3.2. The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure 3.6: Density Performance of Grand Combination vs. Its Components



Notes: The top panel plots the average log score and the bottom panel plots the average CRPS for the grand combination based on the log score weighting scheme and combinations based on the DMS model class, MIDAS model class, and DFM model class, where each individual model class uses the log score weighting scheme. The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

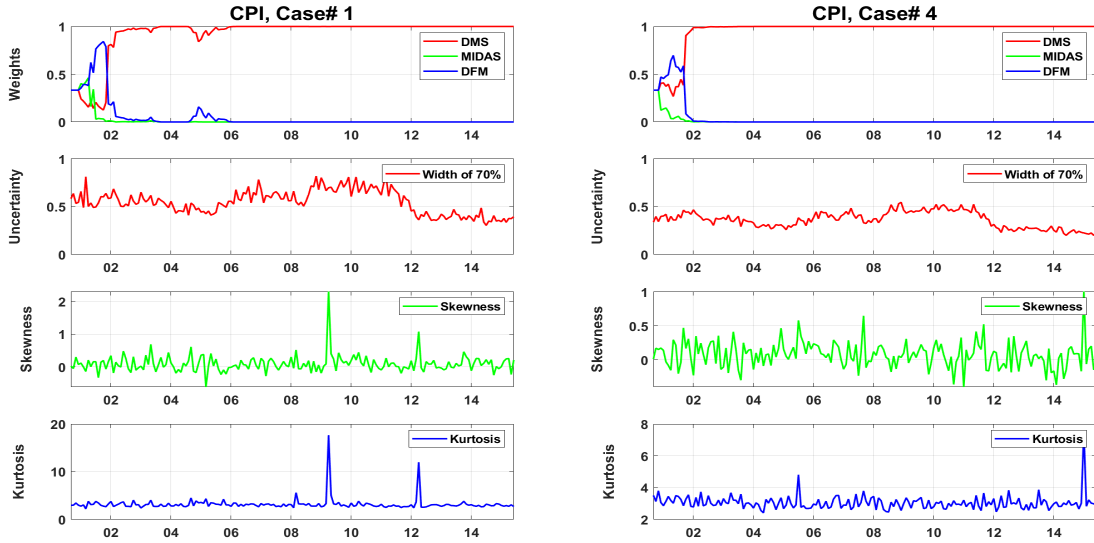
Figure 3.7: Point Nowcasting Performance, Grand Combination vs. DMS



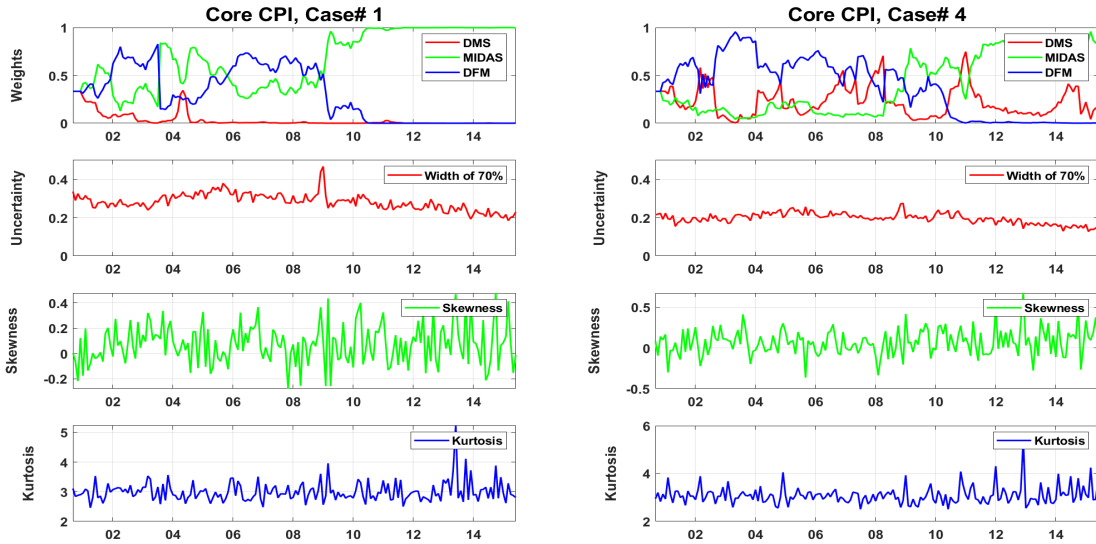
Notes: The figure plots the RMSE for the grand combination based on log score and using the flexible aggregation strategy; the stage 1 combination from the DMS model class; and a single specification from the DMS model class based on Knotek and Zaman (2017). The cases reflect the point in time when each nowcast was made relative to the target nowcast month; see Table 3.2. The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure 3.8: Weights and Higher-Order Moments (CPI and Core CPI)

(a) CPI inflation



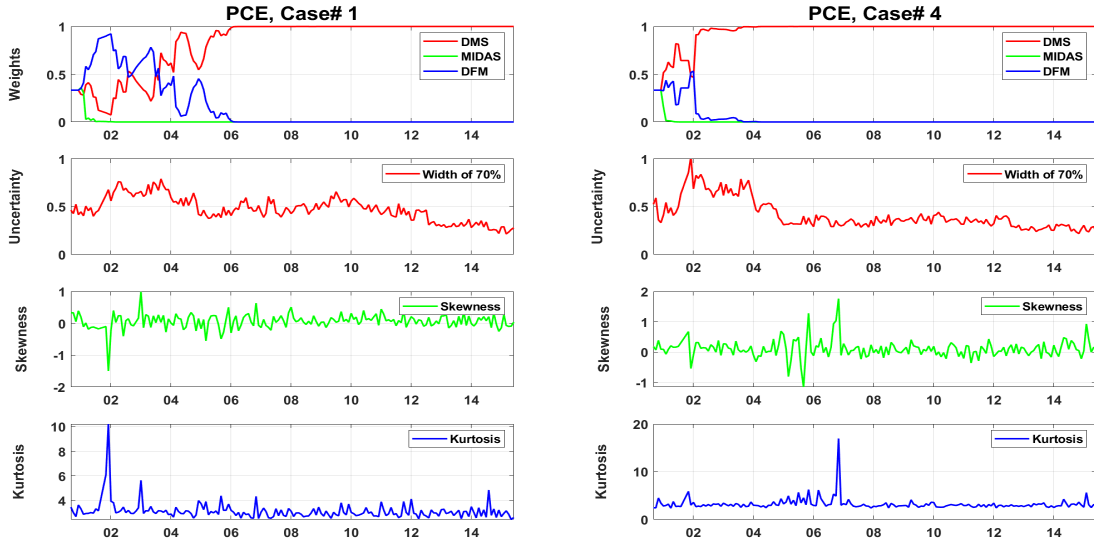
(b) Core CPI inflation



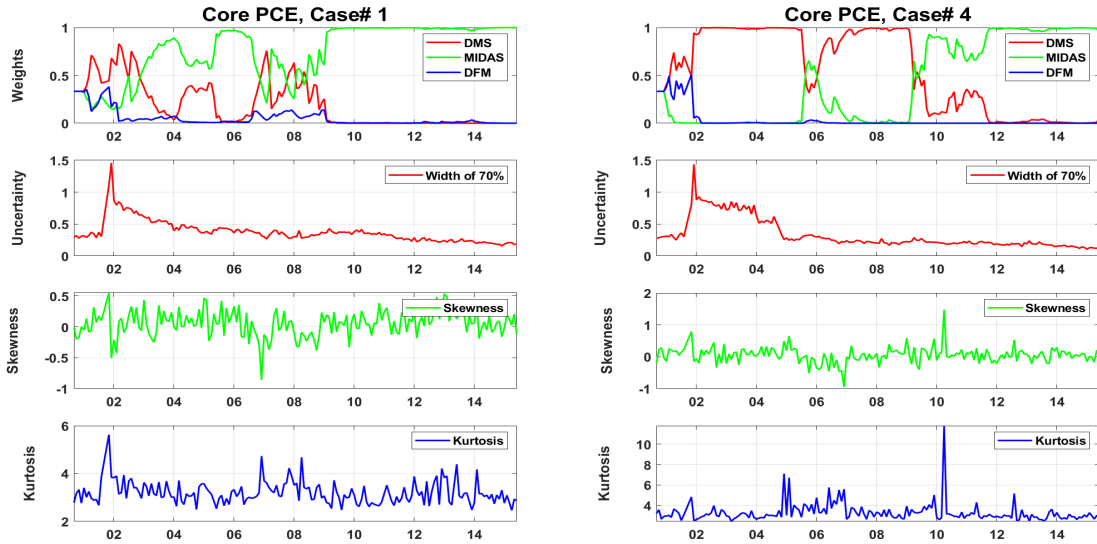
Notes: The first row of each panel plots the evolution of the weights for the three model classes underlying the grand combination, based on the flexible aggregation strategy and log-score weighting scheme. (Each model class is a combination of multiple model specifications.) The second row plots estimates of dynamic uncertainty, defined as the width of the 70% prediction intervals. The last two rows plot time-varying estimates of skewness and kurtosis. The sample period spans September 2000 through June 2015.

Figure 3.9: Weights and Higher-Order Moments (PCE and Core PCE)

(c) PCE inflation



(d) Core PCE inflation



Notes: The first row of each panel plots the evolution of the weights for the three model classes underlying the grand combination, based on the flexible aggregation strategy and log-score weighting scheme. (Each model class is a combination of multiple model specifications.) The second row plots estimates of dynamic uncertainty, defined as the width of the 70% prediction intervals. The last two rows plot time-varying estimates of skewness and kurtosis. The sample period spans September 2000 through June 2015.

Chapter 4

A Unified Framework to Estimate Macroeconomic Stars

4.1 Introduction

The economic concept of a long-run equilibrium level (often denoted with the “star” symbol) is central to macroeconomics. These long-run levels are thought to reflect the fundamental structure of the economy (in the absence of shocks). Hence, they are used as reference points, and deviations from these long-run levels reflect cyclical or idiosyncratic fluctuations. Dynamics of cyclical changes and the pace of their adjustment towards the long-run equilibrium are of interest to macroeconomists. However, effective identification of cyclical fluctuations in any macroeconomic aggregate requires knowledge of its long-run equilibrium.

In this chapter, we estimate jointly range of macroeconomic stars that are of broader interest to macroeconomists and policymakers: the equilibrium rate of productivity growth (p-star), the level of potential output (gdp-star), the growth rate of potential output (g-star), the equilibrium level of the unemployment rate (u-star), the equilibrium level of the real rate of interest (r-star), the equilibrium rate of price inflation (pi-star), the equilibrium rate of nominal wage inflation (w-star).¹ The assumption that a long-run equilibrium exists implies that in the long-run, the economy is growing at potential, inflation equals its trend rate, the unemployment rate reflects a combination of structural and frictional factors, i.e., there is no cyclical pressure, nominal wages grow at a rate equal to the sum of labor productivity growth and price inflation, and the real interest rate reflects the rate consistent with output growing at potential and stable inflation.²

¹The subset of these stars, p-star, g-star, u-star, and r-star, reflect the fundamental structural features of the economy whereas others, pi-star and w-star are thought to be influenced by central banks and monetary policy.

²The literature has referred to the concept of long-run equilibrium using different terminologies, such as “natural”, “neutral”, “trend”, “steady-state”, and “long-run”. There are subtle differences among them, but they can be interpreted as the same for the purpose of this chapter. In some studies, especially using Dynamic Stochastic General Equilibrium (DSGE) models, the concept of the natural rate refers to medium-horizon equilibrium, and in these same models, the concept of steady-state is used to refer to the long-run equilibrium.

As discussed in Weber et al. (2008), conceptually, determining stars' estimates appears straightforward, however, in practice, it is fraught with difficulties. The difficulties stem from the fact that the values of the stars and their determinants are unobserved. To infer the estimates of the stars, a range of statistical and econometric methods are applied to observable historical data. These methods range from statistical univariate filters (e.g., Hodrick and Prescott, 1997; Christiano and Fitzgerald, 2003; Ashley and Verbrugge, 2009) to multivariate models, including semi-structural time-series models (e.g., Pescatori and Turunen, 2016; Morley and Wong, 2020), and fully structural DSGE models (e.g., Del Negro et al., 2017).

Economic theory posits that the structural aspects of the economy, which inform the values of the stars, change slowly. Therefore, methods that produce estimates of stars that change only gradually have more traction than methods that give less smooth estimates. According to this criterion, multivariate unobserved components (UC) models, which are statistical models that use economic theory to frame the empirical specification, have been shown to provide reasonable estimates of the stars (e.g., Kuttner, 1994; Basistha and Nelson, 2007; Laubach and Williams, 2003; Chan, Koop, and Potter, 2016). Hence, they are the dominant methods for obtaining time-varying estimates of the stars.

However, with few exceptions, the popular multivariate UC models that estimate time-varying stars focus on a small number of observables, often just two or three, and minimal structure (e.g., Laubach and Williams, 2003). Studies that entertain more variables have abstracted from important empirical features such as time-varying parameters and stochastic volatility (e.g., Del Negro et al., 2017; Fleischman and Roberts, 2011). A priori, one would expect a framework based on greater information that explicitly permits (contemporaneous) interactions between stars and between cyclical components and a richer structure to provide more reliable estimates of the objects of interest (e.g., stars) than frameworks that ignore them. Indeed, Taylor and Wieland (2016) make a similar argument in the importance of appropriately accounting for trends in the determinants of the r-star to estimate r-star correctly. Specifically, Taylor and Wieland (2016) point out, "What appear to be trends in the equilibrium interest rate may instead be trends in other policy variables that affect the economy."

Accordingly, in this chapter, we take on the challenge of jointly estimating several macroeconomic stars simultaneously, including g-star (and gdp-star), u-star, p-star, pi-star, w-star, and r-star, using a semi-structural time series model (aka multivariate UC model with a particular structure informed from economic theory). For each star, we formulate a rich structure whose elements are guided by past research. For example, econometric estimation of r-star is informed from various sources: Investment-Savings (IS) equation, Taylor-type rule, equation linking g-star and r-star, and equation relating r-star to survey expectations. We allow for time-variation in important macroeconomic relationships and error variances. Fernandez-Villaverde and Rubio-Ramirez (2010), Koop and Korobilis (2010), Carriero et al. (2019), among many others, highlight the importance of allowing for stochastic volatility in macroeconomic models. Incorporating these empirical features should better distinguish between cyclical fluctuations

and lower-frequency movements in the macroeconomic aggregates considered in this paper.

We extend the Chan, Clark, and Koop (2018) – henceforth CCK – approach of using long-run survey expectations to improve pi-star precision to other macroeconomic stars. Specifically, for each macroeconomic variable of interest, we explicitly model the link between the unobserved “star” and the expectations about the star contained in the Blue Chip survey of economic forecasters (or reported by the Congressional Budget Office [CBO] when the survey estimate is not available).³ In a high-dimensional model like ours, the use of long-run survey expectations, which are direct measures of stars, could help anchor model-based estimates of stars to reasonable values (especially in times of heightened uncertainty) and potentially improve the precision of the estimates. We estimate our feature-rich UC model using Bayesian techniques, specifically the efficient sampling techniques developed in Chan, Koop, and Potter (2013) and the precision sampler proposed in Chan and Jeliazkov (2009).

All in all, the combination of time-varying parameters, SV, joint modeling of multiple stars, implementation of an expanded structure, and allowing for a direct connection between stars and long-term survey expectations is what differentiates our UC model from the existing literature. Many popular UC models could be viewed as special cases of our larger UC model, which facilitates model comparison. We note that among the stars considered, w-star has received less attention in the literature. Our UC model’s ability to provide real-time estimates of w-star and model-based decomposition into its determinants p-star and pi-star, as implied by economic theory, is a novel contribution. This specific decomposition is useful to monetary policymakers, who often refer to developments in nominal wages to support their forecasts and related discussions on price inflation and employment.

Our results indicate that there are payoffs to modeling stars jointly using a larger multivariate UC model. The metric of Bayesian model comparison generally favors our larger UC model over smaller-scale UC models. The model yields credible estimates of stars (and the output gap). For example, the output gap estimate is similar to the CBO estimate based on a production function approach. Generally, the contours of stars echo those documented elsewhere in the literature but at times are different, and these differences can matter for policy. For example, let’s consider pi-star, from 2000 to 2010, our UC model has an estimate of the pi-star stable at or close to 2%, whereas pi-star from a popular univariate model (of Stock and Watson, 2007) displays notable fluctuations around 2%, and the bivariate model of CCK indicates a stable pi-star about few tenths below 2%. These differences in pi-star matter for central banks tasked with inflation targeting. Compared to some of the popular UC models and the smaller-scale restricted variants of our larger UC model, the precision estimates of the stars (and the output gap) from our UC model are among the most precise, where precision is measured as the width of 90% credible interval. The model’s reliance on long-term survey expectations data is the

³The long-run survey expectations can be thought of as a hybrid forecast because it combines judgment based on a range of information and forecast derived from a range of modeling approaches. Our use of such a hybrid forecast implicitly serves as an additional channel through which the issue of omitted variable bias is mitigated.

key reason for this improved precision.⁴ Survey expectations played a crucial role in guiding the model-based assessment of stars during the COVID-19 pandemic, a period of heightened uncertainty. Our UC model’s real-time point and density forecasting accuracy rivals and, in some cases, outperform hard-to-beat benchmarks, including small-scale UC models.

We also demonstrate the usefulness of stars’ estimates (from our UC model) as terminal values for external models. Previous research shows that forecasting models, such as (steady-state) vector autoregression (VAR) model, often benefit significantly (in terms of improved forecast accuracy) from external information about steady-states informed by long-run survey expectations (e.g., Wright, 2013). Using a real-time forecasting horse race, we show that if we were to inform steady-states in a VAR with the stars from our UC model, gains in forecast accuracy (for some of the variables) are achieved compared to the standard approach relying on survey expectations. Hence, our framework provides a potential source for obtaining the stars’ estimates in real-time. An advantage of our framework compared to surveys is that it provides estimates of stars (steady-states) for variables not covered by surveys (such as w-star) and offers both point and density estimates for the stars.

We summarize three additional findings. First, we find that the empirical evidence in the link between r-star and g-star, as implied by theory, is weak (consistent with Hamilton et al., 2016; Lunsford and West, 2019), but by bringing survey expectations into the model, the link becomes stronger (providing support to Laubach and Williams, 2016). Second, our results indicate economically and statistically significant evidence of time-variation in our model parameters capturing macroeconomic relationships and strong support for SV’s inclusion in our model equations. It lends support to the popular narratives: “(price) Phillips curve has weakened over time,” “wage Phillips curve is alive,” and “weakening in the procyclicality of labor productivity.” Third, a comparison between final and real-time estimates of the stars indicates that their broad movements have generally tracked each other closely. We view this as a valuable finding because it suggests that we make some progress in mitigating the well-known difficulties associated with the real-time estimation of the stars.

In recent years, several papers have provided estimates of the stars using UC models with more indicators and or an expanded structure. For example, Johannsen and Mertens (forthcoming) [henceforth JM], Pescatori and Turunen (2016), Del Negro et al. (2017), Brand and Mazelius (2019), Gonzalez-Astudillo and Laforte (2020), among others, have examined the roles of additional determinants in explaining r-star. Neither of these studies feature time-varying parameters, and only JM allows for SV, but their model size is significantly smaller than ours. Chan, Koop, and Potter (2016) [henceforth CKP] illustrate the value of modeling u-star and pi-star as bounded random walk processes in a bivariate Phillips curve. More recently, using a fixed-parameter UC model, Crump et al. (2019) estimate u-star by combining a range of labor market indicators across demographic groups and survey expectations of inflation. This chapter

⁴Our precision estimates are on a par with recent studies highlighting the improved precision of stars derived from their approaches (e.g., r-star by Del Negro et al., 2017; u-star by Crump et al., 2019; pi-star by CCK).

goes further by implementing a UC model that estimates long-run equilibrium values for many variables with a rich dynamic structure, including stochastic volatility and time-varying parameters, and uses survey information about the long-run expectations of inflation, unemployment rate, real GDP growth, and interest rates.⁵

The chapter is organized as follows. The next section describes in detail the econometric model and its variants. Section 4.3 describes the data and estimation. Section 4.4 presents and discusses in detail the estimates of stars and other model parameters. Section 4.5 reports the real-time forecasting results and a discussion comparing real-time and final estimates of stars. Section 4.6 illustrates the ability of the model to handle the COVID-19 pandemic data. Section 4.7 concludes. This chapter has a supplementary online appendix that lists detailed Bayesian estimation steps and some additional results.

4.2 Empirical Macro Model and Variants

The ingredients of our macroeconomic econometric model are guided by economic theory, and in part, by features that previous research shows to be empirically relevant. These features include stochastic volatility and time-varying parameters, which in turn imply time-varying predictability. Collectively, these empirical features permit modeling changing macroeconomic relationships in a flexible way.

We represent our empirical model using six sets of equations, which we denote blocks. These six blocks, which allow for *contemporaneous interactions* between them, characterize the joint dynamics of the unemployment rate, output growth, labor productivity growth, price inflation, nominal wage inflation, nominal interest rate, and corresponding stars. To be sure, the model assumes that all innovations are uncorrelated both serially and across equations. However, we emphasize that any assumed current period correlations between the cyclical components and or between stars are directly modeled via the model equations that define the contemporaneous relationships between the components (e.g., cyclical output gap at time t with cyclical unemployment gap at t ; r-star and g-star).

Before we jump into the model description, we describe some basics, including the econometric definition of the star and the usefulness of survey long-run expectations in the estimation of stars.

4.2.1 The econometric notion of a long-run equilibrium

Following CCK, Mertens (2016), Lee and Nelson (2007), Morley (2002), among many, this chapter defines the long-run equilibrium (or star) of a particular macroeconomic series as its

⁵Morley and Wong (2020) and Chan (2019) propose alternative modeling framework based on VARs to estimate the long-run equilibrium values. The advantage of the VAR based framework is the ability to handle larger amounts of information conveniently and flexibly compared to UC models. On the other hand, the advantage of UC modeling, as emphasized by CKP, is the availability of the direct estimates of stars, which in our case proves quite convenient to allow for direct modeling of the relationships between various stars.

infinite-horizon forecast conditional on the current information set. This definition of a star is consistent with the notion of Beveridge-Nelson trend decomposition, and extensive literature has adopted this approach to estimate stars. Equivalently, as commonly defined in the trend estimation literature, the infinite horizon forecast could be viewed as an estimate of trend conditional on the current information set (e.g., CCK, Garcia and Poon (2018), Mertens (2016), Lee and Nelson (2007), Morley (2002)). As discussed in Mertens (2016), among others, different information sets would likely yield different estimates of the infinite-horizon forecast (or trend). Mertens showed that including survey projections of long-term inflation (hereafter survey long-run forecasts) in the information set led to more precise and forward-looking estimates of trend inflation.

The link between the infinite-horizon forecast and the underlying trend is described well by the unobserved components (UC) model (see Laubach and Williams, 2003; Lee and Nelson, 2007; Mertens, 2016; CCK). In a UC model, a series (Y_t) is typically represented as a sum of a nonstationary trend component Y_t^* , which is assumed to evolve slowly and a stationary cycle Y_t^c , whose infinite-horizon conditional expectation is assumed to be zero. Accordingly,

$$Y_t = Y_t^* + Y_t^c. \quad (4.1)$$

The trend component Y_t^* is interpreted as the limiting forecast of the series (conditional on the information set I_t) as the forecast horizon tends to infinity,

$$\lim_{j \rightarrow \infty} \mathbb{E}[Y_{t+j}|I_t] = Y_t^*. \quad (4.2)$$

Differencing the above equation yields,

$$Y_t^* = Y_{t-1}^* + \lim_{j \rightarrow \infty} \mathbb{E}[Y_{t+j}|I_t] - \lim_{j \rightarrow \infty} \mathbb{E}[Y_{t+j}|I_{t-1}] = Y_{t-1}^* + e_t, \quad e_t \sim N(0, \sigma_e^2). \quad (4.3)$$

which suggests a random walk process for the trend Y_t^* .⁶ It also suggests a stationary, ergodic mean-zero process for Y_t^c .

Intuitively, the above set of assumptions imply that once the effects of the shocks have fully played out, the macroeconomic series of interest, Y_t gravitates to its underlying trend level, Y_t^* .

As discussed in CCK, various statistical and econometric models could fit within the above-specified decomposition. This chapter formulates a specific unobserved components time series model and its variants.

⁶The commonly adopted assumption of modeling Y_t^* as a random walk is partly due to consensus among macroeconomists that the factors driving the long-run equilibrium levels are perceived to be quite persistent (e.g., Lee and Nelson, 2007). In practice, the assumption of a (driftless) random walk has generally worked quite well, in that it has provided reasonable estimates of the stars (e.g., Clark, 1987; Kuttner, 1994; Laubach and Williams, 2003; Mertens, 2016).

4.2.2 The role of survey expectations

As discussed in the introduction, an important contribution of this chapter is to provide a direct role for long-run survey expectations in refining the stars' estimates. Specifically, we follow the approach of CCK (and Pescatori and Turunen, 2016). These papers explicitly estimate an equation linking the observed measure of a long-run forecast obtained from external sources (survey in the case of CCK and CBO projection of the output gap in Pescatori and Turunen, 2016) to an unobserved object of interest. We extend their approach to the macroeconomic variables considered in this chapter.

Several papers have documented an essential role of long-run survey (and institutional) forecasts in helping refine the econometric estimation of model parameters, including the latent components (e.g., pi-star: Kozicki and Tinsley, 2012; Mertens, 2016; CCK; Mertens and Nason, 2020; gdp-star: Pescatori and Turunen, 2016; r-star: Del Negro et al., 2017). Specifically, Mertens and Nason (2020), CCK, Mertens (2016), and Kozicki and Tinsley (2012), in using different methodologies (in combining survey data with model forecasts) to estimate trend in US inflation, show that survey long-run forecasts of inflation deliver crucial additional information (beyond the recent inflation history) in refining trend estimates and improving model fit. In a similar vein, Pescatori and Turunen (2016) document the usefulness of CBO's projection of the potential output gap in improving their model's output gap precision. It is this particular literature that motivates us to consider survey long-run forecasts in our large-scale econometric model.

The advantage of survey (and institutional) forecasts stems from the fact that they could be viewed as hybrid forecasts, i.e., a combination of judgment and forecasts derived from various modeling approaches. The fact that human judgment enters into survey expectations is an important reason for their success. As discussed by Kozicki and Tinsley (2012) and others, the good forecasting properties are partly because survey participants have at their disposal a wide range of indicators, including central bank communications, and information about changes in the tax laws, etc. The patterns gleaned from this large information set can help shape opinions, including any perceived structural change, which can immediately influence expectations about the long-run.

The usefulness of this hybrid forecast measure is appealing from both theoretical and practical perspectives. From a theoretical standpoint, the notion of such a type of hybrid forecast is well-aligned with the New-Keynesian view that emphasizes the importance of allowing for interlinkages between the current state of the economic variables and their expectations about the future in determining the relevant stars (see Weber et al., 2008).

From a practical point of view, such a hybrid measure helps limit the number of variables (i.e., information) that need to be brought into the model. Put differently, explicitly utilizing survey long-run expectations through an equation linking these expectations to the corresponding model's latent objects could be thought of as a short-cut to enrich the necessary information set used in model estimation. Furthermore, in high-dimensional models, such as the one de-

veloped in this chapter, the use of survey long-run projections, which are targeted and direct measures of stars, could help anchor model-based estimates of stars to reasonable values and have the potential to improve precision of the estimates.

Accordingly, in this chapter, with the exceptions of nominal wage inflation and labor productivity, for each of the remaining four variables, we model the direct link between long-run survey projection (or the long-run CBO projection in the years for which survey projections are unavailable) and the corresponding star using the following econometric equations⁷,

$$Z_t^y = C_t^y + \beta^y y_t^* + \varepsilon_t^{zy}, \quad \varepsilon_t^{zy} \sim N(0, \sigma_{zy}^2), \quad y = \pi, u, g, r \quad (4.4)$$

$$C_t^y = C_{t-1}^y + \varepsilon_t^{cy}, \quad \varepsilon_t^{cy} \sim N(0, \sigma_{cy}^2), \quad y = \pi, u, g, r \quad (4.5)$$

where π refers to price inflation, u refers to the unemployment rate, g refers to real GDP growth, r refers to the real interest rate, Z_t^y refers to the survey long-run forecast corresponding to the variable y , and y_t^* is the unobserved y star.

C_t^y is the time-varying intercept assumed to evolve as RW process to possibly capture the permanent wedge between survey estimate and the model-based star. This wedge can arise due to several reasons, including the fact that star is assumed to be the infinite-horizon forecast, whereas the survey forecast refers to the average forecast for the five-year period starting seven years into the future in the case of BC and ten-year ahead forecast in the case of SPF (for price inflation).

The above set of equations define a simplistic and flexible relationship between the long-run survey expectations and the star.⁸

4.2.3 Unemployment block

The long-run equilibrium level of unemployment (u-star) is the unemployment rate that prevails when output is growing at potential, and the economy adds jobs so as to maintain the full-employment level. This equilibrium level of unemployment is primarily the result of labor-market imperfections caused by frictional and structural aspects of the labor markets.

As discussed in Crump et al. (2019), two approaches are commonly used to estimate u-star. The first approach applies UC modeling to detailed labor market data (such as job vacancies, firm's recruiting intensity, demographic changes, flows into and out of unemployment) to extract respective trends. These trends are used to construct implied estimates of u-star (e.g., Daly, Hobijn, Sahin, and Valleta, 2012; Davis, Faberman, and Haltiwinger, 2013; Barnichon and Mesters, 2018; Tasci, 2012). The second approach uses a combination of information from prices (and or nominal wages, survey expectations) and the estimated Phillips curve relationship

⁷For the long-run inflation forecast, we use the Survey of Professional Forecasters (SPF)/PTR, and for the long-run forecasts of the other three variables, we use the Blue Chip (BC) survey.

⁸We note that we adopt a relatively less flexible relationship between survey forecasts and stars than CCK.

between price inflation and the aggregate unemployment rate to backout the estimate of u -star.⁹ Various modeling techniques ranging from parsimonious UC modeling to structural DSGE models are applied to estimate the Phillips curve relationship (e.g., Staiger, Stock, and Watson, 1997; Orphanides and Williams, 2002; Lee and Nelson, 2007; CKP; Gali, 2011).

Following CKP, we posit that the observed unemployment rate is decomposed into a (bounded) RW trend component (u -star) and a stationary cyclical component. The cyclical component is modeled as an AR(2) process. The use of a parsimonious (time-invariant) AR2 process to identify the cyclical component of the unemployment rate is a commonly used assumption, in our case motivated by a recent string of empirical studies, e.g., Lee and Nelson (2007), Gali (2011), Stock and Watson (2015), CKP, Tallman, and Zaman (2017), and Gali and Gambetti (2020), who all document reasonable patterns of the cyclical unemployment component (i.e., movements in this component correlate quite well with the NBER business cycle).¹⁰

More generally, the assumption of an AR2 to model the cyclical component of macroeconomic variables has a long tradition, at least going back to Clark (1987). However, because we are also modeling the output gap (i.e., level of real GDP minus level of potential real GDP), we depart from the previous literature by augmenting the AR2 unemployment gap with the output gap (denoted $ogap$) as an additional explanatory variable. Sinclair (2009), Grant and Chan (2017), and Berger, Everaert, and Vierke (2016), among several others, document the empirical importance of jointly modeling the unemployment rate gap and the output gap.

Similarly, as shown later, we add information from the unemployment gap when modeling the output gap. We find that the joint modeling of the output gap and the unemployment gap is empirically and economically useful (confirming previous research findings, e.g., Fleischman and Roberts, 2011). In addition, joint modeling of both the output gap and the unemployment gap allows us to estimate the strength in the relationship between the two cyclical components, popularly known as Okun’s law. In equation (4.7), the coefficient ϕ_u captures the contemporaneous relationship between the output gap and the cyclical unemployment rate gap. The estimate, $\frac{1-\rho_1^u-\rho_2^u}{\phi_u}$ could be interpreted as the Okun’s law coefficient.

We note that when jointly modeling output and the unemployment rate, most researchers assume a common cyclical component between the two. However, in light of the empirical evidence that cyclical unemployment displays more persistence than the output gap (e.g., Berger et al., 2016), we model two separate cycles linked to each other via Okun’s law relationship. As shown in Berger et al. (2016), a specification that entertains two separate cycles (cyclical unemployment and the output gap), the data support a time-invariant parameter describing

⁹As mentioned in Crump et al. (2019), one of the criticisms of this approach is that it will be impacted by the breakdown of the Phillips curve relationship post 2007. However, by allowing time-variation in the coefficients capturing the price and wage Phillips curve relationships, as we do, our approach should face less of a problem. In addition, as illustrated in Del Negro, Giannoni, and Schorfheide (2015) and Clark (2014) bringing information from long-run survey expectations of inflation (as we do) should further help capture the inflation behavior in the post 2007 period.

¹⁰CKP explores the empirical importance of allowing for time-variation in the parameters of AR2 process, and finds that data prefers the time-invariant AR2 process, hence validating the widely used assumption of a simple AR2 process.

the Okun's law relationship. In contrast, a specification with a common cyclical component favored a time-varying Okun's law relationship (adding support to Knotek, 2007).

In a nutshell, the inference of Berger et al. (2016) suggests that time-variation in the parameter capturing Okun's law reflects the sluggish response of the cyclical unemployment to movements in the output gap. Once they allowed for a sluggish response of the cyclical unemployment rate by adding persistence, via one-period lag of the cyclical unemployment rate, evidence of a time-varying Okun coefficient disappears. We found similar evidence. Hence, we adopt the modeling of two separate cycles and a time-invariant Okun's law relationship for our baseline setup; this has the added advantage of requiring estimation of significantly fewer parameters.¹¹

With the exception of CKP, most of the literature models u -star as a driftless RW. The use of an unrestricted RW process has empirically been shown to work well, but CKP shows that modeling u -star as a bounded RW process is even better. Motivating their use of bounds is because, by construction, the unemployment rate is a bounded variable, which implies that the long-run equilibrium in the labor market would restrict the movements in u -star within a bounded interval.¹²

Accordingly, we model u -star as a bounded RW, where the bounds' values are fixed at 3.5% (lower bound) and 7.5% (upper bound).¹³ Following CKP, the variance of the error in the cyclical component is not allowed to vary over time.

$$U_t = U_t^* + U_t^c \quad (4.6)$$

$$U_t - U_t^* = \rho_1^u(U_{t-1} - U_{t-1}^*) + \rho_2^u(U_{t-2} - U_{t-2}^*) + \phi^u \text{ogap}_t + \varepsilon_t^u, \quad \varepsilon_t^u \sim N(0, \sigma_u^2) \quad (4.7)$$

where, $\rho_1^u + \rho_2^u < 1$, $\rho_2^u - \rho_1^u < 1$, and $|\rho_2^u| < 1$; $\phi^u < 0$

$$U_t^* = U_{t-1}^* + \varepsilon_t^{u*}, \quad \varepsilon_t^{u*} \sim TN(a_u - U_{t-1}^*, b_u - U_{t-1}^*; 0, \sigma_{u*}^2) \quad (4.8)$$

where the notation $TN(a, b; \mu, \sigma^2)$ refers to normal distribution with mean μ and variance σ^2 but truncated in the interval (a, b) .

¹¹Bayesian model comparison assessment slightly preferred the approach of two separate cycles with time-invariant Okun's law compared to common cycle with time-varying Okun's law parameter. Also, we note that Sinclair (2009) and Grant and Chan (2017) model two separate cycles, linked through a time-invariant parameter.

¹²CKP argue that economic forces that govern the movements in u -star are slow-moving and those forces would not cause the unemployment rate to fall to levels close to zero or to levels that are higher than the previous peaks caused by recessions.

¹³These values are informed by estimating the CKP model over our estimation sample, and are close to values reported in CKP based on their estimation sample. As a further check, most estimates of the u -star reported in commonly cited literature fall within the bounds we use in this chapter.

Lastly, the equation relating long-run survey projection of the unemployment rate to U^* ,

$$Z_t^u = C_t^u + \beta^u U_t^* + \varepsilon_t^{zu}, \quad \varepsilon_t^{zu} \sim N(0, \sigma_{zu}^2) \quad (4.9)$$

$$C_t^u = C_{t-1}^u + \varepsilon_t^{cu}, \quad \varepsilon_t^{cu} \sim N(0, \sigma_{cu}^2) \quad (4.10)$$

4.2.4 Output block

We are interested in both the potential output (i.e., gdp^*) and the growth rate in potential output (i.e., g^*). To feasibly estimate both these latent processes, we follow the commonly adopted approach, which decomposes the level of aggregate output into the level of potential output and a cyclical component (output gap), where the cyclical component is defined as the deviation of observed aggregate output level from potential output. This simple decomposition has a long tradition going back to Harvey (1985) and Clark (1987).

$$gdp_t = gdp_t^* + ogap_t \quad (4.11)$$

where $gdp \equiv \log(GDP)$ and gdp^* refers to potential output, which is unobserved.

A common approach to model gdp^* is to assume a random walk with a time-varying drift term (i.e., the local level trend process), where the time-varying drift term (interpreted as g^*) is assumed to follow a random walk process (to allow for a stochastic g^*). More recently, Chan and Grant (2017) show that a model where gdp^* is assumed to follow a second-order Markov transition process fits the data better compared to the model where gdp^* is assumed as a random walk with time-varying drift. Hence, we follow Chan and Grant (2017) and model gdp^* as,

$$gdp_t^* = 2gdp_{t-1}^* - gdp_{t-2}^* + \varepsilon_t^{gdp^*}, \quad \varepsilon_t^{gdp^*} \sim N(0, \sigma_{gdp^*}^2) \quad (4.12)$$

Which can be re-written as,

$$\Delta gdp_t^* = \Delta gdp_{t-1}^* + \varepsilon_t^{gdp^*}$$

If we define, $g_t^* \equiv \Delta gdp_t^*$, where Δ is the first difference operator. Then,

$$g_t^* = g_{t-1}^* + \varepsilon_t^{gdp^*} \quad (4.13)$$

An advantage of modeling g^* as a second-order Markov process compared to RW with time-varying drift is that it requires estimating a single shock parameter ($\sigma_{gdp^*}^2$), as opposed to two for the latter (one for the shock to gdp^* and the other for the shock to time-varying drift, aka g^*). This modeling assumption implies that all permanent shocks to output are attributed as shocks to g^* .¹⁴

¹⁴Note: the second-order Markov process for gdp^* could be thought of as a limiting case of the process.

The cyclical component, $ogap$, is assumed a stationary AR(2) process augmented with additional explanatory variables: the real interest rate gap and the unemployment gap,

$$ogap_t = \rho_1^g(ogap_{t-1}) + \rho_2^g(ogap_{t-2}) + a^r(r_t - r_t^*) + \lambda^g(U_t - U_t^*) + \varepsilon_t^{ogap} \quad (4.14)$$

where, $\varepsilon_t^{ogap} \sim N(0, \sigma_{ogap}^2)$, $\rho_1^g + \rho_2^g < 1$, $\rho_2^g - \rho_1^g < 1$, and $|\rho_2^g| < 1$; $\lambda^g < 0$

Equation (4.14) could be interpreted as defining an IS-curve (as in LW and subsequent papers modeling r -star) that allows feedback from the real interest rate gap to the output gap (i.e., the real interest rate gap responds to economic slack). The IS equation is inspired by LW but with two modifications. First, instead of using the interest-rate gap based on the short-term real rate of interest, we use the long-term real interest rate (Gonzalez-Astudillo and Laforte, 2020). Specifically, the long-term real interest rate is constructed as the difference between the nominal yield on a 10-year Treasury bond and the 10-year inflation expectations (i.e., the PTR series for PCE inflation).¹⁵

In theoretical models, the long-term interest rate influences household consumption decisions and business investment decisions. Second, to improve the econometric estimation of the output gap, we enrich the IS equation by bringing in information from the unemployment gap (from the unemployment block) as an explanatory variable.¹⁶ This latter addition is motivated by the approach taken in a long list of papers (e.g., Morley and Wong, 2020; Chan and Grant, 2017; Fleischman and Roberts, 2011; Sinclair, 2009; Clark, 1989) that demonstrate the usefulness of the unemployment rate in improving the econometric estimation of the output gap. As mentioned earlier, in the equation for the unemployment gap, we add the output gap to improve the former's estimation. The coefficient λ^g captures the contemporaneous relationship between the output gap and the unemployment gap. The parameter a^r relates the output gap to the real interest rate gap.

We note that innovations $\varepsilon_{gdp^*}^2$ and ε_{ogap}^2 are uncorrelated. In an important contribution, Morley et al. (2003), who assumes a deterministic g -star, show that this assumption matters for

Assuming gdp^* as a RW with time-varying drift term, where the variance of shock to the gdp^* goes to zero, and shock to the time-varying drift term (g -star) is the only relevant driver governing the evolution of gdp^* and g -star (see Chan and Grant, 2017).

¹⁵We also experimented with an alternative specification, in which the interest rate gap is constructed as the difference between the short-term federal funds rate and first lag of four-quarter trailing PCE inflation, similar in spirit to LW. Based on model fit (log marginal likelihood), this specification was inferior compared to the Base specification. It is worth noting that had the longer history of long-term inflation expectations data available at the time of the writing, LW would have constructed the interest rate gap using the long-term interest rate (see page 1064 in LW).

¹⁶Model fit, the precision metric for u -star and the output gap, and the plausibility of the estimates of output gap strongly supports the joint modeling of output gap and unemployment gap. We note that LW estimated an alternative specification in which they added information from the labor market (hours worked) and found that doing so improved the precision of the estimated output gap, however, that improved precision did not spillover to r -star estimate (the focus of their paper) and therefore in their baseline specification they omit the labor market variable.

estimating potential output. However, Chan and Grant (2017) show that in their specification, once a stochastic g-star is allowed for, the correlation between $\varepsilon_{gdp^*}^2$ and ε_{ogap}^2 goes to zero. They also show that the model without correlation performs comparably to the model with correlated innovations based on Bayesian model comparison. Accordingly, to keep estimation tractable, we assume uncorrelated innovations.

The equation linking survey projection of potential growth rate to g^* ,

$$Z_t^g = C_t^g + \beta^g * 4 * g_t^* + \varepsilon_t^{zg}, \quad \varepsilon_t^{zg} \sim N(0, \sigma_{zg}^2) \quad (4.15)$$

$$C_t^g = C_{t-1}^g + \varepsilon_t^{cg}, \quad \varepsilon_t^{cg} \sim N(0, \sigma_{cg}^2) \quad (4.16)$$

We could potentially bring in additional information from the CBO's projection of the level of the potential GDP to improve the econometric estimation of the level of potential real GDP. However, as shown in the results section 4.4, the implied estimates of the output gap from our multivariate framework (with or without survey data) are remarkably similar to CBO's estimates suggesting the limited value of bringing additional information from CBO projections.

4.2.5 Productivity block

Since the publication of economist Adam Smith's influential book, *Wealth of Nations*, it is widely acknowledged, among various actors, that long-run productivity growth is the most important contributor to long-run changes in the living standards. Therefore, estimates of the long-run level of (labor) productivity growth (p-star) have received considerable discussion in the past decades and are of great interest. Labor productivity is defined here as output per hour worked.

Furthermore, the estimates of p-star are an important input into policymaking, as they are used to gauge the appropriateness of the monetary policy stance. The importance of p-star for monetary policy is because standard macroeconomic models tightly connect p-star to the long-run level of real interest rate (i.e., r-star). In these models, a lower level of p-star implies a lower level of r-star, and a higher level of p-star implies a higher r-star (see Lunsford, 2017). However, based on post-1960 data, Lunsford (2017) found no statistical evidence supporting the link between p-star and r-star.

Several papers have endeavored to estimate the long-run level of productivity growth using various statistical and econometric models. A subset of those papers has documented support in favor of a regime-switching framework to model long-run labor productivity growth (e.g., Kahn and Rich, 2007). Also, to extract more precise and timely estimates of p-star, various authors (e.g., Kahn and Rich, 2007; Roberts, 2001) have proposed using additional variables alongside labor productivity (e.g., real compensation, real consumption, and average hours worked). On the other hand, Edge et al. (2007) show that estimates of long-run productivity growth obtained

from a simple trend-cycle univariate model solved with Kalman filter does an adequate job of mirroring the long-run projections of productivity growth reported in the survey of professional forecasters and institutional forecasts (e.g., CBO).¹⁷

On closer inspection, the ability of the Kalman filter to echo the predictions of the professional forecasters is not surprising, however. Productivity growth is a notoriously volatile series and is subject to extreme revisions from one vintage to another. So, distinguishing highly persistent fluctuations from truly permanent changes is a difficult job for professionals and models alike. Jacobs and van Norden (2012) discuss in detail some of these challenges when working with productivity data.

The combination of findings in Lunsford (2017), Edge et al. (2007), and Jacobs and van Norden (2012) motivates the formulation of a parsimonious structure for the productivity block relative to other blocks of the model. In particular, we abstract from explicit modeling of direct links between p-star and r-star and between p-star and g-star. But our formulation is still in some ways richer than used in the cited literature, as we allow for time-varying parameters, including stochastic volatility. Specifically, the productivity gap, which is defined as (nonfarm) labor productivity growth¹⁸ (quarterly annualized) less p-star, is modeled as a function of a one quarter lag in the productivity gap and the contemporaneous cyclical unemployment gap.

$$P_t - P_t^* = \rho^p(P_{t-1} - P_{t-1}^*) + \lambda_t^p(U_t - U_t^*) + \varepsilon_t^p, \quad \varepsilon_t^p \sim N(0, e^{h_t^p}) \quad (4.17)$$

where, $|\rho^p| < 1$

The inclusion of the cyclical unemployment gap (or output gap) helps tease out movements in productivity associated with the business cycle. The growth in labor productivity (and more generally aggregate productivity) has been shown to be procyclical to some degree (e.g., Roberts 2001); it typically increases sharply at the onset of recoveries and falls during recessions. However, empirical evidence on the strength and the direction of the cyclical relationship is mixed. This mixed evidence stems from the use of different estimation samples and or cyclical indicators (employment-based or output-based). For instance, Gali and van Rens (2020), using split sample estimation, illustrate empirically the significant weakening in the correlation between labor productivity and employment, especially post-1984. They find that the relationship has become countercyclical in the past three decades when using employment as the cyclical indicator.¹⁹ But, it is slightly procyclical when using output as the cyclical indicator. This latter finding motivates our alternative specification that replaces cyclical unemployment with the

¹⁷It is worth emphasizing, in contrast to the regime-switching model (as used in Kahn and Rich, 2007), which allows for deterministic values of p-star, a random walk assumption for p-star (as in Edge et al., 2007) allows for the possibility that p-star may be changing (slowly) in every period.

¹⁸As discussed in Kahn and Rich (2007), the focus outside of the farm sector is primarily on avoiding short-term transitory volatility in the farm sector that is heavily driven by weather and other non-technological factors.

¹⁹Gali and van Rens (2020) using a structural macro model interpret the weakening procyclicality of labor productivity in part to increased flexibility of the US labor market post-1984, which has enabled firms to make adjustment at the extensive margin quickly and easily in response to shocks.

output gap.

$$P_t - P_t^* = \rho^p(P_{t-1} - P_{t-1}^*) + \phi_t^p(\text{ogap}_t) + \varepsilon_t^p, \quad \varepsilon_t^p \sim N(0, e^{h_t^p}) \quad (4.17b)$$

Gali and van Rens (2020) find weakening in the correlation between labor productivity and the cyclical indicator which motivates time-variation in the coefficients λ^p and ϕ^p .²⁰

$$\lambda_t^p = \lambda_{t-1}^p + \varepsilon_t^{\lambda p}, \quad \varepsilon_t^{\lambda p} \sim N(0, \sigma_{\lambda p}^2) \quad (4.18)$$

$$\phi_t^p = \phi_{t-1}^p + \varepsilon_t^{\phi p}, \quad \varepsilon_t^{\phi p} \sim N(0, \sigma_{\phi p}^2) \quad (4.18b)$$

The variance of the error term ε_t^p is allowed to change over time.²¹ Allowing for the time-variation in the cyclical relationship and the error term would better discriminate the cyclical movements and idiosyncratic movements in productivity from those associated with shifts in p-star.

The SV process is defined as a driftless random walk in the log-variance.

$$h_t^p = h_{t-1}^p + \varepsilon_t^{hp}, \quad \varepsilon_t^{hp} \sim N(0, \sigma_{hp}^2) \quad (4.19)$$

P-star is modeled as a driftless random walk component, and the variance of the shocks to this component is assumed to be constant. By modeling p-star this way allows it to capture both unobserved and observed factors that are thought to be persistent but hard to measure. In particular, one factor is developments in fiscal policy; for example, high levels of government debt in the longer-term tend to crowd out private investment, thereby reducing longer-term productivity growth.

$$P_t^* = P_{t-1}^* + \varepsilon_t^{p*}, \quad \varepsilon_t^{p*} \sim N(0, \sigma_{p*}^2) \quad (4.20)$$

Economic theory posits that the long-run nominal wage inflation equals the sum of long-run productivity growth and long-run price inflation. As discussed later in the wage inflation block, this theoretical restriction defines the law of motion for w-star and constitutes an additional channel influencing the dynamics of p-star.

4.2.6 Price inflation block

We use price inflation as measured by the Personal Consumption Expenditures (PCE) price index, the inflation measure that the Federal Reserve targets. Our formulation for price inflation block closely follows CKP and CCK, combining elements from both of these papers. Specifically,

²⁰Fernald and Wang (2016) documents weakening in the cyclicity of productivity at the industry level, suggesting that results of Gali and van Rens (2020) are not due to changes in the industry composition.

²¹Carriero, et al. (2019) and Tallman and Zaman (2020) document superior forecast accuracy of variables including labor productivity from Bayesian VARs with SV compared to VARs without SV; suggesting the usefulness of allowing for SV. We also find empirical support for the inclusion of SV.

as in CKP, the stationary component, inflation gap (defined as the deviation of inflation from pi-star),²² is modeled as a function of one-quarter lagged inflation gap, unemployment gap, and an error term, whose variance is allowed to vary over time.

The coefficient, ρ^π on the lagged inflation gap, which captures persistence in inflation dynamics (i.e., persistence in deviation of inflation from pi-star), is allowed to vary over time.²³ The motivation for time-variation stems from evidence in Weise (2012), who documents that in the period of Great Inflation, deviations of inflation from the desired target were very persistent because the Fed at the time either was unable or unwilling to enact policy to reduce inflation to its desired level. However, in the 1980s, the Fed was aggressive in returning inflation to the desired target. Modeling time-variation in this parameter allows such contrasting behavior to be captured.

$$\pi_t - \pi_t^* = \rho_t^\pi (\pi_{t-1} - \pi_{t-1}^*) + \lambda_t^\pi (U_t - U_t^*) + \varepsilon_t^\pi, \quad \varepsilon_t^\pi \sim N(0, e^{h_t^\pi}) \quad (4.21)$$

$$\rho_t^\pi = \rho_{t-1}^\pi + \varepsilon_t^{\rho^\pi}, \quad \varepsilon_t^{\rho^\pi} \sim TN(0 - \rho_{t-1}^\pi, 1 - \rho_{t-1}^\pi; 0, \sigma_{\rho^\pi}^2) \quad (4.22)$$

The innovations to the AR(1) coefficient, ρ^π are truncated so that $0 < \rho_t^\pi < 1$, ensuring that the inflation gap (in equation 4.21) is stationary at each point in time t .

$$\lambda_t^\pi = \lambda_{t-1}^\pi + \varepsilon_t^{\lambda^\pi}, \quad \varepsilon_t^{\lambda^\pi} \sim TN(-1 - \lambda_{t-1}^\pi, 0 - \lambda_{t-1}^\pi; 0, \sigma_{\lambda^\pi}^2) \quad (4.23)$$

λ^π is the slope of price Phillips curve and is constrained in the interval $(-1, 0)$.

The parameter λ estimates the price Phillips curve relationship (i.e., the relationship between the inflation gap and the unemployment gap at business cycle frequency). There is ample empirical evidence in support of a time-varying price Phillips curve (e.g., Stella and Stock, 2015; CKP; Del Negro et al., 2020) hence our choice of allowing for time-variation in the parameter λ^π .²⁴

$$h_t^\pi = h_{t-1}^\pi + \varepsilon_t^{h^\pi}, \quad \varepsilon_t^{h^\pi} \sim N(0, \sigma_{h^\pi}^2) \quad (4.24)$$

The SV process is defined as a random walk in the log-variance.

Both the theoretical and empirical literature emphasizes the usefulness of the signal from fluctuations in labor costs for inflation dynamics. Post-Keynesian theory posits that excess wage inflation over labor productivity gains puts upward pressure on price inflation, i.e., causality runs

²²Modeling inflation in gap form, where gap is defined as the difference between inflation and slowly-moving trend, was popularized by Cogley, Primiceri and Sargent (2010), and since then has been a widely used approach to modeling inflation in macroeconomic models for policy and forecasting (e.g., Faust and Wright, 2013).

²³Chan et al. (2013), CKP, and CCK have found strong empirical support for the time-variation in the coefficient of inflation gap. Our results reinforce the empirical importance of allowing for time-variation in this coefficient.

²⁴See Del Negro et al. (2020) for a comprehensive literature review of the instability of the Phillips curve in the US data, and similarly Banbura and Bobeica (2020) for the euro area data.

from labor costs to price inflation. In comparison, the Neo-Classical theory suggests causality runs in the opposite direction, from price inflation to nominal wage inflation. The empirical evidence in the US data is inconclusive in that there is no clear evidence in the direction of the causality. If anything, the evidence suggests they co-move together (see Knotek and Zaman, 2014 and references therein).

Given the empirical evidence of co-movement, we explore an alternative specification in which we allow for a connection between two cyclical inflation components, nominal wage inflation gap and price inflation gap, by adding the nominal wage inflation gap as an explanatory variable in the equation describing price inflation gap. The parameter γ^π captures the strength of the relationship between the two cyclical inflation measures. The expression $\frac{\gamma^\pi}{1-\rho^\pi}$ can be interpreted as the passthrough from cyclical wage inflation to cyclical price inflation.²⁵

$$\pi_t - \pi_t^* = \rho_t^\pi (\pi_{t-1} - \pi_{t-1}^*) + \lambda_t^\pi (U_t - U_t^*) + \gamma^\pi (W_t - W_t^*) + \varepsilon_t^\pi, \quad \varepsilon_t^\pi \sim N(0, e^{h_t^\pi}) \quad (4.21b)$$

Similarly, as shown later, we add the price inflation gap to the equation describing the nominal wage inflation gap.

Pi-star is modeled as a driftless random walk component, and the variance of the shocks to this component is assumed to be constant (as in CKP). This latter assumption of homoscedastic errors is in contrast to Stock and Watson (2007), Mertens (2016), and several others. Our choice not to incorporate SV into shocks to pi-star is made to keep the estimation manageable and maintain consistency with our modeling assumptions for the stars.²⁶

$$\pi_t^* = \pi_{t-1}^* + \varepsilon_t^{\pi*}, \quad \varepsilon_t^{\pi*} \sim N(0, \sigma_{\pi^*}^2) \quad (4.25)$$

Following CCK (and others, such as Mertens and Nason, 2020) to improve pi-star's econometric estimation, we bring in information from the long-run survey expectations (of PCE inflation). An important empirical finding of CCK is that long-run survey expectation of inflation is a biased measure of the underlying trend inflation, at least at some times. Hence, simply equating pi-star with the long-run survey expectation or assuming survey expectation as an unbiased measure of pi-star or calibrating econometric estimates of pi-star to surveys (as is commonly done) may not be a reasonable strategy.

Accordingly, we add an equation linking long-run survey expectations of inflation to pi-star, where the intercept, C_t^π , is time-varying to capture possibly time-varying differential between

²⁵We explored the possibility of allowing for time-variation in γ^π but the estimation ran into difficulties hence we resort to time-invariant γ^π .

²⁶Allowing SV in the inflation gap component and not in the trend component is not without precedent. Besides CKP, Chan (2013) is a recent paper modeling SV only in the measurement equation (i.e., cyclical/transitory component). Berger et al. (2016) find support for SV in the inflation gap component but weak evidence for SV in the trend component. Our preliminary results indicate similar findings that adding SV to the pi-star equation neither helps nor hurts the model fit.

the two.²⁷

$$Z_t^\pi = C_t^\pi + \beta^\pi \pi_t^* + \varepsilon_t^{z\pi}, \quad \varepsilon_t^{z\pi} \sim N(0, \sigma_{z\pi}^2) \quad (4.26)$$

$$C_t^\pi = C_{t-1}^\pi + \varepsilon_t^{c\pi}, \quad \varepsilon_t^{c\pi} \sim N(0, \sigma_{c\pi}^2) \quad (4.27)$$

Our model specification modifies and extends the baseline model of CCK in five important ways. First, we allow a time-varying Phillips curve relationship by adding the cyclical unemployment component (similar to CKP).²⁸ Second, we explore a model specification that allows for a link between the nominal-wage inflation gap and the price-inflation gap (capturing the evolving passthrough from labor costs to price inflation). Third, we adopt a more simplistic approach to modeling the link between survey expectations and pi-star. Fourth, the variance of the shocks to the pi-star process does not entertain SV. Fifth, pi-star is restricted to satisfy the long-run restriction informed by theory (see equation 4.28).

4.2.7 Wage inflation block

The long-run equilibrium level of nominal wage inflation (w-star) is the nominal wage growth rate consistent with its fundamentals – p-star and pi-star. As noted earlier, in the long-run, economic theory posits that the nominal wage inflation equals the sum of the long-run growth rate of labor productivity and the long-run level of price inflation. In other words, in the long-run, labor productivity growth is the only fundamental driver of real wages; therefore, price inflation and nominal wage inflation have to adjust relative to each other to maintain the fundamental relationship. In our setup, we impose this relationship to define w-star.

$$W_t^* = \pi_t^* + P_t^* + \varepsilon_t^{w*}, \quad \varepsilon_t^{w*} \sim N(0, \sigma_{w*}^2) \quad (4.28)$$

The above equation implies that W^* is approximately equal to sum of $\pi_t^* + P_t^*$

In order to assess empirical support for the theoretical restriction defined by equation (4.28), we explore an alternative specification that models W^* as a RW process,

$$W_t^* = W_{t-1}^* + \varepsilon_t^{w*}, \quad \varepsilon_t^{w*} \sim N(0, \sigma_{w*}^2) \quad (4.28b)$$

Equation (4.29) relates the nominal wage inflation gap, defined as the difference between the nominal wage inflation and w-star, to its one-quarter lagged gap, the cyclical unemployment

²⁷Our formulation is flexible but less so than the one adopted by CCK. In addition to time-variation in the intercept, CCK add time-variation in the slope coefficient, and moving average (MA) in the error term. For PCE inflation, CCK did not find the addition of the MA in the error term useful. We think our relatively simplistic equation suffices in its aim of influencing econometric estimation of pi-star through the survey expectations.

²⁸CCK explored an alternative specification in which they add cyclical unemployment but as an exogenous variable (constructed using the CBO estimate of the natural rate). CCK found that adding the cyclical unemployment slightly worsened the fit of their model.

gap, and the price inflation gap.²⁹ The variance of the error term, ε_t^w , is allowed to vary over time. The latter feature is motivated by findings in Tallman and Zaman (2020) and Peneva and Rudd (2017), who document better fit of the VAR with SV to nominal wage data compared to VAR without SV.

$$W_t - W_t^* = \rho_t^w (W_{t-1} - W_{t-1}^*) + \lambda_t^w (U_t - U_t^*) + \kappa_t^w (\pi_t - \pi_t^*) + \varepsilon_t^w, \quad \varepsilon_t^w \sim N(0, e^{h_t^w}) \quad (4.29)$$

The SV process is defined as random walk in the log-variance.

$$h_t^w = h_{t-1}^w + \varepsilon_t^{hw}, \quad \varepsilon_t^{hw} \sim N(0, \sigma_{hw}^2) \quad (4.30)$$

The findings in Knotek and Zaman (2014) motivate the inclusion of a one-quarter lagged nominal wage inflation gap, with time-variation in the parameter ρ^w , the latter quantifies the persistence in wage inflation dynamics.

$$\rho_t^w = \rho_{t-1}^w + \varepsilon_t^{\rho w}, \quad \varepsilon_t^{\rho w} \sim TN(0 - \rho_{t-1}^w, 1 - \rho_{t-1}^w; 0, \sigma_{\rho w}^2) \quad (4.31)$$

The innovations to the AR(1) coefficient, ρ^w are truncated so that $0 < \rho_t^w < 1$, to ensure that the wage gap (in equation 4.29) is stationary at each point in time t .

The parameter λ^w in equation (4.29) measures the strength of the cyclical relationship between the nominal wage gap and labor market slack (aka the wage Phillips curve). Many studies, both theoretical (e.g., Gali, 2011) and empirical (e.g., Knotek and Zaman, 2014; Peneva and Rudd, 2017; Gali and Gambetti, 2019), have documented strong support for the existence of a wage Phillips curve in the US data. These studies have also demonstrated the instability of the wage Phillips curve, motivating the need for time-variation in the parameter λ^w .³⁰

$$\lambda_t^w = \lambda_{t-1}^w + \varepsilon_t^{\lambda w}, \quad \varepsilon_t^{\lambda w} \sim TN(-1 - \lambda_{t-1}^w, 0 - \lambda_{t-1}^w; 0, \sigma_{\lambda w}^2) \quad (4.32)$$

where λ^w is the slope of wage Phillips curve and is constrained in the interval $(-1, 0)$.

As discussed earlier in the price inflation block, both theory and empirical evidence point to the connection between price inflation and nominal wage inflation. The standard fully structural models describing the New Keynesian Phillips curve posit a tight relationship between price and

²⁹Following the literature on modeling price inflation in the gap, Knotek and Zaman (2014) apply a similar transformation to modeling nominal wage inflation for the US. In particular, they construct the nominal wage inflation gap as nominal wage inflation less pi-star, where pi-star is survey expectations. They note the competitive forecasting properties of their model is due to modeling in gaps. Following Knotek and Zaman (2014), Bobeica, Ciccarelli, and Vansteenkiste (2019) construct a nominal wage inflation gap for the euro area data and find empirical support for the gap specification.

³⁰Literature has posited various explanations for the instability of the wage Phillips curve, including downward nominal wage rigidities, where the degree of rigidity varies with the phase of the business cycle (see Daly and Hobijn, 2014).

wage inflation via the channel of current and expected future marginal costs. In these models, price inflation today is a function of expected price inflation and expected future marginal costs, where marginal costs are generally linked to wages. Knotek and Zaman (2014) provide empirical evidence of the connection between nominal wage and price inflation. In particular, they show no clear evidence of one Granger-causing the other; instead, both wage and price inflation generally tend to move together. This reasoning would suggest the importance of modeling the direct relationship between wage inflation and price inflation. Hence, the inclusion of the price inflation gap in the measurement equation (4.29).

Several studies document a significant weakening in the empirical link between price inflation and nominal wage inflation since the 1980s (e.g., Peneva and Rudd, 2017; Knotek and Zaman, 2014), motivating time-variation in the parameter κ^w .³¹ The expression $\frac{\kappa_t^w}{1-\rho_t^w}$ could be interpreted as an estimate of the passthrough from price inflation to wage inflation.

$$\kappa_t^w = \kappa_{t-1}^w + \varepsilon_t^{\kappa w}, \quad \varepsilon_t^{\kappa w} \sim N(0, \sigma_{\kappa w}^2) \quad (4.33)$$

To assess empirical support for the inclusion of the price inflation gap in the equation describing the nominal wage gap, we estimate an alternative model that replaces equation (4.29) with the following,

$$W_t - W_t^* = \rho_t^w(W_{t-1} - W_{t-1}^*) + \lambda_t^w(U_t - U_t^*) + \varepsilon_t^w, \quad \varepsilon_t^w \sim N(0, e^{h_t^w}) \quad (4.29b)$$

4.2.8 Interest rate block

We close our model with the interest-rate block characterizing the interest rate dynamics and the law of motion for r-star (the long-run equilibrium real short-term interest rate).

Our first equation of the block brings information from the nominal short-term interest rate via a Taylor-type rule (TR) to aid in identifying r-star. Past research has shown the TR’s usefulness in characterizing the monetary policy reaction function over the past four decades. Specifically, this equation characterizes the dynamics of the short-term nominal interest rate gap, where the gap is the difference between the nominal short-term interest rate i , and the long-run level of the nominal neutral rate of interest, i-star. (i-star = pi-star + r-star). When modeling the nominal short-term interest rate, especially in a framework like ours, one must account for the effective lower bound (ELB) period.

Recent literature provides at least two options to handle the ELB. The first is to explicitly but separately model the observed short-term nominal rate, which cannot go below zero, and the “shadow interest rate,” which is a hypothetical unobserved and unbounded counterpart. Wu and Xia (2016) popularized the concept of the shadow interest rate, and JM and Gonzalez-Astudillo and Laforte (2020) are two recent approaches well suited for inclusion in UC models. The second approach is to treat the estimate of the “shadow rate” obtained from Wu and Xia

³¹Bobeica et al (2019) find that in the euro area the link between labor compensation and price inflation continues to remain strong post-1980.

(2016) as the measure of the short-term nominal interest rate in measurement equations such as the TR (e.g., Pescatori and Turunen, 2016).³²

Given our model’s size and complexity, we adopt the latter approach, which is simpler though not perfect. Using a direct measure of the nominal shadow rate allows us to capture both conventional and unconventional monetary policy effects when the (observed) nominal federal funds rate is constrained at the ELB.³³

Equation (4.34) relates the nominal interest rate gap (based on the shadow federal funds rate) to its one-period lag interest rate gap, current quarter inflation gap (i.e., deviation of inflation from π -star), and unemployment rate gap (i.e., deviation of unemployment rate from u -star). This equation roughly characterizes the monetary policy reaction function as defined by Taylor (1999).³⁴ There is a broad consensus that policy adjustments outside of cyclical turning points are made very gradually (e.g., Carlstrom and Zaman, 2014). Hence, the inclusion of the lagged interest rate gap term.

Recent research documents strong empirical support for constant parameters in the Taylor rule equation while allowing for stochastic volatility in the errors (see Chan and Eisenstat, 2018a; 2018b; and JM). Accordingly, we allow for SV in the interest-rate equation. JM, Gonzalez-Astudillo and Laforte (2020), and Brand and Mazelius (2019) document the usefulness of adding the TR equation to identify r -star. The latter two do not entertain SV, which JM has found to be empirically important. As discussed later, we also found that adding the TR equation improves the precision of the r -star estimates significantly, and data strongly favors allowing for SV in the error process.

$$i_t - \pi_t^* - r_t^* = \rho^i(i_{t-1} - \pi_{t-1}^* - r_{t-1}^*) + \lambda^i(U_t - U_t^*) + \kappa^i(\pi_t - \pi_t^*) + \varepsilon_t^i, \quad \varepsilon_t^i \sim N(0, e^{h_t^i}) \quad (4.34)$$

where, ρ^i is truncated so that $0 < \rho^i < 1$.

$$h_t^i = h_{t-1}^i + \varepsilon_t^{hi}, \quad \varepsilon_t^{hi} \sim N(0, \sigma_{hi}^2) \quad (4.35)$$

The SV process is defined as random walk in the log-variance.

Our second equation motivated by LW heeds to the economic theory suggesting the role of various real factors in influencing movements in r -star. These factors include long-run output growth (and long-run productivity growth), trend labor force growth (reflecting shifts in demographics and net migration), taxation structure, government expenditure shifts, and shifts in

³²The estimates from Wu and Xia (2016) are publicly available and regularly updated. Treating the shadow rate as the measure of the short-term nominal rate in place of the federal funds rate is commonly done, and often academic papers report results indicating robustness to the use of Wu and Xia (2016) shadow rate (e.g., Beyer and Wieland, 2019; Lewis and Vazquez-Grande, 2019)

³³The nominal *shadow* federal funds rate is identical to the nominal federal funds rate when effective lower bound (ELB) is not binding.

³⁴It is worth emphasizing that we denote this equation as a “Taylor-type rule” and not an exact Taylor-rule because in our equation, π -star refers to the estimate of trend inflation which may or may not equal to central bank’s long-run inflation goal.

liquidity preferences (e.g., Del Negro et al., 2017; Bullard, 2018). Accordingly, equation (4.36) expresses r -star as a linear function of g -star and a “catch-all” component D . In our baseline specification, both g -star and D follow random walk processes similar to LW (and many other papers). The RW assumption for D is an appropriate one, given our focus is the long-run r -star that should, in principle, be influenced over time by permanent shifts in aggregate supply and demand (Laubach and Williams, 2016).³⁵

$$r_t^* = \zeta g_t^* + D_t. \quad (4.36)$$

$$D_t = D_{t-1} + \varepsilon_t^d, \quad \varepsilon_t^d \sim N(0, \sigma_d^2) \quad (4.37)$$

In more recent literature on long-run r -star, modeling r -star as a random walk process has performed better empirically than model specifications relying on the link between g -star and r -star.³⁶ Accordingly, we explore an additional specification that models r -star simply as the RW process (similar to g -star, p -star, π -star, u -star). The RW assumption for r -star implies that we are agnostic about the underlying unobserved forces driving r -star but acknowledge those forces reflect persistent structural shifts in aggregate demand and supply that ought to have a bearing on r -star.

$$r_t^* = r_{t-1}^* + \varepsilon_t^{r*}, \quad \varepsilon_t^{r*} \sim N(0, \sigma_{r*}^2) \quad (4.36b)$$

Lastly, the equations linking implied estimate of the long-run survey expectations of real short-term interest rate to r^* is defined as:³⁷

$$Z_t^r = C_t^r + \beta^r r_t^* + \varepsilon_t^{zr}, \quad \varepsilon_t^{zr} \sim N(0, \sigma_{zr}^2) \quad (4.38)$$

$$C_t^r = C_{t-1}^r + \varepsilon_t^{cr}, \quad \varepsilon_t^{cr} \sim N(0, \sigma_{cr}^2) \quad (4.39)$$

All in all, information from six sources and or elements inform the econometric identification of r -star. These sources include: an IS equation (4.14), TR equation (4.34), which allows for SV; an equation linking r -star to survey expectations; shadow rate; and an equation relating r -star to g -star. As we show shortly, all these sources play a role in improving r -star’s precision.

³⁵Researchers have also explored AR process for component D , which would be consistent if the interest is in medium-term r -star (see Lewis and Vazquez-Grande, 2019), as this would allow r -star to be influenced by the transitory shocks to aggregate demand (via the AR process) and permanent shocks to aggregate supply (via the RW process for g -star). In studies focused on the long-run notion of r -star, such as LW, Laubach and Williams (2016), Clark and Kozicki (2005), and Kiley (2020) specification based on RW assumption has shown empirically to be favored by data compared to AR assumption.

³⁶See Kiley (2020), Gonzalez-Astudillo and Laforte (2020), JM, Orphanides and Williams (2002)

³⁷The r -star survey estimates are not direct estimates instead they are inferred from the Blue Chip survey long-run estimates of GDP deflator and short-term interest rates using the long-run Fisher equation. The survey expectations for r -star goes back to 1983:Q1. Please refer to the supplementary appendix C9 for details on the procedure to back cast estimates all the way back to 1959

To reiterate, in our framework, we utilize information from both short-term nominal interest rates (via a TR equation) and long-term nominal interest rates (via an IS equation) to inform the estimation of r-star.³⁸

4.2.9 Base model and its variants

The equations (4.6), (4.7)...(4.39) defines our baseline model formulation (denoted *Base*). Figure 4.1 provides a visual representation of our Base model. And section C1.a. of the supplementary appendix lists all the equations for the Base model for easy reference. To assess the usefulness of survey information in the econometric estimation of our multivariate UC model, we also estimate a variant of the baseline model that excludes the equations linking long-run survey expectations to stars (i.e., excluding equations 4.9, 4.10, 4.15, 4.16, 4.26, 4.27, 4.38, and 4.39). We denote the latter specification as *Base-NoSurv*. The model specifications Base and Base-NoSurv constitute our two main model specifications. To assess the empirical support for numerous additional features (informed from theory and past empirical research) embedded in our modeling framework, we formulate several additional model specifications, each of which is a restricted variant of the Base. For instance, to assess the empirical support of the theoretical restriction defined by equation 4.28 (which defines w-star as the sum of pi-star and p-star), we estimate a variant of the baseline model that replaces equation 4.28 with a random walk assumption for w-star as defined by the equation 4.28b. We denote this specification as *Base-W*RW*.

Similarly, to assess the empirical support for the theoretical restriction defined by equation 4.36 (the link between gstar and rstar), we estimate a model specification that replaces equation 4.36 with a random walk assumption for r-star as defined by the equation 4.36b. Table 4.1 reports the description of model specifications that are formulated to assess various features of importance. We ran few additional specifications to explore the role of priors in influencing r-star, and for brevity purposes they are relegated to the appendix. To keep the length of chapter manageable, we report selected results from the auxiliary model specifications in the main part of the chapter with additional results included in the supplementary appendix.

4.3 Data and Bayesian Estimation

4.3.1 Data

We estimate the empirical model using the following quarterly data: (1) the unemployment rate; (2) Real GDP growth; (3) Nonfarm labor productivity growth; (4) inflation rate in Personal Consumption Expenditures (PCE) price index; (5) Average Hourly Earnings (AHE) of

³⁸Del Negro et al. (2017), JM, Bauer and Rudebusch (2020), Gonzalez-Astudillo and Laforte (2020) are recent studies that have highlighted the usefulness of exploiting information from both short-term and long-term interest rates in the identification of r-star.

production and nonsupervisory workers (total private industries);³⁹ (6) Federal funds rate; (7) Nominal yield on 10-year Treasury Bond; (8) Shadow federal funds rate from Wu and Xia (2016); (9) Blue Chip (real-time) long-run projections of 3-month Treasury Bill, real output growth, the unemployment rate, and GDP deflator inflation; (10) Long-run inflation expectations of PCE inflation (PTR series). We also collect the real-time long-run CBO projections of real output growth, level of real potential output, and the natural rate of unemployment. For forecast evaluation exercises, the real-time data vintages of real GDP growth, PCE inflation, the unemployment rate, AHE, and nonfarm labor productivity spanning 1998Q1 through 2019Q4 are downloaded from the ALFRED database maintained by the St. Louis Fed and the real-time database maintained by the Federal Reserve Bank of Philadelphia. For the data series labeled (1) through (7), which comprises our core dataset, we collect two vintages of revised data: 2020Q2 and 2020Q4 vintages, respectively. We use data starting 1959Q4 through 2019Q4 from the 2020Q2 vintage (which includes the third estimate of 2019Q4) as a featured sample for this chapter. To show the implications of the COVID-19 data on our model estimates, we estimate our model(s) using the 2020Q4 vintage, which has data spanning from 1959Q4 through 2020Q3. The vintages corresponding to the revised data are downloaded from Haver Analytics.

4.3.2 Bayesian estimation

We use Bayesian estimation methods to fit our Base model and its variants. The use of inequality restrictions on latent parameters in our model(s) setup leads to a non-linear state-space model, which renders estimation using standard Kalman filter methods infeasible. Accordingly, we implement our Markov chain Monte Carlo (MCMC) posterior sampler based on computational methods developed in Chan, Koop, and Potter (2013) and CKP, which uses the band and sparse matrix algorithms detailed in Chan and Jeliazkov (2009). Specifically, the MCMC posterior sampler is a significantly scaled-up version of the sampler employed by CKP. The CKP posterior sample developed for a relatively smaller-scale nonlinear state-space model is carefully extended to accommodate the additional structure and numerous features of our model(s). Since the computational methods used in this chapter are based on CKP, we relegate the specific details of the sampler to the supplementary appendix C1.

In a methodological sense, this chapter’s novelty is in assembling the existing sampling algorithms based on the fast band and sparse matrix routines to solve a large nonlinear and a high-dimensional UC model. We found that the use of inequality restrictions such as bounds on the u -star and other parameters is crucial to estimate the model, especially in the Base-NoSurv model. Intuitively, features such as truncated distributions that we implement for some of the time-varying parameters, e.g., the Phillips curve (price and wage), persistence, and bounds on u -star facilitate estimation by guiding the estimation procedure to the credible regions of the

³⁹Average Hourly Earnings (AHE) of production and nonsupervisory workers in total private industries goes back to 1964Q1. From 1959Q4 through 1963Q4, we use the AHE of production and nonsupervisory workers in goods producing industries. We splice them together.

parameter space.

For each model, we simulate 1 million posterior draws from the MCMC posterior sampler. We then discard the first 500,000 draws, and of the remaining, we keep every 100th draw. Accordingly, all the reported results for the Base model and its variants are based on 5000 retained draws.

As emphasized by Chan (2017), the MCMC algorithm is considered efficient if the draws it produces have low autocorrelation – they are autocorrelated by construction – and the time it takes to sample a given number of posterior draws is reasonable. In the appendix (see section C3), we report efficiency diagnostics of our MCMC algorithm. Those diagnostics, which include inefficiency factors and convergence metrics, indicate good convergence properties (and low autocorrelation) of our sampler for both Base and Base-NoSurv models.⁴⁰

Bayesian model comparison is based on the marginal likelihood metric. In computing marginal likelihood for various models, we use the approach proposed by CCK, which decomposes the marginal density of the data (e.g., inflation) into the product of predictive likelihoods; see appendix C1.d for details.⁴¹ Our preference to use the CCK approach is because it allows us to separately compute marginal data density for each variable of interest: inflation, nominal wages, interest rate, real GDP, the unemployment rate, and labor productivity. The variable specific marginal densities prove useful for us because it allows for deeper insights about the source of the deficiencies, which helps differentiate models at a more granular level.

We note that our prior settings are similar to those used in CKP, CCK, Gonzalez-Astudillo and Laforte (2020). As discussed in CCK, UC models with several unobserved variables, such as the one developed in this chapter, require informative priors. That said, our priors settings for most variables are only slightly informative. The use of inequality restrictions on some parameters such as the Phillips curve, persistence, bounds on u-star could be viewed as additional sources of information that eliminates the need for tight priors, something also noted by CKP. The parameters for which there is a strong agreement in the empirical literature on their values, such as the Taylor-rule equation parameters, we use relatively tight priors, such that prior distributions are centered on prior means with small variance. In model comparison exercises, the priors are kept the same for the common parameters across models. We also perform some prior sensitivity analysis reported in the appendix C2.

⁴⁰Regarding computational time, given the high-dimensionality of our model(s) and the number of posterior simulations we require, the speed is quite fast (in our assessment). When applied to the Base model, the MCMC algorithm, which is implemented in Matlab, takes about 350 seconds to generate 10,000 posterior draws using a laptop computer with an Intel(R) Xeon(R) E-2176M CPU @ 2.70 GHz processor. To generate 1 million posterior draws, it takes less than 10 hours.

⁴¹Alternatively, one could use the cross-entropy (Kullback-Leibler divergence) approach proposed by Chan and Eisenstat (2018a,b). We leave this for future extension.

4.4 Full Sample Estimation Results

This section sequentially discusses results for each of the six blocks, particularly the full-sample estimates of stars. Here, we briefly highlight some of the noteworthy findings. Comparing estimates of the stars between model specification Base and Base-NoSurv, a clear pattern emerges. The precision of the estimates, as measured by the width of the 90% credible intervals, indicates that the Base model, which explicitly accounts for the links between stars and survey data, yields more precise estimates of the stars than Base-NoSurv. Although the broader contours seen in the estimates of stars from our two main model specifications, Base and Base-NoSurv, are comparable, at times, the differences can be notable, especially in the case of r-star. In the case of g-star (and in-turn the output gap), our two model specifications yield very similar trajectories as implied by the posterior mean estimates. This result suggests that data is very informative about g-star, and not so much in the case of r-star, confirming Kiley’s finding (2020).

Furthermore, the Bayesian model comparison indicates marginally higher support in data for Base over Base-NoSurv, as shown in table 4.3. The breakdown of the marginal data density by variables suggests that the Base model’s improved fit over Base-NoSurv is due to its improved fit to inflation and nominal interest rate data. But that improved fit is offset mainly by worsening fit to the unemployment and wage data. The table also reports the marginal data density for two additional Base model variants that the Bayesian model comparison indicates a comparable fit to the data. Overall, we find that bringing additional information from surveys – by directly modeling the connection to the stars – leads to more reasonable values, provides more precise estimates, and marginally improves the model fit.

4.4.1 Estimation results for u-star

Figure 4.2 plots the evolution of u-star (and its uncertainty) covering the period 1960 through 2019. Panel (a) plots the posterior estimates from the Base model and panel (b) from the Base-NoSurv model. Also plotted are the corresponding 90% credible intervals. Both models imply smooth evolution of u-star. In the past six decades, the (posterior mean of) u-star has fluctuated between 4.4% and 5.7%, peaking in the early-1980s and early-2010s; and troughs in the late-1990s and at the end of our sample period. The contours of u-star from both models are generally similar (as can also be seen conveniently in panel d); however, the level of u-star can differ notably in some periods. From 1960 through the late-1970s, the u-star has gradually increased (from 5.4% to 5.7% in Base and 5.0% to 5.6% in Base-NoSurv). But, since the mid-1980s through the late-1990s, u-star has steadily drifted lower (to 4.5%). This downward trend in the later period is also documented in the u-star literature based on job-flows data, which attributes the decline in u-star to declining trends in job-separation and job-finding rates (e.g., Crump et al., 2019; Tasci, 2012).

From early-2000 through early 2010, the u-star has trended higher, with a sharp pickup

during the Great Recession period. Since 2010, u-star has steadily drifted lower. By the end of 2019, Base has u-star at 4.4% (with a 90% interval covering 3.7% to 5.2%) and Base-NoSurv at 4.5% (with a 90% interval covering 3.7% to 5.4%). As shown in figure 4.3, at the end of 2019, the unemployment gap implied by both models is negative, i.e., the unemployment rate is below the estimated u-star.

The use of survey information in the Base model mainly contributes to the difference in the levels of the u-star across the two models. To facilitate comparison, panel (a) also plots u-star from the survey. As is evident from the plot, u-star from the survey displays more pronounced shifts in u-star than the model-based estimates. However, due to a strong estimated relationship between the survey u-star and Base u-star (i.e., posterior mean of $\beta^u = 0.988$; see table 4.2), the Base estimate of u-star reflects the contours in survey u-star. As can be seen in panel (e), which plots the precision (measured as the width of the 90% intervals), taking onboard survey information improves the precision of u-star notably (comparing Base and Base-NoSurv).

Surprisingly, based on the Bayesian model comparison, the Base model has an inferior fit to the unemployment data compared to Base-NoSurv. As shown later, this result contrasts with the results for pi-star, r-star, and g-star, for which survey information helps improve the model fit or at least does not worsen the fit.⁴²

Sensitivity of u-star to modeling assumptions including information set

Panel (c) plots additional estimates of u-star (posterior mean) from the variants of the Base model to highlight the sensitivity of u-star to modeling assumptions and the informational aspect of joint modeling. The plot denoted Base-NoBoundU* represents the Base model variant that eliminates the bound on the random walk process describing the u-star. Doing so has a trivial effect on the estimates of u-star, the precision of u-star, and model fit. Comparing between panels (a) and (c), the posterior mean estimate of u-star is quite similar across Base and Base-NoBoundU*. Similarly, there is little change in the u-star estimate's precision across the two models, with Base only marginally better in the latter part of the sample (as shown in panel e). Not surprisingly, the Bayesian model comparison suggests equal support for both Base and Base-NoBoundU*.

We highlight two noteworthy comments in regards to the implementation of bounds on u-star. First, the trivial difference in the estimates between Base and Base-NoBoundU* is because the bounds defined on u-star are wide. Put differently, the values of the bound we have set are not binding on the Base model. Second, we find that using bounds on u-star is extremely important in the Base-NoSurv, as it helps keep the estimation tractable. In other words, the advantages of using bounds on the random walk processes stressed in CKP were in full display in the estimation of Base-NoSurv. Hence, our preference to keep bounds on u-star in our main models, including Base.

⁴²Wright (2013) and Tallman and Zaman (2020) use long run survey information in an attempt to improve VAR model forecasts. In both cases, survey information did not improve unemployment rate forecast accuracy, although they found survey forecasts useful in improving accuracy for a range of other macroeconomic variables.

When writing this chapter, we got hit with a COVID-19 shock, an extreme and an unprecedented global health shock along various dimensions, leading several analysts to call it a “once-in-a-lifetime upheaval.” As we will show in section 4.6, the implementation of bounds on u -star is part of the story in preventing our models from blowing up in response to COVID-19 data.

The other model variants plotted in panel (c) are all nested specifications of the Base model: the Bivariate model of GDP and the unemployment rate (is a Base model that excludes survey information and everything else except the equations describing the dynamics of GDP and unemployment rate); Bivariate+Surv, which is Bivariate but adds survey data for GDP and unemployment; and CKP Adjusted, which is a bivariate model of inflation and the unemployment rate as in CKP but with no bounds on π -star. For visual reasons (to limit the number of plots), u -star from Base in panel (c) is not shown. However, for the sake of discussion, we could treat the plot representing Base-NoBoundU* as the estimate for the Base model, since they are identical to each other (as discussed in the preceding paragraph).

These plots show that different model specifications could provide very different signals about the level of long-run unemployment, indicating the sensitivity of u -star to modeling assumptions. As evident from the figure, the small-scale model specifications indicate u -star range-bound between 5.3% and 6.5% over the sample. In contrast, the Base specification has u -star fluctuating over a broader range. A model specification that infers the estimate of u -star from inflation and unemployment data only, i.e., the price Phillips curve (CKP Adj. model), has a higher trajectory of u -star compared to Base. The story is similar in the case of the model specification that infers the estimate of u -star from GDP and the unemployment data only, i.e., the Okun’s law relationship (Bivariate model), though the trajectory of u -star is lower than implied by the CKP-Adj model. Once the Bivariate model is augmented with survey data for GDP and unemployment, the trajectory is revised higher to resemble CKP-Adj, but with contours similar to Base (because of the survey data).

Panel (d) compares the u -star estimates from our main model specifications with the CBO estimate of the long-run unemployment rate. Interestingly, except for the 2000-2007 period, the CBO u -star’s contour is similar to our model-based estimates, though the level of CBO u -star estimate is significantly higher from 1960 through the early-2000s. From 2000 to 2007, both Base and Base-NoSurv indicate a steadily rising u -star, whereas CBO has u -star trending lower. At the onset of the Great Recession and through the early phase of the economic recovery, all three have u -star continuing to move higher. Whereas Base and Base-NoSurv peak in late 2010 at 5.5% and 5.2%, respectively, CBO has u -star peaking in late 2011 at 5.8%. Since then, u -star has steadily moved lower, with the pace of decline quite similar across CBO and Base. CBO has u -star at the end of 2019 at 4.4%, identical to Base and just a tenth shy of Base-NoSurv.

Precision of u -star

The panel (e) plots the precision of u -star estimates for Base, Base-NoSurv, Base-NoBoundU*,

and Bivariate+Surv. The plots indicate couple of observations. First, comparing between Base and Bivariate+Surv, both these model specifications use information from the surveys and the Okuns relationship. However, Base relies on greater information and additional structure (e.g., wage Phillips curve, price Phillips curve, cyclical productivity, monetary policy stance via the Taylor-type policy rule) to infer the u-star compared to Bivariate+Surv, hence the more improved precision of the resulting u-star. The latter reasoning contributed to the very different estimate of u-star from the Base than Bivariate+Surv discussed earlier (and shown in panel c). And the model comparison indicates a significantly higher fit of the Base model to the unemployment data compared to the Bivariate+Surv.

Second, comparing with Base and Base-NoSurv, additional information from survey forecasts improve the precision of u-star, but this improved precision does not translate into the improved model fit, which worsens somewhat (as shown in table 4.4).

Panel (f) shows the precision of u-star for Base-NoSurv, Bivariate (GDP and unemployment), and CKP-Adj (which is bivariate model of price inflation and unemployment rate). As indicated earlier, the u-star from Base-NoSurv is inferred from a broader information set and structure than the other two small-scale models. Accordingly, Base-NoSurv estimate of u-star is, for the most part, more precise and the model comparison indicates a substantially higher fit to the unemployment data than the other two. The plots also show that u-star inferred from the Okun's law relationship (i.e., Bivariate model) is less precise than inferred from the price Phillips curve (i.e., CKP-Adj model). In contrast, the Bayesian model comparison lends support to the Bivariate model over the CKP-Adj model.

The results also provide evidence that adding survey data to the Bivariate model (Bivariate+Surv) improves further both the precision of the u-star (comparing Bivariate and Bivariate+Surv in panels e and f) and the fit to the unemployment data (Bivariate: -56.5 vs. Bivariate+Surv: -46.5, as shown in table 4.4). This latter finding of improved fit from adding survey data is interesting because in the case of Base, adding survey data worsens the model fit (Base: -24.6 vs. Base-NoSurv: -21.7). Importantly, it suggests that survey forecasts of u-star are likely useful in the case of parsimonious models but of limited use for models that already are utilizing various sources of information to infer u-star.

Cyclical unemployment

Figure 4.3 presents the posterior mean estimate of the unemployment gap (i.e., the cyclical component of the unemployment rate) and the corresponding 90% credible intervals. The top panel plots the estimates from the Base model, and the bottom panel from the Base-NoSurv model. A visual inspection indicates that the movements in the cyclical unemployment correspond quite well with the NBER's business cycle dating. For instance, cyclical unemployment falls in economic expansions and rises during recessions. Both models show a significant spike in the cyclical unemployment rate in the 1982-83 and 2007-09 recessions. And a sharper recovery following the 1982-83 recession but a more gradual recovery following the Great Recession. The

figure also highlights that both models produce similar estimates of cyclical unemployment. Comparing estimates of (smoothly evolving) u -star in the previous figure to estimates of the cyclical unemployment rate indicates that fluctuations in the observed unemployment rate are attributed mainly to cyclical unemployment.

4.4.2 Estimation results for g -star and the output gap

The panels (a) and (b) in figure 4.4 plot the g -star estimates from several sources, and panel (c) plots the corresponding precision. As is evident, g -star estimates from all sources shown indicate a steady decline throughout the sample, except a temporary rise in the late 1990s, which literature has attributed to the technology boom. According to the posterior mean estimates of g -star from our Base and Base-NoSurv models, the growth rate of potential output has continuously drifted lower from an annualized rate of close to 4.5% in early 1960 to 1.4% by the end of 2019. The estimate of g -star fell to 1.2% in 2012 and remained there through 2015 and then began very slowly to move up. The story is generally similar based on the inference from simpler (nested) specifications of the Base model: (1) univariate model (which is Chan and Grant 2017 model); (2) the bivariate model of real GDP and unemployment rate; (3) and the bivariate model augmented to include survey data for g -star and u -star (denoted Bivariate+Surv).

This continuous reduction in the growth rate of potential output has been extensively documented elsewhere (e.g., Berger et al., 2016; Chan and Grant, 2017; Coibion et al., 2018) and in particular the decline since 2009 has been of great concern among the policymakers. Several other researchers including, Summers (2014), Eggertson, Mehrotra, and Summers (2016), Pescatori and Turunen (2016), LW (2016), Antolin-Diaz, Drechsel, and Petrella (2017) have also documented the secular decline in g -star over the past two decades.

Not surprisingly, the precision of the g -star estimates (and of the output gap) display patterns that align well with intuition. For instance, model specifications that incorporate survey expectations (i.e., Base and Bivariate + Surv) yield more precise estimates than specifications that ignore survey data. As discussed earlier, previous researchers have shown unemployment data to be the most critical indicator for the estimation of g -star and the output gap, and the plots provide evidence to that effect. For example, the model specification Bivariate, which builds on the univariate GDP model by adding the unemployment rate, yields a substantial improvement in the precision of the g -star estimate. Interestingly, the g -star estimate's precision from Bivariate is about the same as the Base-NoSurv, suggesting that unemployment is the most crucial variable influencing estimation of g -star. However, as we show shortly, model comparison results indicate support for Base-NoSurv over Bivariate in the output gap case.

Output gap estimates

Next, we examine the estimates of the output gap. In figure 4.4, panels (d) and (e) plot the output gap estimates from the same sources as in the case of g -star, and panel (f) plots the

corresponding precision. Overall, the estimates of the output gap from Base and Base-NoSurv are quite similar. They accord well with the NBER recession dates. Furthermore, as noted by Morley and Piger (2012) and Johannsen and Mertens (2015), our model-based estimates provide evidence of asymmetry in that recessions are shorter in duration but deeper than expansions in the US. It is instructive to highlight that estimates imply a more negative output gap (of -10.5% ; posterior mean) during the 1981-82 recession compared to the Great Recession period (-7%) when output fell more dramatically. At a first pass, this may seem odd. But a closer inspection reveals that in comparison to the 1981-82 recession during the Great Recession, g -star fell significantly (as can be seen in panel a), resulting in a smaller negative output gap; in contrast, during the 1981-82 recession, g -star is estimated to have remained stable.

Bivariate models are consistent with a similar story, but there are notable differences in the estimates implied from a univariate model. For instance, the latter model suggests a less dramatic fall in the output gap in 1973-74, 1981-82, and 2007-09 recessions. In the mild 2001 recession, the univariate model estimates a positive output gap, although less positive than before the recession. Also, as of 2019, according to this model, the output gap remains negative, which is in sharp contrast to other models and the consensus view (e.g., CBO output gap).

The precision estimates and the Bayesian model comparison indicate the inferior quality of the univariate model's output gap estimate. Table 4.5 reports the assessment of model fit to the GDP data for the various model specifications discussed in this section. As in the case of g -star, the output gap estimate from the univariate model is subject to a great deal of uncertainty. The bivariate model, that brings additional information from the unemployment rate helps improve the precision significantly. This improved precision is also reflected in the bivariate model's substantially improved fit to GDP data compared to the univariate model (Bivariate: -280.4 vs. Univariate: -296.5). In our subjective assessment, the estimate of the bivariate model's output gap is more reasonable because its trajectory aligns well with Base and Base-NoSurv, and as shown shortly, with the consensus view.

Further improvements in precision are realized by bringing in additional information from the surveys (comparing Bivariate vs. Bivariate+Surv); however, the model fit, which reflects uncertainty about other model parameters in addition to g -star and the output gap, is little changed – slightly deteriorates. The usefulness of survey data in improving the quality of the g -star and output gap estimates is also apparent comparing precision between Base and Base-NoSurv; however, model fit is nearly similar. Contributing to comparable model-fit between the two models is that beginning 1990 onwards, the precision of output gap across the two models is identical. Unlike, in the case of g -star, the Base-NoSurv yields significantly more precise estimates of output gap than Bivariate, suggesting the usefulness of additional information and structure embedded in Base-NoSurv; hence, on-net the slightly better fit of the Base-NoSurv model to GDP data compared to Bivariate (Base-NoSurv: -279.1 vs. Bivariate: -280.4)

Posterior parameter estimates output block

Next, we discuss the Base model’s parameter estimates of the output block that drive the dynamics of g-star and the output gap. The posterior mean estimates of parameters ρ_1^g and ρ_2^g indicate a high degree of persistence ($\rho_1^g + \rho_2^g = 0.74$) and suggest a hump-shaped response of output gap to shocks (as $\rho_1^g > 1$). These parameters are precisely estimated as evidenced by tight posterior credible intervals. The posterior estimate of parameter λ^g (the coefficient on the unemployment gap in the output gap equation) is negative and highly significant statistically. The estimated posterior mean of λ^g is -0.46 (with 90% interval -0.58 to -0.34). Similarly, the parameter ϕ^u (the coefficient on output gap in the unemployment equation), discussed earlier, is also negative and highly significant statistically. Together, these estimates indicate a strong Okun’s law relationship in the data. The implied posterior mean estimate of the Okun’s law coefficient, $\frac{(1-\rho_1^u-\rho_2^u)}{\phi_u}$ is -2.1 , with 90% credible intervals spanning -2.3 to -1.8 . This estimated coefficient is strikingly identical to the conventional estimate often discussed in macroeconomic textbooks. Therefore, not surprisingly, both the estimated output gap and unemployment gap (shown earlier) reveal similar cyclical dynamics. For instance, according to both cyclical measures, the 1981-82 recession is estimated to have been deeper than the Great Recession.

The parameter a^r , which relates the output gap to the real rate gap (characterizing the IS relation), is negative and much smaller than the prior mean. The estimated posterior mean of a^r is -0.07 (with 90% interval -0.14 to -0.00).

The posterior mean estimate of $E(\sigma_{gdp*}^2)$, the variance parameter of the innovations to the process governing the evolution of the g-star (and gdp-star), is nearly identical across the Base and Base-NoSurv models: 0.02^2 . For comparison, the prior mean $E(\sigma_{gdp*}^2)$ is 0.01^2 . Similarly, for the output gap, the posterior mean estimate of $E(\sigma_{ogap}^2)$, the variance parameter of the innovations to the IS equation, is also identical across the two models: 0.72^2 (compared with a prior of 1^2).⁴³ The estimation results suggest that data is quite informative in influencing the dynamics of both output gap and g-star, confirming Kiley (2020).

Model-based estimates of output gap vs. CBO and others

Figure 4.5 presents estimates from outside sources, CBO, and LW model to gauge how our model-based output gap estimates compare to measures from other sources. Also plotted are estimates from Base and Base-NoSurv in panel (a) and from the Bivariate and Bivariate+Surv models in panel (b) to facilitate comparison with the outside estimates. A few observations immediately stand out. First, the output gap estimate implied from the LW model is notably different over most of the sample period. In particular, during the Great Recession period, the output gap from LW turned slightly negative, while other estimates implied larger negative gaps. The slight negative gap in the LW model is the result of the LW model estimating a

⁴³Note, our prior setting implying a high ratio of $\frac{\sigma_{ogap}^2}{\sigma_{gdp*}^2}$ is consistent with the high noise to signal ratio suggested in Kamber et al. (2016). Specifically, they recommend fixing the noise-to-signal ratio to a high value to obtain large and more persistent cycles when estimating output using Morley et al. (2003) UC model with maximum likelihood estimation.

dramatic fall in potential output, in line with the collapse in the actual output.

Second, both Base and Base-NoSurv models produce estimates of the output gap generally similar to the CBO estimate. This close similarity is notable because the CBO approach, which uses a production function approach (see Shackleton, 2018), differs from our multivariate modeling approach.⁴⁴ Note that the Bivariate models of real GDP and the unemployment rate (with and without survey), which are nested specifications of our bigger Base model, produce estimates of the output gap broadly similar, though not identical, to our two main model specifications. Interestingly, in periods when the output gap estimates from our Base (and Base-NoSurv) differ from the CBO estimates, the Bivariate models' estimates are identical to the CBO's. And in periods when Bivariate models differ from CBO, the Base (and Base-NoSurv) estimates are identical to CBO's.

The estimation results suggest an important takeaway: that joint modeling of real GDP and the unemployment rate is the key to obtaining credible output gap estimates. Morley and Wong (2020), who estimate the output gap using a large BVAR, also found that the unemployment rate is the most crucial indicator for the output gap. (In section 4.6, we compare our model estimates with additional estimates, including Morley and Wong's). Recently, Barbarino et al. (2020) use a range of small-scale UC models to estimate the output gap and similarly find that the unemployment rate is the most valuable indicator.

Our chapter's result indicating a close resemblance of our models' output gap estimates to the CBO's output gap provides evidence supporting the common practice of using output gap estimates from the CBO as an exogenous variable in empirical macroeconomic models (e.g., JM; Stock and Watson, 2020). We view this result as a useful contribution to the applied macroeconomics literature.

On the one hand, the fact that model-based estimates of the output gap bear a strong resemblance to institutional forecasts, i.e., CBO, are encouraging and lends credibility to our model(s). However, on the other hand, in light of the evidence reported in Coibion et al. (2018), the strong resemblance to outside estimates is an unfortunate outcome. We say this for the following reason. Coibion et al. (2018) examine estimates of potential output taken from a variety of model-based and external sources, including CBO and survey forecasts, and based on a range of shock measures, find that (in real-time) the estimates of potential output are unable to distinguish between transitory and permanent shocks effectively. Put differently, they find that their estimates of potential output respond "gradually and similarly" to both supply shocks and demand shocks that drive cyclical fluctuations in real GDP. This is unfortunate, since, by definition, potential output (and g-star) should only adjust in response to permanent shocks.

Coibion et al. (2018) in their conclusion postulate whether a framework that jointly estimates the dynamics of potential output with other relevant stars (as theory would imply) better distinguish between permanent and transitory components and hence lead to more credible es-

⁴⁴We note that in recent years, more and more papers using UC models, which jointly models real GDP growth and unemployment rate, yield estimates of output gap similar to CBO output gap (e.g., Johannsen and Mertens, 2015; Berger et al., 2016; Kiley, 2020; Gonzalez-Astudillo and Laforte, 2020).

timates of potential output. Unfortunately, our model estimates suggest otherwise.

4.4.3 Estimation results for p-star

Figure 4.6 presents posterior estimates of p-star and other parameters of the productivity block. The panels (a), (b), and (c) present p-star estimates from the Base, Base-NoSurv, and Base-W*RW models, respectively. Both the posterior mean and the 90% credible intervals are shown. Also plotted is the actual labor productivity series. A visual inspection of the actual series indicates the unusually high volatility of the quarterly productivity data. In addition, this series is subject to a high-degree of revisions in subsequent data vintages, suggesting the extreme difficulties of its measurement in real-time (see Jacobs and van Norden, 2016). Perhaps, a quote from the former Chair of the Federal Reserve, Alan Greenspan (courtesy of Jacobs and van Norden, 2016), would be instructive to reflect a general sentiment about the productivity data, *“The productivity numbers are very rough estimates because we are measuring a whole set of production outputs from one set of data and a whole set of labor inputs from a different set. That they come out even remotely measuring actual labor productivity is open to question...”* (Transcript: Meeting of the Federal Open Market Committee, March 25, 1998, p. 96)

Not surprisingly, researchers have emphasized that these difficulties of extreme volatility, extensive revisions, and real-time measurement issues with productivity data complicate its trend-cycle decomposition (e.g., Edge, Laubach and Williams, 2007; Kahn and Rich, 2007). Our model-based estimates reflect these challenges. For instance, the estimate of the parameter ρ^p , reported in table 4.2, indicates close to zero persistence in the labor productivity data, defined as the difference between the growth rate in labor productivity and p-star. Similarly, our estimation indicates that labor productivity data has very little influence on the estimate of p-star. Put differently, the data is so volatile to allow for a meaningful identification of trend in the productivity data. The posterior mean of $E(\sigma_{p^*}^2)$, the variance of the shock process for p-star, is essentially the same as the prior mean.⁴⁵ As a result, the degree of time-variation in p-star is primarily influenced by the prior setting. So conditional on our prior belief, which allows p-star to evolve slowly from one quarter to the next, we find considerable evidence of gradual time-variation in p-star over the post-war sample. The evidence of time-variation is economically significant and is consistent with the findings of Roberts (2001), Benati (2007), Edge, Laubach and Williams (2007), and Fernald (2007).

Comparing across the three top panels in the figure, the paths of p-star (posterior mean estimates) from Base and Base-NoSurv are lower than Base-W*RW. The latter model removes the restriction that the long-run w-star grows at a rate equal to the sum of pi-star and p-star. So removing this restriction eliminates the direct influence on p-star from wages and prices and helps raise the level of p-star. As evident from the Bayesian model comparison reported in table

⁴⁵We tried different values for the prior mean on this parameter, and found that posterior moves with the prior.

4.6, the elimination of this restriction improves the fit of the model to the productivity data but reduces the overall model fit to other data, particularly interest rates – via the changes in π -star and u -star in the Taylor-type rule equation. That said, all three models generally indicate similar broad patterns in p -star.

After averaging between 2% and 3% in the 1960s, the models indicate that p -star experienced a sharp deceleration in the 1970s through mid-1980s, mirroring the dramatic fall in productivity growth. Both Base and Base-NoSurv estimates show p -star trending lower from 2.4% (2.3%) in early 1970 to 0.5% by mid-1980, whereas Base-W*RW has it falling close to 1.5%, with wide 90% credible intervals that range from 0.4% to 2.3%. From there on through to the late 1990s, p -star increased sharply, at a pace roughly equivalent to its deceleration in prior periods, to reach a level of 2.0 to 2.4% by 1999. The literature attributes part of this acceleration in the latter half of the 1990s to the information technology boom. Roberts (2001), Edge, Laubach, and Williams (2007), and Benati (2007) document estimates of trend productivity generally similar to the p -star implied from the Base-W*RW model.⁴⁶

In the 2000s, the models have p -star gradually declining to a level close to 1.0%-1.2% by 2010. It remained close to that level through most of the past decade, but since 2018, it has steadily increased. At the end of our sample, all three models estimate the posterior mean of p -star at, or close to, 1.5%. As we show in appendix C12, these estimates of p -star are consistent with the narrative implied by the two-regime Markov-switching model of Kahn and Rich (2007), an influential contribution in the trend productivity literature.

The uncertainty around the posterior mean estimates of p -star is large. Panel (d) quantifies this uncertainty by reporting the width of the 90% credible intervals corresponding to all three models. The plots provide evidence that the theoretical restriction (defined by eq. 4.28) contributes to a substantially improved precision of p -star (just like it does for π -star and w -star), as evident by Base and Base-NoSurv plots lying below the Base-W*RW. Interestingly, the estimate of p -star from the Base is more precise than Base-NoSurv in 1960 through the mid-1980s, even though we do not utilize survey based long-run expectations of productivity in the estimation. Based on the Bayesian model comparison, the Base model's fit to productivity data is marginally better than Base-NoSurv. The improved precision and better fit of the Base model to productivity data suggest there are important spillover effects in the estimation from survey forecasts of other stars.

Cyclical dynamics of labor productivity

Panel (e) plots the estimate of the parameter λ^p , which relates cyclical unemployment to the productivity gap, from all three models. The plots indicate a high level of uncertainty around the estimates of λ^p . The 90% credible intervals are wide, such that they include both positive and negative values complicating reliable inference. Going just by the posterior mean

⁴⁶Edge et al. (2007), who collect real-time estimates of long-run productivity from various sources, including historical Economic Reports of the President, document a similar pattern in the trend productivity estimates.

estimate, the evidence suggests a counter-cyclical behavior of labor productivity, which weakens over time.

The model specification Base-P*CycOutputgap, which relates productivity to output gap instead of unemployment gap (through parameter ϕ^p), provides generally similar inference about the cyclical nature of labor productivity. The plot for the time-varying parameter ϕ^p is relegated to appendix C13 to conserve space. The credible intervals are wide and include both positive and negative values. Based on the posterior estimate of parameter ϕ^p , productivity is procyclical in the 1960s and post-2010, but it is either acyclical or countercyclical in other periods. Overall, the empirical evidence generally corroborates the evidence presented in Gali and van Rens (2020). The model comparison results indicate a slightly inferior fit of the Base-P*CycOutputgap model compared to Base (-608.1 vs. -606.9 in the case of productivity data and -1773.8 vs. -1771.7 for the overall fit).

The importance of SV in productivity equation

An essential element of our decomposition of labor productivity into the trend, cyclical, and idiosyncratic components, which others have abstracted from, is that we permit a time-varying variance of the idiosyncratic component. Empirically, our results show this to be an important extension, as shown in the SV plots included in panel (f) and the model comparison results reported in table 4.6. The plots indicate statistically significant evidence of time-variation in the volatility of the idiosyncratic component. The model comparison further provides evidence supporting SV inclusion, as the Base-NoSV is the worst performing and has a significantly worse fit to productivity data compared to Base.

We also explored the possibility that the SV may be soaking up the variation in productivity, which otherwise would have been attributed to the cyclical component of productivity. The model specification Base-NoSV, which shuts down SV in the idiosyncratic component of productivity (and other model equations), yields estimates of λ^p similar to Base, suggesting that SV is not contributing to the ambiguous result on the cyclical nature of labor productivity.

4.4.4 Estimation results for π -star

Figure 4.7, panels (a) and (b) plot the posterior mean estimates of π -star along with the 90% credible intervals from the Base and Base-NoSurv model specifications, respectively. Panel (c) plots the corresponding precision estimates, defined as the width of the 90% intervals. A quick visual inspection shows that the π -star from the Base specification is significantly more precise than Base-NoSurv, as evidenced by narrower credible bands and the precision plot corresponding to Base lying below the Base-NoSurv plot. Based on the marginal likelihood criteria, the fit of the inflation equation to the data in the case of Base is greater than Base-NoSurv (as reported in Table 4.8). The fit of the overall Base model to the data is also higher than Base-NoSurv. Our finding that adding survey expectations improves both the model fit and the precision of π -star is consistent with CCK.

The broad contours reflected in the posterior mean pi-star from the two models are similar to those documented elsewhere in the literature (e.g., CCK). For instance, pi-star was low in the 1960s, high in the 1970s, fell sharply in the 1980s, continued a steady deceleration in the 1990s, fluctuated in a narrow range between 2.0% and 2.5% in the 2000s, and has been below 2% since 2012. This general pattern is consistent with the widely held view. Focusing on the specifics, unlike some papers (e.g., Stock and Watson, 2007; Mertens, 2016), which show two peaks in pi-star, one in the mid-1970s and another in the early 1980s, our model-based estimates (both with and without survey data) do not show the earlier peak (similar to CCK). Relatedly, in those same papers, the pi-star is estimated to peak at a level of 10% or higher; in contrast, the mean estimate of pi-star in our model specifications peak at a lower level (similar to CCK and Mertens, 2016 – in his model spec that augments survey data).⁴⁷

Comparing estimates from Base and Base-NoSurv specifications, the level of pi-star is similar in the 1960s but starting in early 1970, pi-star from the Base specification sharply accelerates to peak at 6% in early 1980, whereas, while the estimates of pi-star from Base-NoSurv specification also accelerate but peaks at a lower level of 4.5%. As shown, uncertainty about pi-star increases sharply in early 1980, with the Base-NoSurv estimates experiencing a much more dramatic rise. It is the case that uncertainty around pi-star (as measured by the width of 90% credible intervals) inferred from the Base-NoSurv is higher compared to the Base throughout the estimation sample. But in early 1980, the differential in uncertainty is twice as large, as can be seen comparing dotted and solid plots in panel (c).

A similar rise in model-based estimates of pi-star uncertainty in the late 1970s through early 1980 (known as the Great Inflation period) has been noted elsewhere (e.g., Mertens, 2016). A subset of literature attributes the rise in pi-star uncertainty to un-anchoring of inflation expectations during the Great Inflation period. Beginning early 1980 through early 2000, both models have (posterior mean of) pi-star steadily declining to 2%. Between 2000 and 2012, whereas in the case of Base, pi-star is flat at 2%, in Base-NoSurv, it is stable at a slightly higher level of 2.3%. Since 2012, pi-star has slowly moved down to reach 1.5% (in Base) and 1.4% (in Base-NoSurv).

Inflation gap persistence

Panel (d) shows the posterior mean estimates of parameter ρ^π for Base (solid line) and Base-NoSurv (dotted line). Also plotted are the 90% credible intervals from the Base model. As is evident, the uncertainty around the posterior mean is high. That said, both models indicate quite similar estimates of persistence in the deviations of inflation from pi-star. There is also strong evidence of time-variation in inflation gap persistence. For example, gap persistence was low (0.25 to 0.35) in early 1960, but from thereon began to increase steadily, reaching close

⁴⁷Since both Stock and Watson (2007) and Mertens (2016) endow SV to the RW process governing the pi-star whereas we do not, this difference in the modeling assumption may explain why the difference in the pi-star estimates around the Great Inflation period. However, CCK, who also allows SV in the pi-star process, yields pi-star generally similar to our pi-star estimates suggest that the SV assumption is likely not the answer.

to 0.85 by early 1970, and remained at that level through the early 1980s. Subsequently, the persistence declined steadily to 0.4 by the early 1990s. From the late 1990s to the early 2000s, persistence fell further to 0.3, and it remains at that level.

Price Phillips curve

Panel (e) plots the posterior mean estimate of parameter λ^π , which is the slope of the price Phillips curve. As before, the plots from the main model specifications are shown. (The 90% credible intervals are from the Base model). The plots indicate both model specifications yield very similar estimates of the parameter linking the inflation gap to the unemployment gap. The estimates also indicate strong evidence of time-variation in the slope of the Phillips curve. For example, both models estimate a steeper Phillips curve in the 1960s that subsequently weakens (becomes less negative) over time through 2010. From thereon, it slowly begins to become steeper (more negative), ending 2019 at -0.23, which is still weak historically speaking and is surrounded by wide intervals spanning -0.05 to -0.52. This pattern is consistent with the familiar narrative (also documented in several other papers) that the “*Phillips curve has weakened over time.*”

SV in price inflation equation

Panel (f) plots the volatility estimates (i.e., the standard deviation of the shocks to the inflation gap, $e^{h_t^\pi}$) corresponding to the two model specifications. The figure shows that the posterior mean estimates are quite similar. The estimates imply high volatility during the period of Great Inflation that fell subsequently. Inflation volatility increased sharply again during the Great Recession but has trended lower since then. By 2019, inflation volatility had returned to the low levels of the early 2000s but remains shy of historic lows of the mid-1960s and mid-1990s. The contours of the volatility in the inflation gap are generally similar to those reported in CCK and Chan, Koop and Potter (2013). Bayesian model comparison indicates strong evidence supporting the inclusion of SV in the price inflation equation.

Link between survey and pi-star

In figure C6 (appendix), panels (a) and (b) plot the posterior estimates of the coefficients C^π and β^π that provide a sense of the estimated relationship between survey forecast and pi-star. The prior assumes an unbiased relationship between survey forecasts and pi-star (i.e., prior mean $\beta^\pi = 1$ and at all periods $C_t^\pi = 0$). The model estimation yields posterior mean estimate of 0.99 for β^π , with 90% credible interval spanning 0.91 to 1.07 (also reported in table 4.2). The posterior mean of C_t^π displays considerable time-variation; however, the 90% credible intervals, for the most part, include zero. Taken together, the posterior estimates of β^π and C_t^π imply that the survey forecast, on average, is somewhat a biased measure of pi-star. This latter result confirms the findings of CCK.

In the supplementary appendix C11.a, we include the results and discussion comparing pi-star estimates from the Base model to external models: CCK, CKP, and UCSV. The estimates indicate that the CCK model generates the most precise pi-star, followed by the Base model, CKP, and UCSV. The latter model yields volatile and erratic estimates of pi-star (and precision).

Summary: pi-star

All told, we summarize the analysis of the inflation block as follows. First, the Base model and its variants indicate contours of pi-star that corroborate the narrative documented elsewhere in the literature. However, in some periods, pi-star estimates can differ notably across models, and as emphasized in CCK, these differences can have important implications for monetary policy. Second, comparing across Base and Base-NoSurv specifications, and comparing the Base specification with outside models, strongly suggests the usefulness of survey forecasts in improving the econometric estimation of pi-star (i.e., survey forecast information yields sensible estimates of pi-star and improved precision); hence, corroborating evidence in CCK, Mertens (2016), and Nason and Smith (2020). Third, we find evidence in support of incorporating time-variation in the price Phillips curve. Specifically, suggesting, in a broad sense, weakening of the relationship between inflation and labor market slack since the 1960s, something also documented by several others (e.g., CKP; Stella and Stock, 2015; Del Negro et al., 2020).

Fourth, as found by CKP, CCK, Mertens and Nason (2020), among others, we find evidence supporting time-variation in the persistence of the inflation gap. Fifth, as noted by many, including Stock and Watson (2007) and Clark and Doh (2014), we find strong evidence of SV in the innovations of the inflation equation. Sixth, as in CCK, we find evidence that the survey forecast of PCE inflation is a biased measure of pi-star. Lastly, the Base model’s improved fit to inflation data compared to some of its variants provides evidence supporting the long-run theoretical restriction defined by equation (4.28), which imposes $w\text{-star} = \text{pi-star} + \text{p-star}$.

4.4.5 Estimation results for W-star

In modeling w-star, a novel feature of our framework is the decomposition of w-star into its fundamental components, pi-star and p-star. Figure 4.8 presents posterior estimates of w-star along with the decomposition. The first row in the figure plots estimates from the Base model, and the second-row plots estimates from the Base-NoSurv. Third-row plots p-star estimates from other model variants alongside Base and Base-NoSurv models, and also presents precision estimates of the w-star.

The estimates imply w-star increased steadily in the 1970s and peaked in the early 1980s. This increasing w-star reflected upward drift in pi-star that more than offset the downward drift in p-star, as evidenced by the widening in the shaded area representing pi-star and slight narrowing of the shaded area representing p-star. Base model implies w-star peaked at 7.3% in the early 1980s, while Base-NoSurv has w-star peaking at 6%. Both models have w-star sharply drifting lower through much of the 1980s to reach near 4% by early 1990. From thereon, the

path of w-star across the two models is very similar and indicates a gradual slowing to 2.5% by the end of 2017. W-star moved up to 3.0% by the end of 2018, only to fall back to 2.8% in 2019.

Not surprisingly, w-star is more precisely estimated in the Base model than in the Base-NoSurv model, as shown in panel (f). The considerable uncertainty around w-star, implied by the Base-NoSurv model during the 1970s, is mostly driven by pi-star. As noted earlier, the pi-star estimate from the Base-NoSurv was highly imprecise during the 1970s. Interestingly, despite the inferior precision of the Base-NoSurv compared to Base, the Bayesian model comparison suggests a better fit of the Base-NoSurv model to the nominal wage data than Base (see Table 4.7). However, the overall fit of the Base-NoSurv to data (which includes data beyond nominal wage) is slightly worse than Base.

Sensitivity of w-star to modeling assumptions

Panel (e) plots estimates of w-star from two additional Base model variants, Base-W*RW and Base-NoPT. The Base-W*RW model eliminates the long-run restriction that w-star is the sum of p-star and pi-star (on average) and instead models w-star as a RW process. Interestingly, the path of w-star implied by Base-W*RW is similar to Base-NoSurv through the mid-1980s. From thereon, w-star from Base-W*RW is below the Base-NoSurv through early 2010. Since then, it is identical to Base and Base-NoSurv. Although in the first half of the estimation sample, the w-star from Base-W*RW is less precisely estimated than Base, in the second half of the sample, it is more precise. According to the Bayesian model comparison, the fit of the Base-W*RW to the nominal wage data is inferior to both Base-NoSurv and Base. However, compared to Base, the degree of inferiority is only slight. The overall fit of the Base-W*RW model to the data is substantially worse than achieved by either the Base and Base-NoSurv model.

The Base-NoPT model eliminates the passthrough from prices, i.e., it removes the price inflation gap from the equation describing the wage inflation gap. In other words, the direct link between the cyclical components of prices and nominal wages is eliminated, but the connection between the permanent components pi-star and w-star remains. Doing so has notable implications for the estimate of w-star and the model's fit. As shown in panel (e), the w-star implied from Base-NoPT is higher than that implied by the other models through the first half of the sample. While the Base model has w-star peaking at a little above 7% in the early 1980s, the Base-NoPT model implies a higher peak of 8.2%. The acceleration in w-star implied by the Base-NoPT model during the 1970s is much stronger than that implied by the Base model estimates. This stronger path of w-star is associated with more precise estimates of w-star in the 1970s compared to the Base and other models, as can be seen in panel (f). However, according to the Bayesian model comparison reported in table 4.7, removing the connection between the cyclical components negatively impacts the model fit, as evidenced by the substantially inferior fit of the Base-NoPT model to data compared to the Base model.

Wage persistence, Wage Phillips curve, Passthrough from prices, and SV in wage equation

Figure 4.9 presents the time-varying posterior estimates of the parameters describing the persistence in the nominal wage gap, the wage Phillips curve, the short-run passthrough from prices to wages, and stochastic volatility of the shocks to the nominal wage equation. Estimates from three models, Base, Base-NoSurv, and Base-W*RW are presented. A quick visual inspection indicates both statistically and economically significant evidence of time-variation in these parameters that capture important empirical relationships.

Wage gap persistence

Panel (a) plots the posterior estimates of the parameter ρ^w , capturing the persistence in the nominal wage inflation gap. Also included are the 90% credible intervals from the Base model. The posterior mean estimates for the three models indicate increasing persistence in the wage inflation gap beginning in early 1960 and peaking in mid-1980. The Base model shows a notably higher peak at 0.45 compared to 0.39 and 0.33 for Base-NoSurv and Base-W*RW, respectively. The higher peak suggests that the Base model attributes a higher share of fluctuations in the cyclical nominal wages to the persistence component than that implied by the other two models. From the mid-1980 to the early 1990s, the persistence steadily declines but after that increases through the mid-2000s; from there on through the early 2010s, the estimated persistence in the nominal wage gap falls to levels seen in the mid-1970s. Since then, it has been slowly increasing. It is worth noting that credible intervals around the posterior mean are wide, suggesting high uncertainty in the inference about the estimated persistence.

Wage Phillips curve

In Figure 4.9, panel (b) plots the estimates of parameter λ^w , which captures the wage Phillips curve relationship in the US data. The plot provides strong evidence supporting the existence of the wage Phillips curve in the post-war data. Notably, the plot also offers convincing evidence of a time-variation in this relationship. According to the posterior mean estimate, from the early 1960s through mid-1970s, our models imply the strength of the wage Phillips curve at a moderate level, but from thereon through the mid-1980s, the wage Phillips curve steepened sharply in Base and Base-NoSurv (less steep in Base-W*RW). By the mid-1980s, in the Base model, the posterior mean of the wage Phillips curve parameter is estimated to be -0.5 , with 90% credible intervals ranging from -0.23 to -0.75 . From thereon, it gradually flattened until the mid-2000s, but soon afterward, it began to flatten more rapidly through to early 2010. In 2010, all three models estimate the posterior mean of the Phillips curve parameter at -0.2 . The prevalence of the downward wage rigidities during the Great Recession is among the primary explanations for the flattening wage Phillips curve (see Daly and Hobijn, 2014).

From 2010 onwards, with an improving economy, the estimated wage Phillips curve has steadily steepened. Our empirical evidence on the wage Phillips curve is consistent with the

findings of Knotek and Zaman (2014), Peneva and Rudd (2017), and Gali and Gambetti (2019), who all document strong support for the continuing existence of a wage Phillips curve in the US data.

Passthrough from prices

Panel (c) plots the estimates of the short-run passthrough from prices to wages defined as $\frac{\kappa_t^w}{1-\rho_t^w}$. The posterior estimates of passthrough indicate a weakening relationship between cyclical nominal wage inflation and cyclical price inflation over the estimation sample, confirming the evidence presented in Peneva and Rudd (2017) and Knotek and Zaman (2014). The relationship between the two was strong in the 1970s through to the mid-80s, but since then, it has gradually weakened such that it has been nonexistent (i.e., the passthrough is estimated to be zero) for the past decade. Coincidentally, this period of the breakdown in the relationship between the two cyclical components has coincided with low and stable price inflation, and the Federal Reserve has adopted an explicit target of 2%. Literature has attributed various explanations to this breakdown in the relationship that includes improved anchoring of inflation expectations (Peneva and Rudd, 2017) and amplification of downward wage rigidities during low levels of price inflation (Daly and Hobijn, 2014).

Our empirical finding also supports the results of Bobeica et al. (2019), who show using euro area and US data that the link between labor compensation and price inflation in the short-run importantly depends on the prevailing inflation regime. They find a high inflation regime associated with a tighter connection between labor costs and price inflation and a low inflation regime associated with a weaker link. All three of our models imply similar inference. Both Base-NoSurv and Base-W*RW indicate a stronger passthrough than the Base model in the 1970s and 1980s, periods associated with high inflation in the US. Comparing between panels (a) and (c) suggests that during that period, Base attributes more of the increase in nominal wage inflation to increase in persistence than to the passthrough from price inflation, hence the less strong passthrough estimate than seen in the Base-NoSurv and Base-W*RW models.

SV in the wage equation

Panel (d) plots the time-varying standard deviation of the innovations to the nominal wage inflation gap. The plot indicates the importance of allowing for SV in the equation for nominal wage inflation. All three models show very similar posterior mean estimates of SV. The estimates suggest that volatility in the early 1960s was high but then subsequently fell to rise back up again in the late 1960s and early 1970s. From thereon through the mid-1970s, it fell steeply. It increased in the early 1980s, but less sharply than in the previous decades. From the mid-1980s through the early 1990s, volatility moderated to low levels, and since then, it remains flat at that low level. The Bayesian model comparison indicates further evidence supporting the inclusion of SV in the wage equation, as evident by the substantially inferior fit of the Base-NoSV model to nominal wage inflation data compared to Base and Base-NoSurv models (Base-NoSV: -344.3

vs. Base: -277.5; see table 4.7).

4.4.6 Estimation results for r-star

Figure 4.10 presents r-star and the “catch-all” component D estimates for our two main model specifications. The top row of the figure plots the estimates of r-star (panel a) and component D (panel b) from the Base model. Also included in panel (a) are the survey expectations of r-star (which enters our Base model). As can be seen, the contours of (posterior mean) r-star from the Base model track the survey estimate, this suggests that survey data plays an influential role in guiding the model’s assessment of r-star. The posterior mean estimate from the Base model shows r-star staying relatively flat at 3.5% in the 1960s, and then slowly trends down through the 1970s to reach 3% by early 1980. Thereafter, it fluctuates in a range between 2.8% and 3.5% through to the beginning of 2000. From there on, it steadily declines to reach 1.1% at the end 2019.

Panel (b) plots the estimate of component D, whose dynamics are shaped by the survey expectations data and by information from the Taylor rule and IS equations. As can be seen in the figure, component D is imprecisely estimated. According to the posterior mean estimate, in the 1960s, component D exerts slight upward pressure on r-star that is mostly offset by downward pressure coming from g-star (via equation 4.36), helping keep r-star relatively flat. After that, with D remaining flat through 2000, developments in g-star shape the trajectory of r-star. Beginning in 2000 and onwards, all forces (as captured through the model structure) work in the same direction to push r-star steadily downwards. The estimated link between r-star and g-star is of moderate strength (posterior mean of parameter $m = \frac{\zeta}{4} = 0.701$); see table 4.2); therefore, movements in g-star play an influential role in driving r-star.

Moving on to the Base-NoSurv model specification, in panel (c), the mean estimate shows r-star rising from 2% in early 1960 to 3.5% through early 1980 and then remaining stable through early 2000. This trajectory is similar to Gonzalez-Astudillo and Laforte (2020), who also utilize the Taylor-rule and information from long-term interest rates in their estimation. Beginning 2000 onwards, r-star steadily declines to reach 1.4% at the end 2019. The trajectory of r-star from 2000 onwards is similar to that from the Base model. It is worth noting that our models’ indication of a secular decline in the r-star beginning in 2000 is also documented elsewhere in the literature tackling r-star (the exception being JM).⁴⁸ However, the extent of decline varies considerably across studies. The literature attributes this secular decline in r-star to various explanations, including: a trend decline in g-star (e.g., LW, 2016); rising premiums for convenience yield, i.e., increased demand for safety and liquidity Treasury bonds (see Del Negro et al., 2017; Bullard, 2018); and excess global savings (Pescatori and Turunen, 2016).

The uncertainty around the r-star estimate from the Base-NoSurv model is substantially

⁴⁸JM document that r-star from their preferred specification (which allows for SV in TR equation) is generally flat over their sample spanning 1960 through 2018. However, in an alternative specification, which does not permit SV, the r-star estimate exhibits decline in r-star similar to that documented elsewhere. In our examination, r-star trajectory is little changed comparing between Base specification and Base without SV (i.e., Base-NoSV).

higher than Base, as can be seen by comparing panel (a) and panel (c) of figure 4.10, and also shown in panel (b) of figure 4.11. The increased uncertainty in r-star comes from component D, which is imprecisely estimated without the survey data. Furthermore, based on the marginal likelihood criteria, the Base model is favored over the Base-NoSurv model (see table 4.9).

Without the survey information about r-star, the estimated link between g-star and r-star is significantly weaker (posterior mean of $m = \frac{\zeta}{4} = 0.390$; see table 4.2), which is consistent with the evidence documented in Hamilton et al. (2016) and Lunsford and West (2019). Therefore, the movements in component D significantly dominate the contours of r-star in Base-NoSurv model. Both the IS curve and Taylor rule equations shape the evolution of component D. The hump-shaped patterns in both D and r-star reflect the trends in real long-term interest rates (informed from the IS equation) and short-term interest rates (from the Taylor-rule equation). It is interesting to note that the r-star estimate from JM (and Gonzalez-Astudillo and Laforte (2020)) – who use the Taylor-rule equation and information from both short and long-term interest rates – also exhibit hump-shaped behavior though, in the case of JM, it is only slight. As we show shortly, the prior setting on the shock process for r-star (in our Base and Base-NoSurv cases, component D) plays an essential role in shaping the contours of r-star.

As discussed earlier, the estimates of g-star from both Base and Base-NoSurv models are quite similar. Therefore, the primary source of the differential in the r-star estimates between Base and Base-NoSurv is the quantitatively weaker relation estimated between r-star and g-star in Base-NoSurv than Base.

Assessment of policy stance

Panel (e) of figure 4.10 provides an assessment of the stance of monetary policy. Following Pescatori and Turunen (2016), we gauge monetary policy’s stance as the deviation of the short-term nominal interest rate from the long-run nominal neutral rate of interest (defined as the sum of r^* and π^*) – this is the interest rate gap from the Taylor rule equation. A positive interest rate gap characterizes a restrictive monetary policy stance, and a negative interest rate gap implies a stimulative stance. The solid line corresponds to the policy stance inferred from the Base model and the dashed line to that inferred from the Base-NoSurv model. Even with notable differences in the estimates of r-star across the two models, the assessment of the policy stance is remarkably similar throughout the sample. So why are they so similar?

The answer lies in the differences in π -star estimates across two model specifications, as shown in figure 4.7 (panels a and b). In other words, the differences between r-star estimates across the two models are compensated (i.e., offset) by the differences between π -star, such that assessment about the stance of monetary policy across the two models is strikingly similar. For instance, in the 1960s, the r-star estimate from the Base is on average 1.34 percentage points (ppts) higher than that of Base-NoSurv. However, over the same period, the π -star estimate from the Base model is 0.54 ppt lower than that of the Base-NoSurv model, which reduces the difference between their associated assessments of policy stance.

According to our model(s) estimates, the policy stance appeared to be slightly restrictive before the Great Recession, but at the onset of the Great Recession, the policy stance immediately turned accommodative. Since then, it has remained very accommodative (reflecting the effects of unconventional monetary policy). After peaking in late 2015, the degree of accommodation has gradually declined (i.e., the interest rate gap has become less negative), such that, by the end of 2019, it has edged closer to the neutral threshold.

A closer inspection of the figure reveals an interesting insight. Since 1990, both the degree and duration of policy accommodation in response to the recession have been more significant than the previous recession. For instance, the monetary policy stance was more accommodative both in terms of level and duration following the 2001 recession than the 1990-1991 recession. Similarly, during and following the Great Recession, the policy stance in terms of level and duration was more significant than following the 2001 recession. Broadly speaking, our model implied assessment of the stance of monetary policy mirrors that in the assessment documented in Brand and Mazelis (2019).

Random walk assumption for r^ vs. Base (and Base-NoSurv)*

As is commonly done when estimating stars, a random walk assumption for r^* is a popular choice (e.g., Kiley, 2020; JM). Accordingly, we explore this particular modeling choice's fit to the data and empirical properties by replacing the equation linking r^* to g^* and the RW component D with an equation that assumes a random walk process for r^* in our main model specifications (Base and Base-NoSurv). We denote these specifications, Base-R*RW and Base-NoSurv-R*RW. By adopting a RW assumption for r^* , the models will be unable to attribute any specific causes of movements in estimates of r^* .

Figure 4.11, panel (a) plots the posterior mean r^* estimates from the Base-R*RW and Base-NoSurv-R*RW models. To facilitate comparison, also shown are estimates from Base and Base-NoSurv models. Panel (b) plots the corresponding r^* precision estimates (defined as the width of 90% credible intervals). A few observations immediately stand out. First, the estimated r^* from model specifications with the RW assumption, although exhibiting broadly similar contours, are higher than those obtained from the specifications that impose a relation between r^* and g^* . Second, beginning 2000 and onwards, r^* estimates from the specifications with the assumed link between r^* and g^* experience a more stark decline than those from the specifications with the RW assumption. For instance, by late 2019, the estimates of r^* from both the Base-R*RW and Base-NoSurv-R*RW models settle at close to 2%, whereas those from the Base and Base-NoSurv models fall further to the range of 1.2 to 1.4%. This differential is mostly explained by the lack of direct downward pressure from g^* in the specifications with RW assumption.

Third, bringing information from the survey improves the precision of the r^* estimates substantially, irrespective of whether r^* is modeled simply as RW process or the combination of RW component and a component linking r^* to g^* . Fourth, the specifications that

allow for a link between r-star and g-star are more precise than those that do not. The Base specification generates the most precise r-star estimates relative to those obtained from the other three specifications.

Based on the model comparison metric reported in table 4.9, the model indicating the most precise r-star does not necessarily rank as the best fitting model. The fit of the Base specification is slightly inferior to that of both the Base-R*RW specification and the Base-NoSurv-R*RW, but not by very much. This evidence nicely illustrates that the marginal likelihood metric has a built-in penalty that increases as the model complexity goes up. When comparing between Base-NoSurv-R*RW and Base-R*RW specifications, the addition of survey data increases the model's fit to the interest rate data slightly from -215.58 to -214.04 (but reduces overall model fit marginally from -1769.3 to -1770.8). In comparison, moving from the Base-NoSurv to the Base specification, the fit to the interest rate data increases, from -221.98 to -216.4 (and the overall model fit from -1772.8 to -1771.7).

Taken together, the evidence described above suggests that the RW assumption for r-star is a viable option, and bringing in information from surveys helps improve precision of the estimates of r-star substantially. Another result worth highlighting from the model comparison exercise concerns the marginal value of survey information in the estimation of r-star. In a model specification that imposes a relationship between r-star and g-star (e.g., Base; Base-NoSurv), adding survey information is crucial in making the model a competitive alternative. By bringing in survey information, which has a high correlation between g-star and r-star survey expectations (0.8), the estimated link between g-star and r-star in the model becomes stronger ($m = \frac{\zeta}{4} = 0.701$ in Base vs. $m = \frac{\zeta}{4} = 0.390$ in Base-NoSurv).

Given the evidence on model comparison and the precision of r-star estimates, our preferred choice is the Base model because its fit to the data is competitive (just marginally inferior to spec with RW assumption). And it produces the most precise r-star estimate compared to the alternative specifications. An additional factor that influences our choice is the ability to provide an explanation for movements in r-star, which is made possible due to its direct link to g-star.

In a supplementary set of exercises, we illustrate the usefulness of the Taylor-type rule equation and the equation linking r-star to survey expectations for identifying r-star. The addition of the Taylor-type rule equation turns out to be crucial to yield plausible and precise estimates of r-star (see appendix C10.d).

We also explored the role of data versus prior in determining r-star, and in the interest of brevity, the results of the exercises are relegated to the appendix (see C10.a). Here, we briefly mention that results indicate that in the case of the Base-NoSurv model, the prior views determine the shape of the posterior for r-star, confirming Kiley (2020). However, in the Base model, which links r-star to survey expectations, the data does have some influence. In the case of the Base-NoSurv model, we also find that if we loosen the prior on the variance of the r-star process (to values similar to Kiley, 2020), then the data begin to shape the posterior. But, this

comes at the cost of worsening fit to the interest rate data.

Overall, we find that when it comes to r-star, specification choices matter a lot (something also highlighted by Clark and Kozicki (2005), Beyer and Wieland (2019), and Kiley (2020)). The model specifications that include survey expectations yield estimates that are both reasonable and the most precise. In our examinations, the Base specification, which allows the link between g-star and r-star, is equally preferred by the data to variant of the Base specification that assumes r-star as a RW. Since 2000, the best fitting model specifications indicate a steady decline in r-star, similar to documented elsewhere in the literature.

4.5 Real-time Estimates and Forecasting

In this section, we perform two real-time, out-of-sample, forecasting exercises. In the first exercise, we compare the real-time forecasting performance of our two main models, Base and Base-NoSurv. We evaluate both the point and density forecast accuracy for real GDP growth, PCE inflation, the unemployment rate, nominal wage inflation, labor productivity growth, and the shadow federal funds rate. We show that the Base model is more accurate on average compared to Base-NoSurv for all variables of interest except the unemployment rate. We also document our Base model’s superior forecasting properties relative to “hard to beat” benchmarks, including some of the recently proposed UC models for inflation forecasting. By-products of our real-time forecasting exercise are the real-time estimates of the stars from 1999 through 2019. We compare these real-time estimates to the final (smoothed) estimates – based on the entire sample spanning 1959Q4 through 2019Q4.

In the second forecasting exercise, we illustrate the efficacy of the stars’ estimates produced from our models by demonstrating their usefulness in forecasting with external models (e.g., steady-state VARs). We find that the quality of our estimates of the stars from the Base model is generally competitive to the survey estimates, which are commonly used as proxies for stars in VAR forecasting models. For brevity purposes, we relegate the discussion of this latter exercise to the appendix C5.

Table 4.10 presents the results from comparing the out-of-sample forecasting performance (both point and density) between the two models over the forecast evaluation sample spanning 1999Q1 through 2019Q4. The forecast evaluation is based on real-time data vintages and uses a recursively expanding estimation window, where each recursive run uses an additional quarterly data point in the estimation sample.⁴⁹ The forecast accuracy (point and density) is computed from one-quarter ahead to 20 quarters out. Partly due to our focus on the medium-term horizon

⁴⁹Going back in time means that we are using relatively fewer observations to estimate model(s). As is commonly done when performing real-time forecasting using multivariate UC models, we need to tighten priors on the shocks’ variances driving the latent components (see, for instance, Barbarino et al., 2020). Accordingly, we devise a systematic approach to adjusting the prior on the Scale parameters of the inverse gamma distributions defining the variances of the stars. We multiply the Scale parameter with the $factor = (\frac{2T}{N} - 1) * (\frac{T}{N+5(N-T)})$, where N is total sample size from 1959Q4 through 2019Q4, and T refers to the number of data points in a given data vintage. At the end of the sample, the $factor = 1$ because $T = N$.

and partly in the interest of space, we report accuracy metrics for four, eight, twelve, sixteen, and twenty quarters ahead. We evaluate the forecast accuracy using real-time data; specifically, we treat the “actual” as the *third* release of a given quarterly estimate.⁵⁰ For instance, in the case of real GDP, the third estimate for 2018Q4 corresponds to the GDP data available in late 2019Q1. The point forecast accuracy is assessed using the root mean squared error (RMSE) metric, and the density forecast accuracy is evaluated using the log predictive score (LPS). The statistical significance of the point forecast accuracy is gauged using the Diebold-Mariano and West test and in the case of density forecast accuracy based on the likelihood-ratio test of Amisano and Giacomini (2007).

The top panel of the table reports the results corresponding to the point forecast accuracy, while the lower panel for the density forecast accuracy. The numbers reported in the table correspond to relative RMSE – RMSE Base relative to RMSE of Base-NoSurv – in point forecast comparison, and relative mean LPS – LPS Base minus LPS Base-NoSurv – in the case of density forecast comparison. Hence, numbers less than one in the top panel suggest that point forecast accuracy of the Base forecast is more accurate on average, and positive numbers in the bottom panel suggest that density forecast accuracy of the Base forecast is more accurate than Base-NoSurv forecast.

As is evident by the numbers reported in the table, except for point forecast accuracy of the unemployment rate, the evidence generally favors the Base model as more accurate than Base-NoSurv. The evidence in support of the Base model is strongest for PCE inflation and labor productivity growth. In the case of the unemployment rate, as was the case with the in-sample model fit (see table 4.3), the Base-NoSurv model outperforms the Base model in point forecast accuracy. However, over the forecast evaluation sample, the uncertainty around the point forecast of unemployment rate implied by the Base-NoSurv model was higher than the Base model (on average). This higher uncertainty contributed to inferior density forecasts from the Base-NoSurv model compared to the Base model, as is evident by the positive numbers in the row corresponding to the unemployment rate. It is generally the case across all variables, except the real GDP growth, the accuracy of the density forecasts from the Base model are more accurate than Base-NoSurv. In the case of real GDP growth, even though, for the most part, the point forecast accuracy between the two models are similar on average, the density forecasts from the Base-NoSurv model are more accurate than the Base model.

We also compared the forecasting performance of our Base model to the outside benchmark models, which forecasting literature has shown to be useful forecasting devices. Specifically, we compare the accuracy of the inflation forecasts from our Base model to the following three models, UCSV of Stock and Watson (2007) [UCSV], Chan, Koop, and Potter (2016) [CKP], and Chan, Clark, and Koop (2018) [CCK]. We compare the accuracy of the unemployment rate forecasts from our Base model to the Chan, Koop, and Potter (2016), and the accuracy

⁵⁰Results are qualitatively similar if we instead use the revised data (2020Q1 vintage data) as the actual values in the forecast evaluation exercises. The results are available on request from the author.

of the nominal wage inflation from the Base model to the UCSV model applied to the nominal wage inflation – motivated by Knotek (2015). Our forecasting results provide strong evidence in support of our Base model’s competitive forecasting properties. For the sake of brevity, the results are included in the appendix C4 table C3.

Real-time versus final estimates

Up to this point, we only examined the smoothed estimates of the stars inferred using all the sample data, i.e., from 1959Q4 through 2019Q4, which we denote here as final estimates. As discussed in CKP and Clark and Kozicki (2005), the examination of final estimates is beneficial for “historical analysis,” such as the evaluation of past policy. But for real-time analysis, such as forecasting and policy-making, real-time estimates at time t – estimates based on data and model estimation through time t (instead of through 1:T) – are the relevant measures. In estimating the stars, a voluminous number of papers have documented the typical pattern of notable differences between real-time and final estimates, e.g., see Clark and Kozicki (2005) and Beyer and Wieland (2019) for r -star and Tasci (2019) for u -star.⁵¹

Relatedly, several researchers have attributed the inability to precisely know the location of these stars in real-time to past policy mistakes, see Powell (2018) and references therein. The documented differences between the real-time and final estimates, which at times could be dramatic, and the recognition of these differences by policymakers have been the primary reason limiting the usefulness of real-time estimates of the stars in policy discussions in recent years (see Powell, 2018). Hence, there is a strong preference for methods that can provide more credible inferences about stars in real-time.⁵²

Comparing real-time and final estimates of the stars from our Base model suggests that we make some progress in mitigating the difficulties in previous real-time estimation of the stars. However, in the Base-NoSurv model, there is less success in mitigating this issue. We believe a big reason for this lack of success in the latter case is that we estimate a very high-dimensional model with a lot less data (as will be the case when stopping estimation at earlier periods). An artifact of this is that it requires the imposition of very tight priors in earlier periods than when estimating with more recent periods, which, in turn, impacts on the posterior estimates of model parameters and the stars. This latter fact mechanically contributes to more considerable observed differences between real-time and final estimates in the first half of the sample period analyzed. In the case of the Base model, the use of survey information helps anchor the estimates to more reasonable values even in the face of tight priors.

⁵¹Both revisions to past data and the accrual of additional data could contribute to the observed differences between the real-time and final estimates. The estimation with additional data has been found by many to be the primary factor causing revisions to historical estimates of the stars and contributing to divergence between the real-time and final estimates (see Tasci, 2019; Clark and Kozicki, 2005).

⁵²The issue of imprecision in the estimation of stars is an important one. It has been long recognized that considerable uncertainty surrounds the estimated stars complicating reliable inference (e.g., Cross, Darby and Ireland, 2005). Despite the stars’ imprecision, they continue to be used as inputs into policymaking and for other purposes (see Williams, 2018; for r -star). After all, as discussed in Mester (2018), uncertainty about the stars is just one source of uncertainty among many that confront policymakers.

Figure 4.12 plots the real-time posterior mean estimates of the stars from 1999Q1 to 2019Q4. Also plotted are the corresponding final (posterior mean) smoothed estimates and the 68% and 90% credible bands, respectively. The real-time estimates are the end-sample posterior mean (of the smoothed) estimates at any given period, e.g., 1999Q1 estimate corresponds to estimating the model(s) from 1959Q4 to 1999Q1; similarly, 1999Q2 estimate corresponds to estimating the models from 1959Q4 to 1999Q2. As can be seen, for the most part, the real-time estimates of the stars generally remain within the credible intervals, especially the 68% credible sets implied based on full sample information. The exception is the g-star, and in turn, the output gap, the latter shown in figure C3 of the appendix for brevity purpose. In the case of g-star, the real-time estimate, which before 2007 is not aware that the Great Recession (GR) is about to occur, is relatively upbeat like the survey estimate (at 2.9% vs. 2.25% for the final). Between 2007 and 2011, the real-time estimate gradually decelerates to 2.5%, but from thereon through 2015, it declines very rapidly, catching up to the final estimate.

Interestingly, the BC survey (which is real-time data) fell by only two-tenths over this period, but between 2015 and 2017, it dropped another three-tenths to 2.0% and remained at that level through the end of 2019. Whereas beyond 2017, the survey estimate remained flat, the real-time model estimate gradually moved back up partly in response to stronger realizations in real GDP growth than the implied g-star. The more gradual decline in the real-time g-star during the GR implies a larger (negative) output gap estimate than is indicated by the final estimate.

In the case of u-star, interestingly, in the first half of the forecast evaluation sample, the real-time estimate generally fell outside the 68% credible intervals but inside the 90% intervals. In contrast, in the second half, it remained inside the 68% intervals and tracked quite closely the posterior mean of the final estimate. In the case of p-star, the real-time posterior mean estimate largely remained inside the 68% intervals, and post-2007, the implied inference on p-star is consistent with the final estimate, as evident by similar contours of the real-time and final estimates. It is worth noting that both the real-time and final estimates of p-star are notably lower than the survey estimate of p-star; to be sure, survey estimate of p-star is not used in the estimation of the Base model.

In the case of pi-star, the real-time estimate closely tracked the final estimate from 1999 through 2004, and from thereon through 2016, it averaged three tenths higher. From 2017 through 2019, the real-time estimate once again closely tracked the final estimate. The contours of the real-time and final estimates are similar, and the real-time estimate remained within the 90% credible intervals throughout the sample period analyzed. Consistent with a generally higher real-time estimate of p-star and pi-star than the final estimate, the real-time estimate of w-star is higher than the final. Just as in the case of p-star and pi-star, the contours of the real-time and final estimates of w-star are similar, and the real-time estimate of w-star remained within the 90% credible intervals. Also plotted are the implied survey estimates of the w-star, constructed by adding p-star and pi-star survey estimates. Arguably, these implied estimates are implausibly high given that the actual realizations of nominal wage inflation in the post-

2007 period were mostly below these implied estimates. The results of a forecasting exercise presented in the appendix further confirm the implausibility of the implied survey estimate.

In the case of r -star, the contours of the real-time and final estimates are remarkably similar. Between 2000 and 2005, and post-2014, the real-time estimate closely tracked the final estimate. From 2006 through 2013, the real-time estimate averaged 35 basis points higher than the final estimate. In our assessment, the magnitude of the gap between real-time and final estimates is relatively small compared to the uncertainty estimate around the posterior mean and estimates of uncertainty typically reported in papers estimating r -star, e.g., Clark and Kozicki (2005), Laubach and Williams (2016), and Lubik and Matthes (2015). Furthermore, the real-time estimate of r -star remained within the 68% credible intervals throughout the sample period considered. The width of the estimated 68% (and 90%) intervals from the Base model has been remarkably stable between 0.9 and 1.2 percent (1.5 and 2.0 percent) in the last 25 years. For reference, the typical estimates of 90% bands from popular models such as LW and Lubik and Matthes (2015) have a width averaging more than 4% and 3.5%, respectively.⁵³ Given that the 68% and 90% credible intervals are significantly narrower compared to typical estimates reported elsewhere in the literature, we view the evidence of our real-time r -star remaining inside the estimated credible intervals as encouraging.

Overall, the real-time estimates of stars and forecast evaluation based on the past twenty years of data provide empirical evidence supporting the competitive forecasting and real-time properties of the Base model. Ideally, we would have liked to evaluate our models' real-time properties over the more extended historical sample, but are restricted due to the high-dimensional nature of our models and also to availability of the real-time data on nominal wage inflation.

4.6 The implications of COVID-19 Pandemic on Stars

At the time of writing this chapter, the global economy is in the midst of an ongoing global pandemic crisis (GPC), which has continued to inflict significant disruption to economic activity both here in the US and globally. The GPC, which started in early 2020, contributed to extreme movements in many US macroeconomic indicators, including those used in this chapter. For instance, the US real GDP growth in quarterly annualized term declined from -5% in Q1 to -31% in Q2, the deepest contraction in post-war data (COVID-19 recession). And in Q3, growth rebounded to +33%, a record increase in the post-war data. These extreme movements, which were several standard deviations away from their historical averages, contributed to the breakdown of many conventional time series models, especially the time-invariant VAR models estimated with monthly data, see Lenza and Primiceri (2020) and Carriero et al. (2021).

Up to this point, our analysis has focused on the pre-GPC data. In light of the recent work

⁵³We note that recent approaches to model r -star such as JM and Del Negro et al. (2017) also generate precise intervals similar to ours, with JM marginally less precise and Del Negro and all measurably more precise than ours.

documenting the difficulties of the standard time-series models handling pandemic data, we are naturally curious to see how our two main models respond to the COVID-19 GPC data? What do data during the pandemic imply for the estimates of the stars, or, in other words, what are the estimated long-run consequences of the GPC data? As we illustrate shortly, the short answer to the first question is that both models, Base and Base-NoSurv, handled the pandemic data well, but the Base model, in our judgment, did a much better job than the Base-NoSurv model.

The answer to the second question is a bit complicated. But, briefly, the models indicate that, at the moment, the long-run consequences of the economic developments related to the pandemic are highly uncertain. We view this latter characterization as a reasonable one. Let's take long-term productivity growth. Based on past research, the pandemics are associated with the hit to long-term productivity growth; think of disruptions to schools and universities.⁵⁴ On the other hand, given the state of the technological advances and the unique nature of the GPC, which has raised prospects of recurrent pandemics, the recent research argues will lead to an acceleration in automation (see Leduc and Liu, 2020) and, in turn, boost long-term productivity. The net effect of these opposing factors on p-star is uncertain, and our model-based estimates of p-star reflect this via increased assessment of uncertainty around p-star, as shown in the appendix C8.

We believe that the rich features of our models helped position our models, especially the Base model, to handle the pandemic data quite well.⁵⁵ In the interests of brevity, we refer the reader to appendix C8 for the results comparing the pre-pandemic and pandemic periods, and results comparing estimates from the Base model to outside sources including CBO and Morley and Wong (2020).

4.7 Conclusion

This chapter takes up the challenge of developing a large-scale UC model to jointly estimate the dynamics of inflation, nominal wages, labor productivity, the unemployment rate, real GDP, interest rates, and their respective survey expectations to back out the estimates of long-run counterparts of these variables. These long-run counterparts include potential output (gdp-star), the growth rate of potential output (g-star), the equilibrium level of the unemployment rate (u-star), the equilibrium real rate of interest (r-star), the trend in inflation (pi-star), the trend in labor productivity (p-star), and the trend in nominal wage inflation (w-star). The structure of our UC model is guided by economic theory and past empirical research. Past research has highlighted strong evidence of changing macroeconomic relationships and allowing for stochastic volatility in the shocks to cyclical components of a range of macroeconomic

⁵⁴Aucejo et al.(2020) document negative impact of the COVID-19 on the students' academic performance.

⁵⁵The rich features include: (1) modeling the changing economic relationships via the implementation of time-varying parameters; (2) allowing for changing variance of the innovations to various equations (i.e., SV); (3) imposing bounds on some of the random walk processes; (4) joint modeling of the output gap and unemployment gap in particular; (5) and the use of survey forecasts;

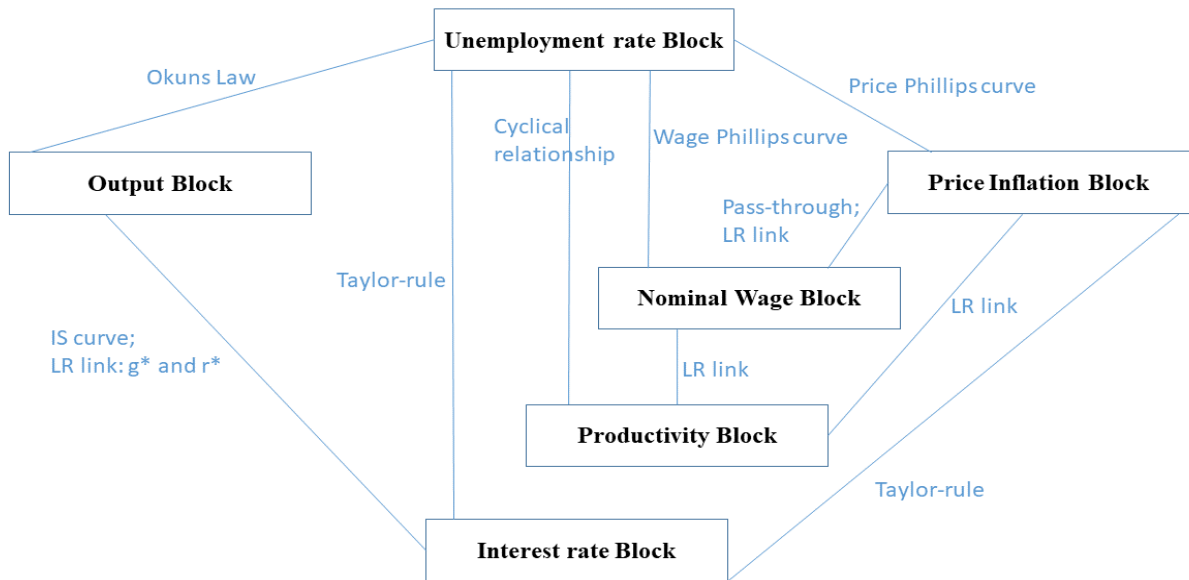
indicators. Accordingly, our model structure permits time-varying parameters and stochastic volatility in most model equations.

An essential feature of our model structure is the explicit role of the survey long-run forecasts in possibly informing the econometric estimation of the stars. Like other researchers, we find adding information from surveys to be an important element in yielding reasonable and credible estimates of the stars. Incorporating a rich set of empirical features leads to a very flexible but heavily parameterized model. To feasibly estimate this model (and its variants), we use Bayesian techniques, specifically the efficient sampling techniques developed in Chan, Koop, and Potter (2013), and the precision sampler proposed in Chan and Jeliazkov (2009).

We explore the empirical relevance of various features incorporated in our baseline model by estimating several variants of the baseline model. The Bayesian model comparison results provide strong support to model features informed based on past research and confirm findings documented elsewhere. For instance, we find that allowing for SV in the model equations to be very important. Similarly, we find economically and statistically significant evidence of time-varying price Phillips curve, wage Phillips curve, the evolving cyclicalness of labor productivity, a changing passthrough relationship between wages and prices, and evolving persistence in price inflation and wage inflation gaps. Given the richness of our model, we document an expansive set of empirical results that we hope will prove helpful for both applied and theoretical macroeconomists alike.

The estimates of the stars from our modeling framework generally echo the contours of stars documented elsewhere in the literature utilizing smaller-scale models. However, they are more precise, a consequence of the use of surveys, a flexible multivariate approach, and jointly modeling the dynamics of several stars. We also show that our baseline model held up well when including the COVID-19 pandemic data. The rich set of features endowed in our UC model helped handle the pandemic data without any difficulties. Lastly, we document the competitive real-time forecasting properties of both our main model and, separately, the estimates of the stars, if they were to be used as steady-state values in external models.

Figure 4.1: Visual Overview of Interactions Between Blocks



Notes: The solid lines represent contemporaneous relationship between the element(s) of the blocks. LR link denotes long-run relationship, i.e., link between stars.

Table 4.1: Description of Model Specifications

Panel (a): Main Models	
Model	Description
Base	Full model as formulated in equations (4.6) through (4.39)
Base-NoSurv	Model as defined by baseline but excluding equations linking surveys to stars (i.e., excluding 4.9, 4.10, 4.15, 4.16, 4.26, 4.27, 4.38 and 4.39). Aim: assess the usefulness of survey data
Panel (b): Auxiliary Models	
Base-W*RW	Base but replaces eq. (4.28) with a RW assumption for W^* as defined eq. (4.28b). Aim: assess the support for the theoretical restriction $W^* = \pi^* + P^*$
Base-PT-Wages-to-Prices	Base but replaces eq. (4.21) with eq. (4.21b) . Aim: asses the usefulness of allowing for pass-through from wages to prices (for reference, Base allows for pass-through from prices to wages)
Base-NoPT	Base but replaces eq. (4.29) with eq. (4.29b). Aim: assess the usefulness of inclusion of wage inflation in the price equation AND the inclusion of price inflation in the wage equation.
Base-NoSurv-NoPT	Base w/o surveys and replaces eq. (4.29) with eq. (4.29b). Aim: usefulness of inclusion of wage inflation in the price equation AND the inclusion of price inflation in the wage equation in a Base-NoSurv.
Base-P*CycOutputGap	Base but replaces eq. (4.17) with eq. (4.17b). Aim: assess empirical link between output gap and productivity gap and whether data support inclusion of output gap instead of unemployment gap.
Base-NoR*Surv	Base but excludes survey expectations of r^* ; i.e., removes eq. (4.38) and eq. (4.39). Aim: assess specifically the marginal value of survey expectations of r^*
Base-NoR*Surv-NoTRule	Base but excludes (1) survey expectations of r^* , i.e., removes eq. (4.38) and eq. (4.39); and (2) policy rule eq. (4.34). Aim: assess the marginal value of the policy rule equation to the model.
Base-R*RW	Base but replaces eq.(4.36) and eq.(4.37) with a RW assumption for r^* as in eq. (4.36b). Aim: assess the support for the theoretical link between r^* and g^* .
Base-NoSurv-R*RW	Base-NoSurv but replaces eq.(4.36) and eq.(4.37) with a RW for r^* as in eq. (4.36b). Aim: assess support for the link between r^* and g^* in a spec. w/o survey data.
Base-R*AR	Base but replaces eq. (4.37) with AR assumption for D. Aim: support of r^* defined as comb. of permanent and transitory components.
Base-NoSurv-R*AR	Base-NoSurv but replaces eq. (4.37) with AR assumption for D. Aim: support of r^* defined as comb. of permanent and transitory components.
Base-NoSurv-R*TightPrior	Base specification but tighter prior for std. of the shock to process D. Aim: assess the impact on r^* estimate of using a tighter prior for D.
Base-NoSV	Base specification but with no SV in any of the measurement equations. Aim: assess the importance of stochastic volatility in shock variances.
Base-NoBoundU*	Base specification without the bounds on the U^* process in eq. (4.8). Aim: assess empirical support for imposing bounds on U^* .

Table 4.2: Parameter Estimates

Parameter	Parameter description	Posterior estimates					
		Base			Base-NoSurv		
		Mean	5%	95%	Mean	5%	95%
a^r	Coefficient on interest-rate gap	-0.069	-0.141	0.002	-0.054	-0.130	0.015
$\rho_1^g + \rho_2^g$	Persistence in output gap	0.738	0.678	0.797	0.732	0.669	0.794
ρ_1^u	Lag 1 coefficient on UR gap	1.283	1.238	1.327	1.277	1.232	1.322
ρ_2^u	Lag 2 coefficient on UR gap	-0.504	-0.542	-0.466	-0.503	-0.540	-0.466
$\rho_1^u + \rho_2^u$	Persistence in UR gap	0.778	0.736	0.821	0.774	0.733	0.814
ρ^p	Persistence in productivity gap	-0.005	-0.118	0.110	-0.001	-0.119	0.115
$m = \frac{\zeta}{4}$	Relationship between r^* and g^*	0.701	0.621	0.784	0.390	0.270	0.560
ρ^i	Persistence in interest-rate gap	0.879	0.838	0.917	0.860	0.816	0.904
λ^i	Interest rate sensitivity to UR gap	-0.229	-0.278	-0.180	-0.243	-0.293	-0.194
κ^i	Interest rate sensitivity to inflation	0.061	0.014	0.107	0.085	0.035	0.134
λ^g	Output gap response to UR gap	-0.457	-0.580	-0.339	-0.454	-0.575	-0.333
ϕ^u	UR gap response to Output gap	-0.108	-0.130	-0.086	-0.118	-0.140	-0.096
$\frac{(1-\rho_1^u-\rho_2^u)}{\phi_u}$	Implied Okuns Law	-2.054	-2.327	-1.803	-1.929	-2.179	-1.699
β^g	Link between g^* and survey	0.929	0.774	1.082	—	—	—
β^u	Link between u^* and survey	0.988	0.919	1.055	—	—	—
β^r	Link between r^* and survey	1.018	0.923	1.117	—	—	—
β^π	Link between π^* and survey	0.991	0.912	1.070	—	—	—
$\sigma_{\pi^*}^2$	Variance of the shocks to π^*	0.121 ²	0.100 ²	0.141 ²	0.127 ²	0.084 ²	0.182 ²
$\sigma_{p^*}^2$	Variance of the shocks to p^*	0.145 ²	0.111 ²	0.183 ²	0.141 ²	0.109 ²	0.176 ²
$\sigma_{u^*}^2$	Variance of the shocks to u^*	0.075 ²	0.064 ²	0.089 ²	0.084 ²	0.071 ²	0.100 ²
$\sigma_{gdp^*}^2$	Variance of the shocks to gdp^*	0.018 ²	0.014 ²	0.021 ²	0.021 ²	0.016 ²	0.026 ²
σ_d^2	Variance of the shocks to d	0.093 ²	0.077 ²	0.110 ²	0.114 ²	0.084 ²	0.148 ²
$\sigma_{w^*}^2$	Variance of the shocks to w^*	0.158 ²	0.112 ²	0.215 ²	0.158 ²	0.111 ²	0.220 ²
σ_{ogap}^2	Variance of the shocks to Output gap	0.725 ²	0.672 ²	0.781 ²	0.723 ²	0.669 ²	0.780 ²
σ_u^2	Variance of the shocks to UR gap	0.268 ²	0.248 ²	0.289 ²	0.265 ²	0.245 ²	0.286 ²
σ_{hp}^2	Var. of the Volatility – Productivity eq.	0.274 ²	0.219 ²	0.336 ²	0.273 ²	0.220 ²	0.335 ²
σ_h^2	Var. of the Volatility – Price Inf. eq.	0.297 ²	0.237 ²	0.364 ²	0.298 ²	0.238 ²	0.365 ²
σ_{hw}^2	Var. of the Volatility – Wage Inf. eq.	0.301 ²	0.237 ²	0.374 ²	0.303 ²	0.239 ²	0.373 ²
σ_{hi}^2	Var. of the Volatility – Interest rate eq.	0.377 ²	0.293 ²	0.469 ²	0.369 ²	0.287 ²	0.454 ²
$\sigma_{\lambda\pi}^2$	Var. of the shocks to TVP λ^π	0.041 ²	0.032 ²	0.052 ²	0.041 ²	0.032 ²	0.051 ²
$\sigma_{\lambda w}^2$	Var. of the shocks to TVP λ^w	0.041 ²	0.032 ²	0.051 ²	0.040 ²	0.032 ²	0.050 ²
$\sigma_{\lambda p}^2$	Var. of the shocks to TVP λ^p	0.044 ²	0.034 ²	0.057 ²	0.045 ²	0.034 ²	0.057 ²
$\sigma_{\kappa w}^2$	Var. of the shocks to TVP κ^w , PT	0.041 ²	0.032 ²	0.052 ²	0.041 ²	0.032 ²	0.052 ²
$\sigma_{\rho w}^2$	Var. of the shocks to TVP ρ^w	0.042 ²	0.033 ²	0.053 ²	0.042 ²	0.032 ²	0.052 ²
$\sigma_{\rho\pi}^2$	Var. of the shocks to TVP ρ^π	0.048 ²	0.037 ²	0.062 ²	0.047 ²	0.036 ²	0.060 ²

Table 4.3: Bayesian Model Comparison: Main Models and selected variants

	Base	Base-NoSurv	Base-R*RW	Base-NoSurv-R*RW
MDD of Inflation	-366.6	-368.6	-366.8	-370.1
MDD of Productivity	-606.9	-607.1	-607.5	-607.4
MDD of Nominal Wage	-277.5	-274.3	-278.0	-274.9
MDD of Unemployment	-24.6	-21.7	-24.7	-21.8
MDD of Interest rate	-216.4	-222.0	-214.0	-215.6
MDD of GDP	-279.7	-279.1	-279.8	-279.5
MDD	-1771.7	-1772.8	-1770.8	-1769.3

Table 4.4: Model Comparison: Variants focused on unemployment rate

	Base	Base-NoSurv	Base-NoBoundU*	Bivariate	Bivariate+Surv	CKP-Adj
MDD of UR	-24.6	-21.7	-25.0	-56.5	-46.5	-61.9

Table 4.5: Model Comparison: Variants focused on GDP

	Base	Base-NoSurv	Univariate	Bivariate	Bivariate+Surv
MDD of GDP	-279.7	-279.1	-296.5	-280.4	-281.9

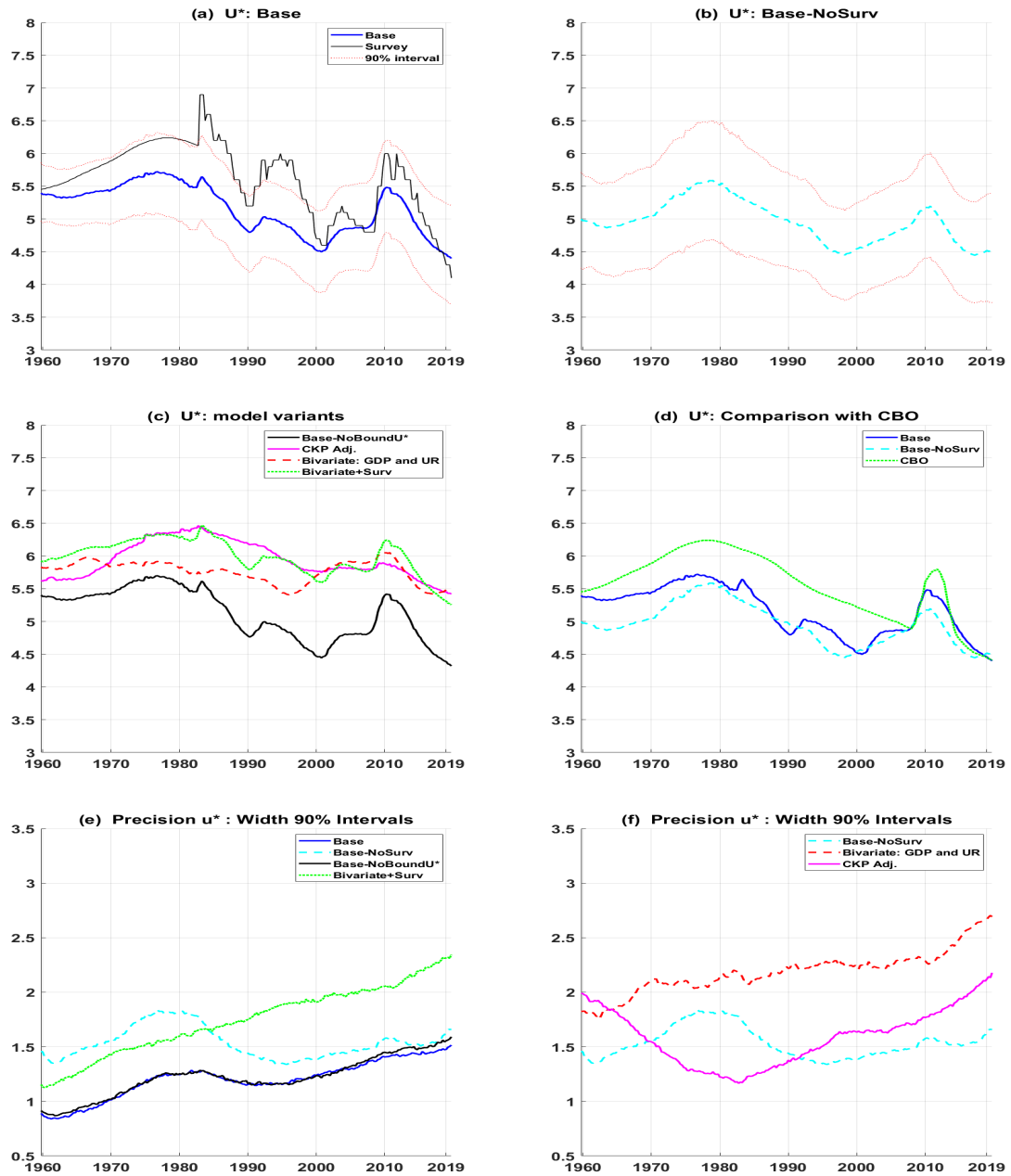
Table 4.6: Model Comparison: Variants focused on labor productivity

	Base	Base-NoSurv	Base-W*RW	Base-P*CycOutputGap	Base-NoSV
MDD of Productivity	-606.9	-607.1	-603.5	-608.1	-633.4
MDD of model	-1771.7	-1772.8	-1790.8	-1773.8	-2024.3

Table 4.7: Model Comparison: Variants focused on nominal wages

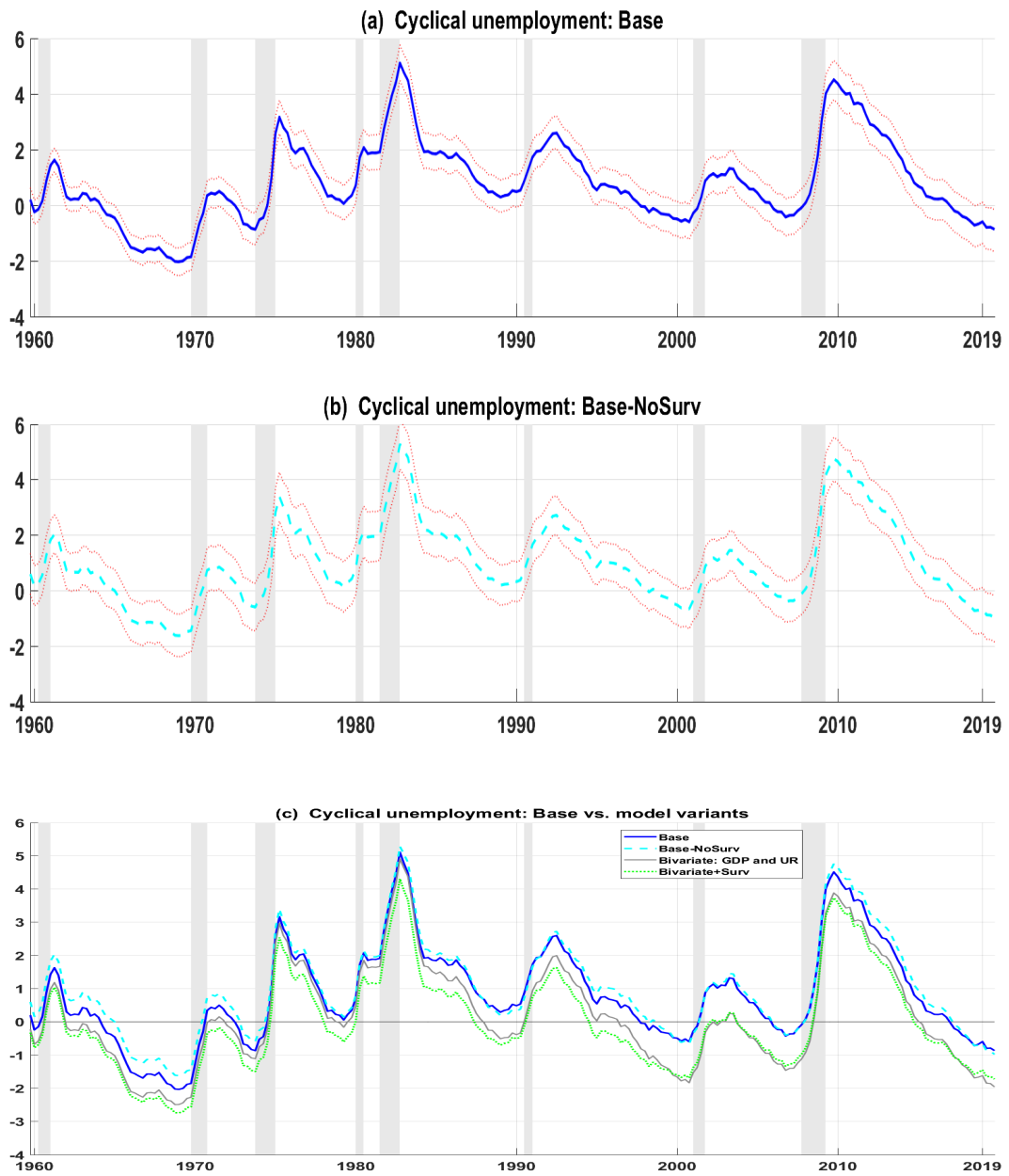
	Base	Base-NoSurv	Base-W*RW	Base-NoPT	Base-NoSV
MDD of Nominal wages	-277.5	-274.3	-277.7	-281.5	-344.3
MDD of model	-1771.7	-1772.8	-1790.8	-1776.3	-2024.3

Figure 4.2: Full Sample Estimates for Unemployment Rate block



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4). In panel (e), CKP Adj. refers to the bivariate model of inflation and the unemployment rate as in CKP but with no bounds on π -star.

Figure 4.3: Full Sample Estimates for Cyclical Unemployment



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4). The shaded areas represent the NBER recession dates.

Figure 4.4: Full Sample Estimates for Output block

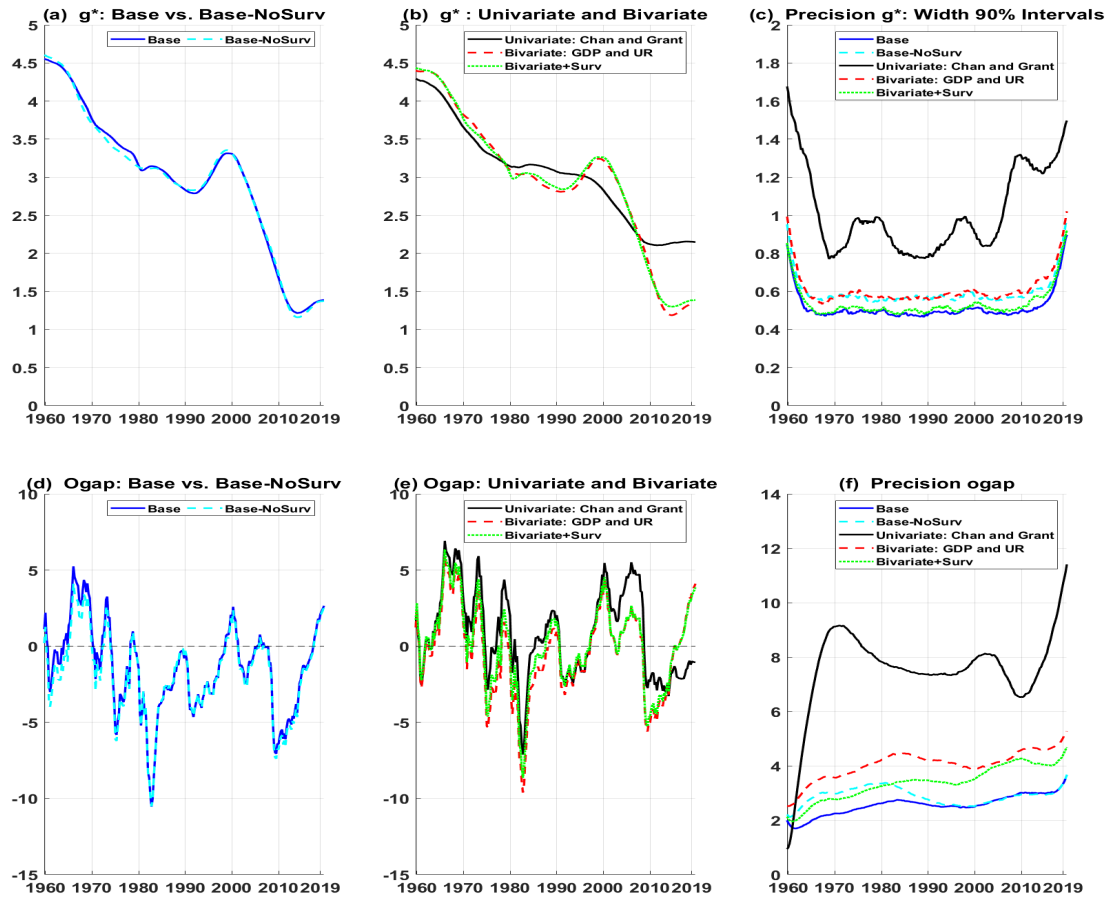


Figure 4.5: Model-based output gap vs. CBO

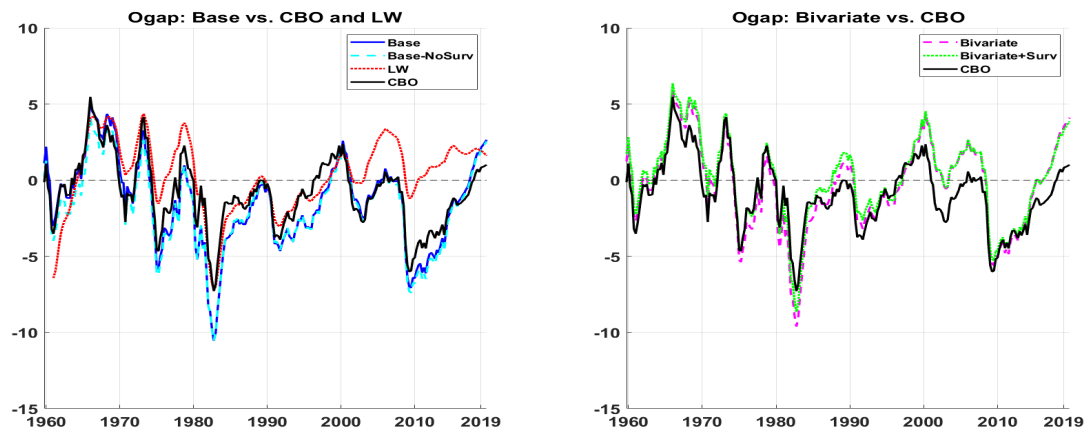
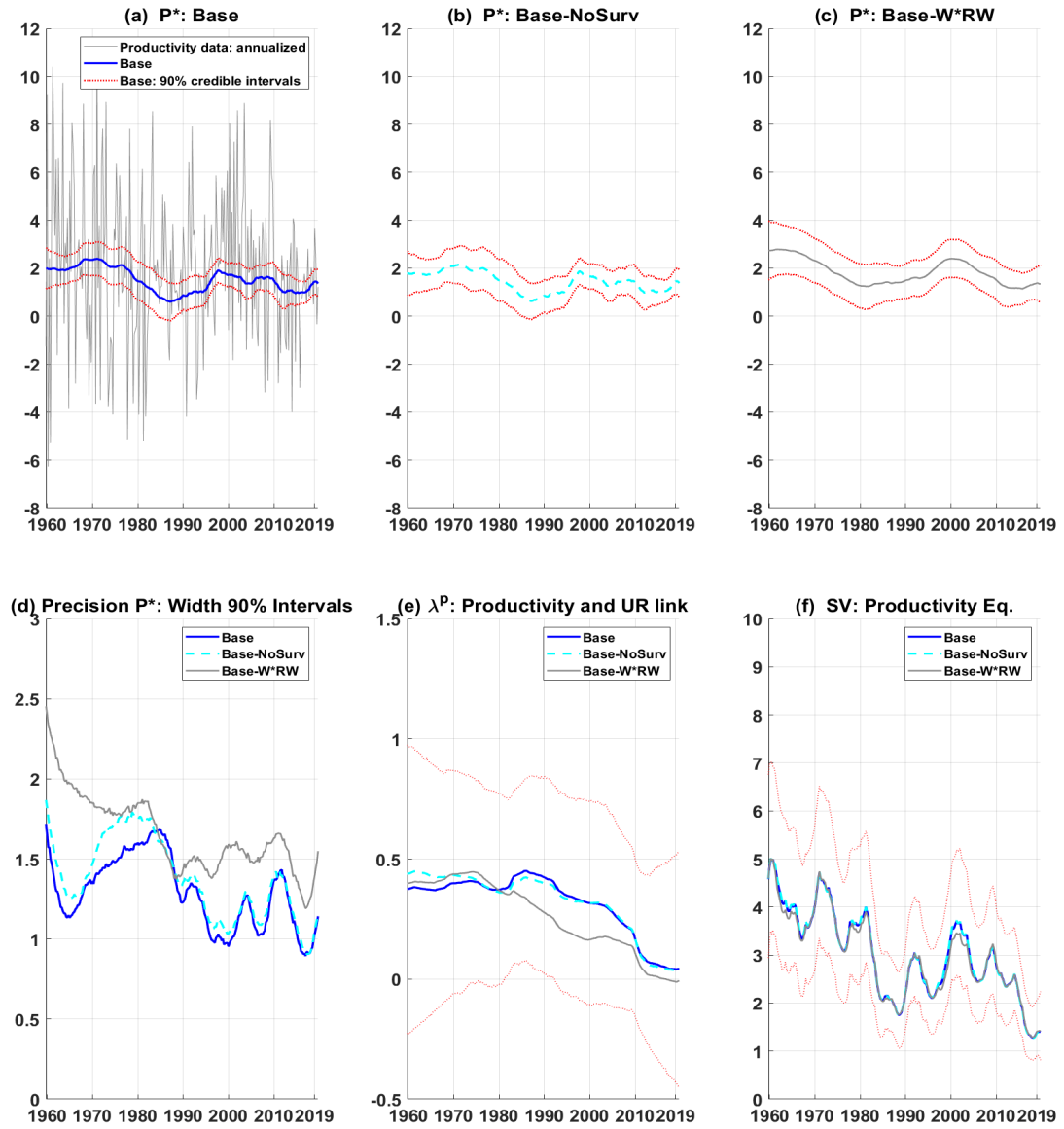
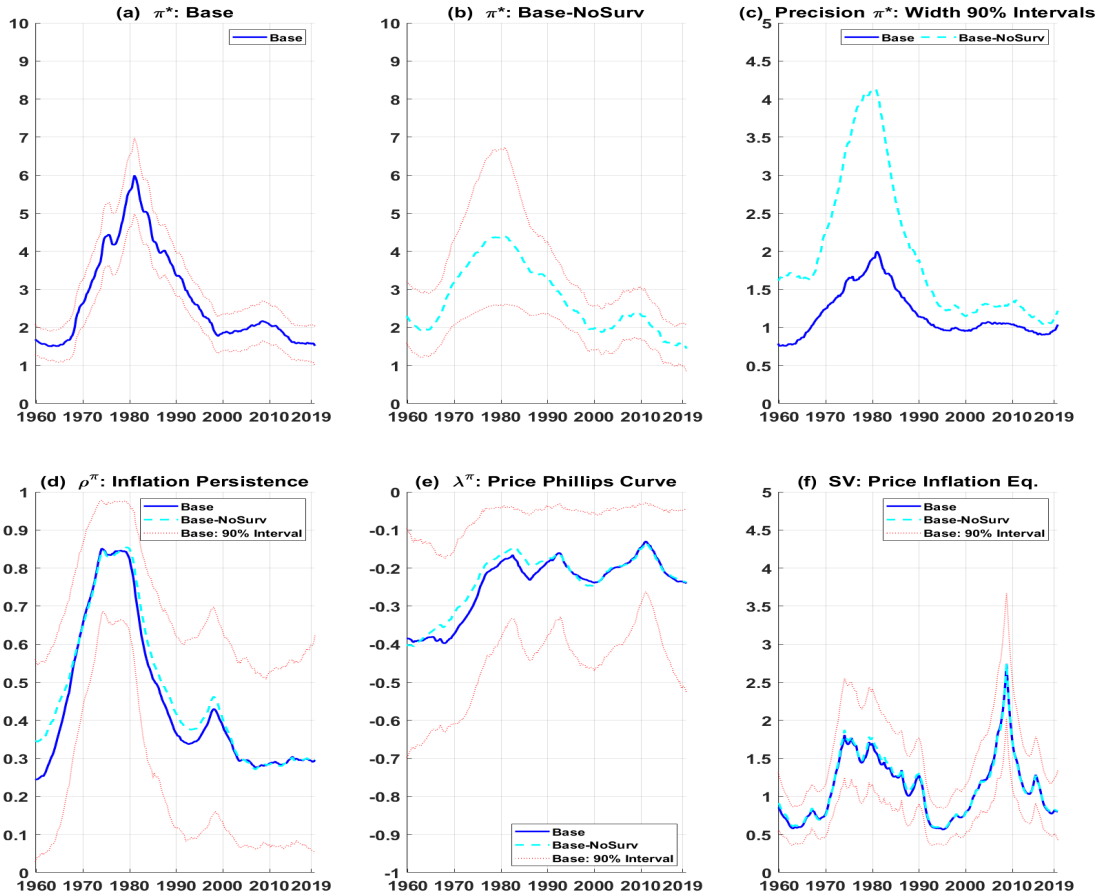


Figure 4.6: Full Sample Estimates for Productivity block



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

Figure 4.7: Full Sample Estimates for Price Inflation block

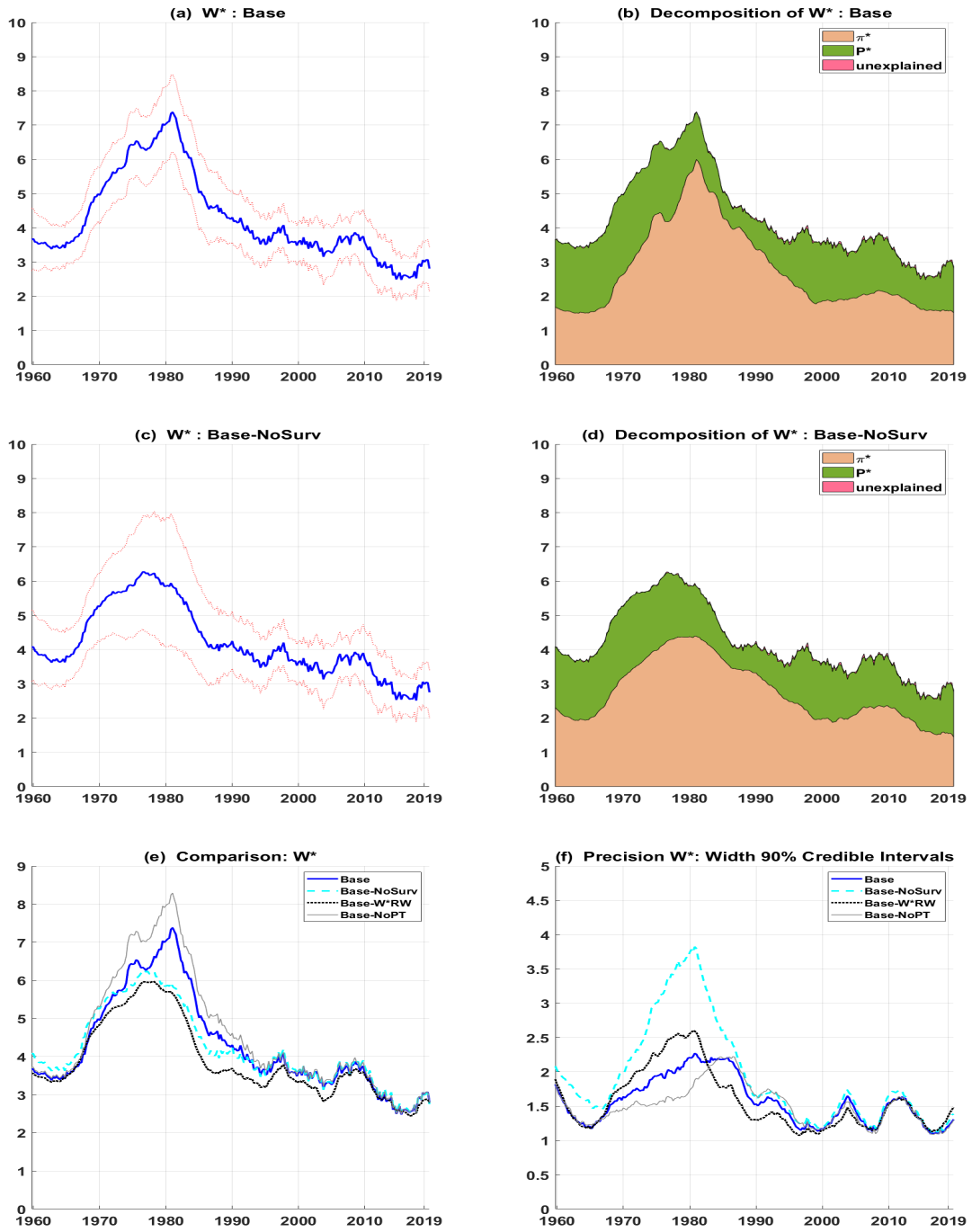


Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

Table 4.8: Model Comparison: Variants focused on price inflation

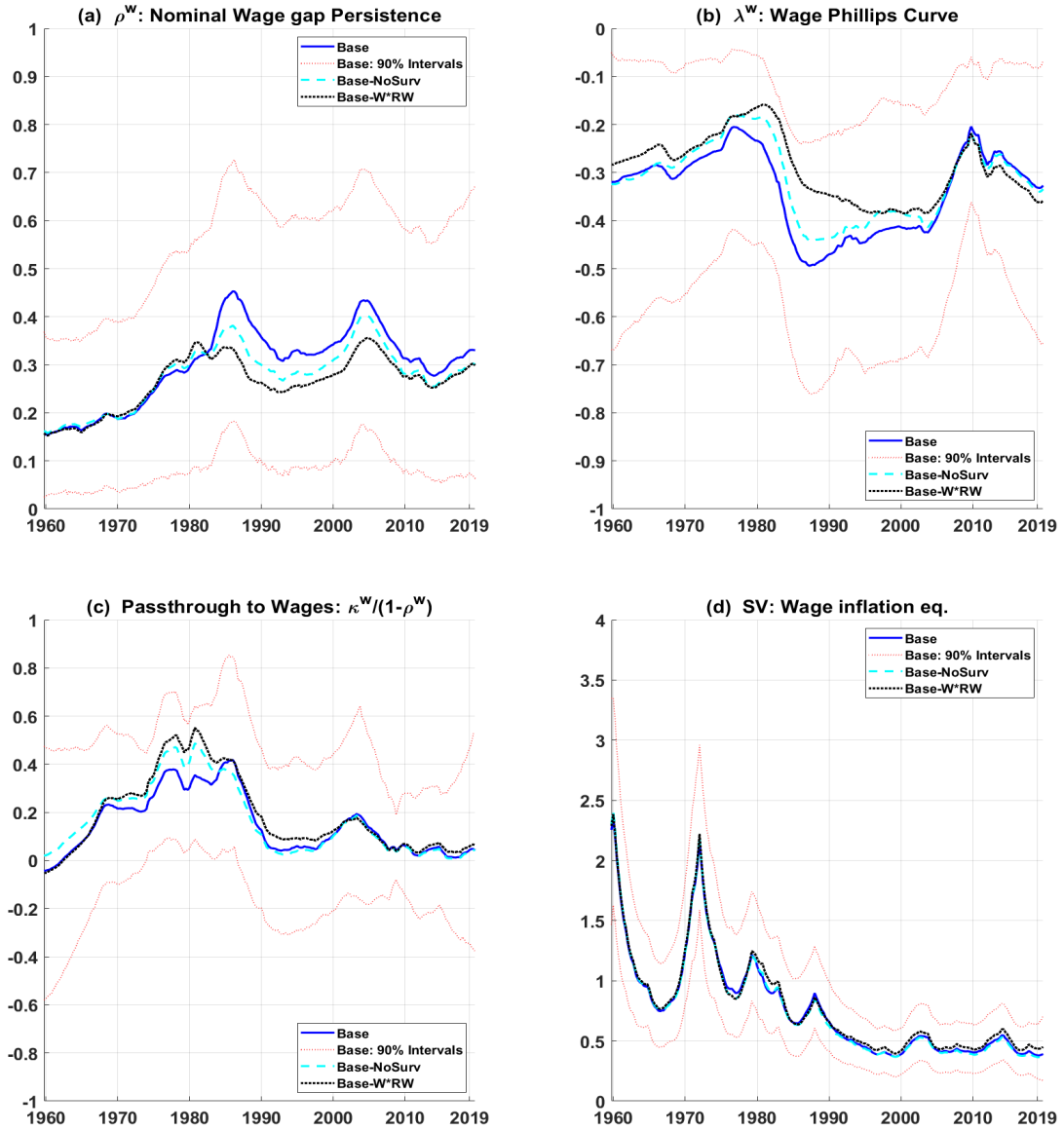
Model variant	MDD of Price inflation	MDD Model
Base	-366.6	-1771.7
Base-NoSurv	-368.6	-1772.8
Base-W*RW	-367.0	-1790.8
Base-PT-Wages-to-Prices	-366.7	-1771.4
Base-NoPT	-365.2	-1776.3
Base-NoSV	-413.9	-2024.3

Figure 4.8: Full Sample Estimates for Nominal Wage block



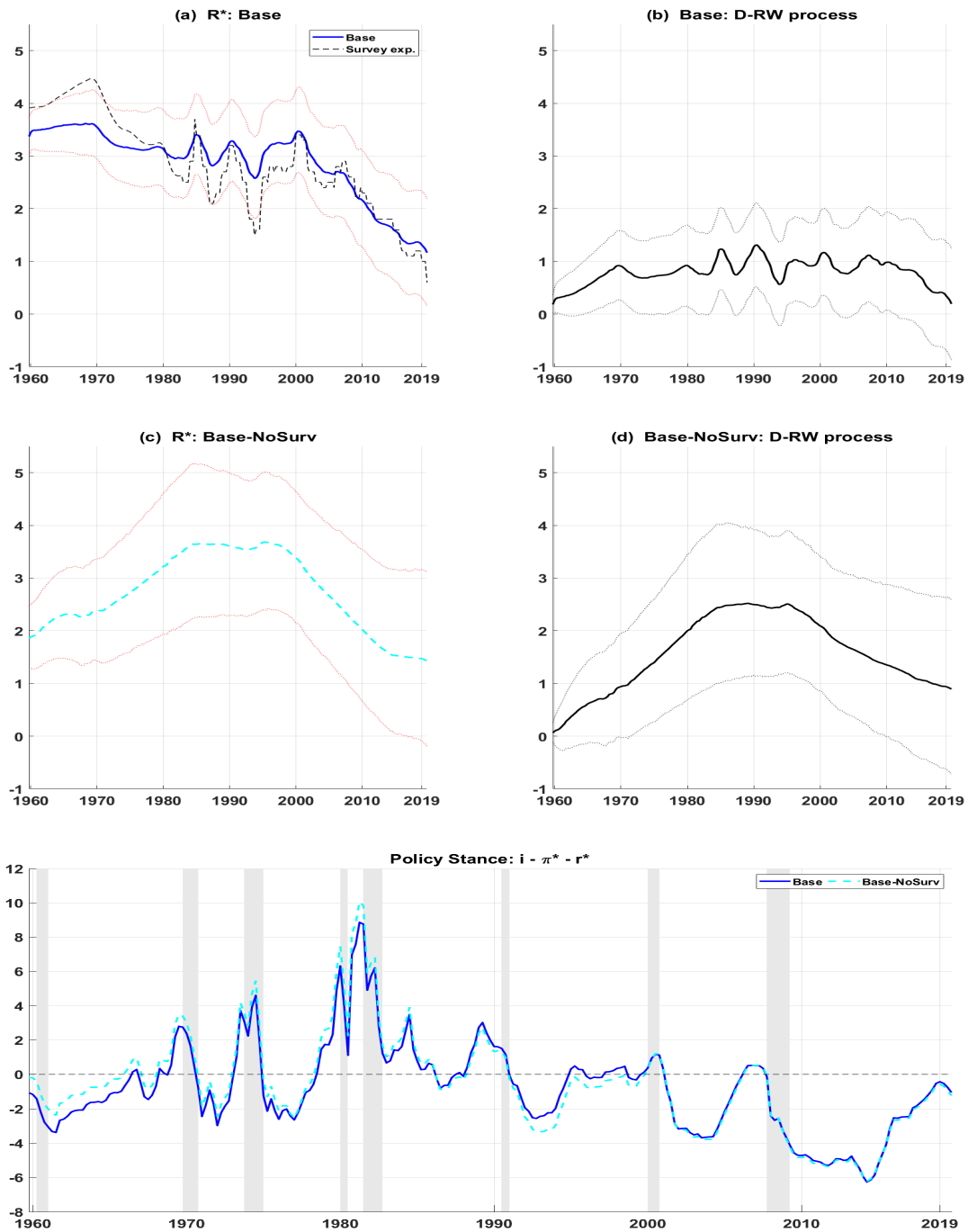
Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

Figure 4.9: Wage Persistence, Wage Phillips Curve, Passthrough from Prices, and SV



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

Figure 4.10: Full Sample Estimates for Interest Rate block



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4). In top panel, the plot labeled "survey exp." is an implied estimate: inferred from the Blue Chip survey long-run estimates of GDP deflator and short-term interest rates (3-month Treasury bill) using the long-run Fisher equation. Specifically, the long-run forecast of 3-month Treasury bill less the long-run forecast of GDP deflator. To this differential, we add +0.3 to reflect the average differential between the federal funds rate and the 3-month Treasury bill.

Figure 4.11: More Estimates for Interest Rate block

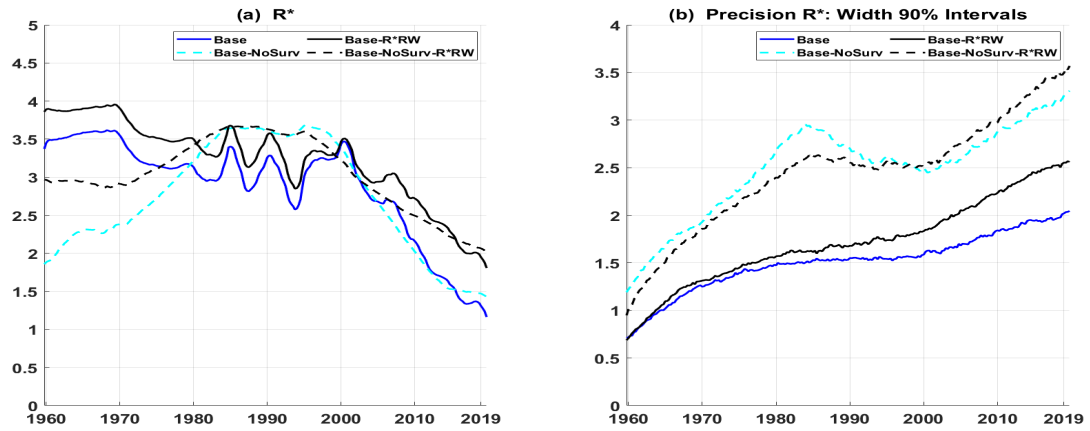
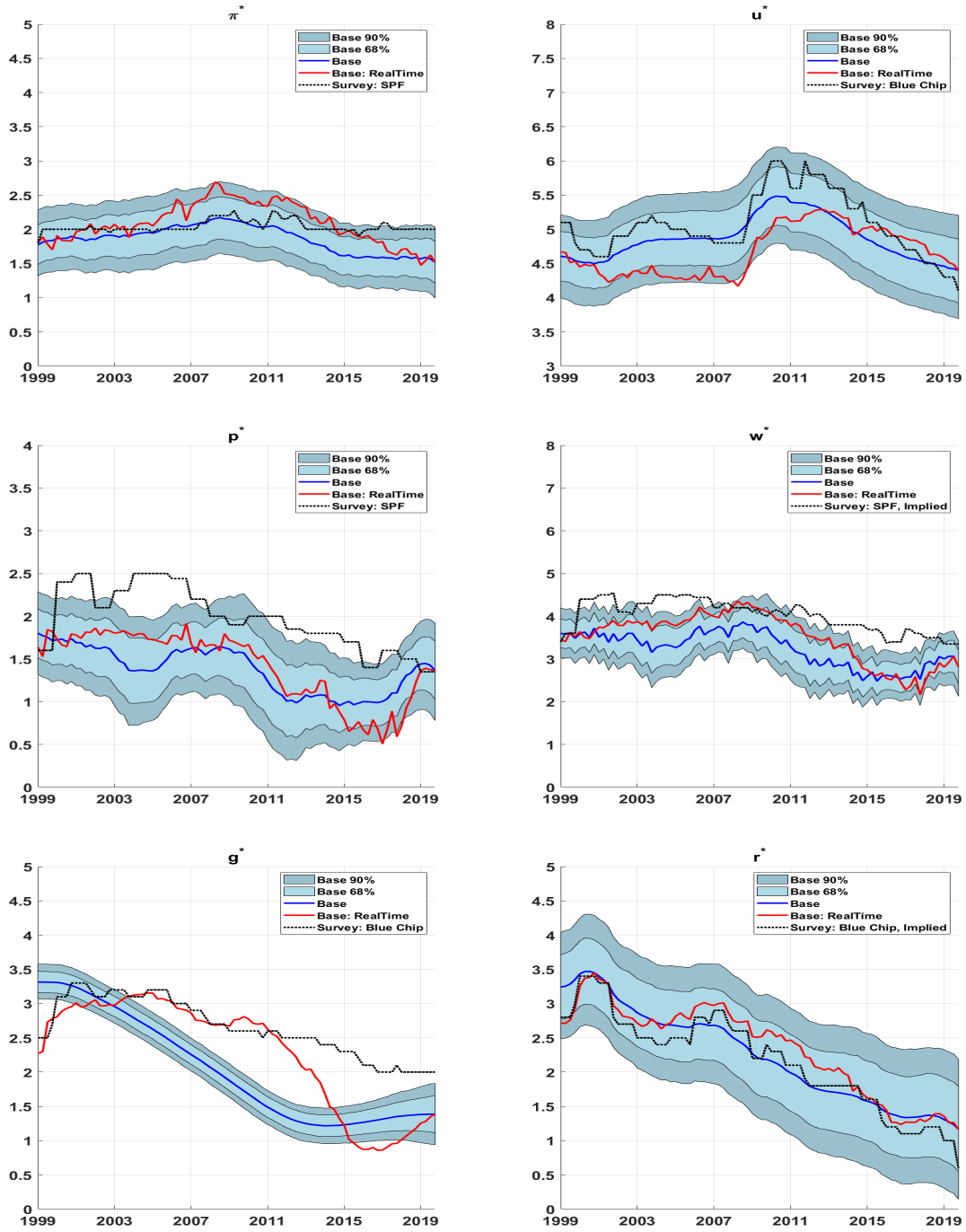


Table 4.9: Model Comparison: Variants focused on interest rate

Model variant	MDD of Interest rate	MDD Model
Base	-216.4	-1771.7
Base-NoSurv	-222.0	-1772.8
Base-R*RW	-214.0	-1770.8
Base-NoSurv-R*RW	-215.6	-1769.3
Base-NoSurv-R*TightPrior	-232.0	-1773.6

Figure 4.12: Real-time Recursive Estimates of Stars: Base model



Notes: The plots labeled Base are posterior estimates reflecting information in the full sample (from 1959Q4 through 2019Q4). The plots labeled Base:RealTime are posterior estimates reflecting information available at a given point in time (i.e., truly real-time).

Table 4.10: Real-Time Forecasting Accuracy: Base vs. Base-NoSurv

Panel A: Point Forecast Accuracy (Recursive evaluation: 1999.Q1-2019.Q4)

Relative RMSE: RMSE Base / RMSE BaseNoSurv					
	h=4Q	h=8Q	h=12Q	h=16Q	h=20Q
Real GDP	1.07*	1.03	1.01	1.00	0.99
PCE Inflation	0.96	0.94*	0.91*	0.89	0.88*
Productivity	0.98*	0.98	0.97	0.98*	0.97
Nominal Wage (AHE)	0.98	1.01	1.05	1.05	1.02
Unemployment Rate	1.07*	1.05*	1.03*	1.02*	1.00
Shadow FFR	0.98	1.00	1.00	0.99	0.98

Panel B: Density Forecast Accuracy (Recursive evaluation: 1999.Q1-2019.Q4)

Relative Log Predictive Score (LPS): LPS Base - LPS BaseNoSurv					
	h=4Q	h=8Q	h=12Q	h=16Q	h=20Q
Real GDP	-0.003*	-0.002*	-0.002*	-0.001*	-0.001*
PCE Inflation	0.018*	0.020*	0.025*	0.027*	0.026*
Productivity	0.001	0.002*	0.002*	0.002*	0.003*
Nominal Wage (AHE)	0.017*	0.016*	0.018*	0.018*	0.019*
Unemployment Rate	-0.003	0.005	0.009*	0.012*	0.013*
Shadow FFR	0.020*	0.007*	0.003	0.000	0.000

Notes for Table: The top panel compares the point forecast accuracy of Base model with Base-NoSurv model. The numbers less than 1 indicates that Base model is more accurate on average. The bottom panel reports the corresponding density forecast accuracy performance. A positive value (for the relative mean predictive log score) suggests that Base model is on average more accurate. The log predictive scores are computed using parametric normal approximation. The table reports statistical significance based on the likelihood-ratio test of Amisano and Giacomini (2007) for the density forecast accuracy, and based on the Diebold-Mariano and West test (with the lag $h - 1$ truncation parameter of the HAC variance estimator) for the point forecast accuracy. The test statistics for likelihood-ratio test use two-sided t-test. In the case of the Diebold-Mariano and West test, the test statistics use two-sided standard normal critical values for horizons less than equal to 8 quarters, and two-sided t-statistics for horizons greater than 8 quarters. *up to 10% significance level.

Chapter 5

Conclusions

5.1 Summary of contributions and policy implications

In recent years, a consensus is developing in the crucial role of “good” nowcasts and reasonable estimates of steady states for obtaining accurate macroeconomic forecasts. My thesis, through a series of chapters, enhances this consensus view. First, it adds support to the consensus via empirical exercises applied to a range of VAR models. The subsequent chapters take the consensus view as given and develop approaches that improve the accuracy of the nowcasts (of inflation) and the estimation of the steady states for a range of macroeconomic variables. Specifically, the second main contribution of the thesis is that it provides a framework to obtain highly accurate nowcasts of inflation, a variable that is of considerable importance to various economic agents, including central bankers. The third main contribution is this thesis develops a *unified* framework to estimate time-varying steady states (i.e., “stars”) for a host of macroeconomic variables. The estimates of these steady states are of value to macroeconomic forecasters. They can also be used for various other purposes, such as the construction of cyclical indicators, guiding fiscal policy by separating fiscal balances into their cyclical and structural components, and much more.

Chapter 2 proposes an approach to adjust the medium- and long-horizon forecasts from a VAR toward plausible values informed by judgmental forecasters. Specifically, by applying the flexible and powerful technique of relative entropy, we tilt the VAR forecast both in the near term with the survey nowcast and in the long run with the survey long-run projection. We denote the resulting forecast as the “hybrid” forecast. The horizon at which the long-run survey projection is combined with the VAR forecast is variable-specific and is determined by the variable’s estimated persistence at the forecast origin. We take as given the fact that survey nowcasts and survey long-run projections are reasonable estimates of starting and terminal conditions (steady states), respectively. Our approach is an alternative to the one proposed by Wright (2013), which uses steady-state BVAR where survey long-run forecasts inform the steady states. However, in contrast to Wright’s approach, which requires specifying VAR in a certain way – specifically in deviations from means – our approach is more general because it

is a post-estimation procedure. Therefore, it can be applied to any VAR model, including a steady-state VAR, or as a matter of fact, to any empirical model that can generate a density forecast. Our results echo many of those reported in Wright (2013), in particular, confirming the importance of reasonable steady states in generating accurate forecasts. Our work is also related to Kruger, Clark, and Ravazzolo (2017), who use relative entropy to tilt VAR forecasts to survey nowcasts. We extend their work: in addition to tilting the VAR forecast in the nowcast quarter, we also tilt the medium-to-long-horizon forecast from the VARs to match the long-horizon forecast reported in the survey of professional forecasters.

Unlike Wright (2013) and Kruger, Clark, and Ravazzolo (2017), who consider only a single Bayesian VAR model, we consider the efficacy of our proposal to a variety of Bayesian VAR models. We summarize our four main findings as follows. First, all the models examined benefit through improved accuracy, with the greatest gains in forecast accuracy seen in models estimated with longer histories and the smallest gains achieved in models that attempt to accommodate structural changes. Second, the gains in forecast accuracy are achieved for most variables. Still, the most significant gains are for variables believed to have undergone marked shifts in their permanent components (e.g., inflation and the federal funds rate). Third, the hybrid inflation forecasts from simple VAR models (such as the fixed-parameter VAR model) rival those of relatively accurate univariate benchmark models and the Federal Reserve's Greenbook. We view this as a valuable and practical contribution because among the many frustrations of monetary policymakers include the inability of multivariate models, which allow for feedback effects from policy to the real economy and inflation, to match the forecasting performances of univariate forecasting models. Fourth, all models considered display demonstrable improvements in the forecast accuracy of hybrid forecasts for real GDP growth and CPI inflation over the last decade (i.e., from the Great Recession to the present), which coincides with the period associated with possible structural change. These results lead us to view our proposal as a low-cost method for mitigating model instability issues that may arise from structural shifts caused by moving endpoints.

Chapter 3 develops a flexible framework based on the model and density combinations to produce density nowcasts for US inflation. By combining individual density nowcasts from three classes of parsimonious mixed-frequency models (MIDAS, Dynamic factor model, and Dynamic model switching of Knotek and Zaman, 2017), our framework generates nowcasts at trading-day frequency and updates as information accumulates over the course of a month or a quarter. Combining densities require a functional form for aggregation and weights to apply to the different densities. Previous research has used either the linear opinion pool or the logarithmic opinion pool as the functional form for aggregation. However, some researchers use both methods and present results for the more accurate approach over a chosen evaluation period. Instead of enforcing a particular functional form for aggregation at the outset, we devise and implement a novel and flexible aggregation strategy that allows the data to determine which of the two functional forms is preferred dynamically. A potential advantage of this flexibility is

that it allows for the possibility of switching between the two functional forms at different points within a month or a quarter based on the nowcast origin or at different points in time during the sample as more observations become available. We complement this flexible aggregation strategy by examining various dynamic model averaging approaches, where weights used to combine the nowcasts can be updated based on learning from past performance. An important feature of this proposed framework is its ability to accommodate non-Gaussian and time-varying properties of the variance, skewness, and kurtosis in the nowcast density estimates. These dynamic features are essential in improving the accuracy of density nowcasts for headline inflation.

Our framework allows us to incorporate a range of density combination methods proposed in the literature in a comprehensive empirical examination. Our examination reveals how the densities are combined matters: not all the combination methods improve accuracy compared to the best-performing individual densities. Hence, our findings should serve as a guide to practitioners on which combination methods may or may not work for nowcasting US inflation.

We also show in a horse race with the survey of professional forecasters that the density nowcasts from our framework provide superior (implied) point and density nowcasts for CPI inflation and PCE inflation. For core inflation, the performance of our framework is competitive to the survey. The ability of our proposed framework to generate a highly accurate point and density nowcasts of inflation is a valuable outcome for practitioners.

Chapter 3 focuses on inflation because recent work on *density nowcasting (based on mixed-frequency models)* has generally focused on real GDP growth and other indicators of real economic activity. The empirical work of chapter 2 and Kruger, Clark, and Ravazzolo (2017) demonstrate conditioning quarterly macroeconomic models with nowcast means and nowcast densities (informed from external sources) leads to improvements in the accuracy of multistep point and density forecasts, especially inflation. The framework developed in chapter 3 provides a potential source of both point and density nowcasts of US inflation that is shown to be more accurate than survey nowcasts, which is a popular source for obtaining nowcast estimates.

Chapter 4 takes on the challenge of jointly estimating several macroeconomic stars (i.e., steady states) simultaneously, using a semi-structural time series model (aka a multivariate UC model with a particular structure informed by economic theory) estimated with Bayesian methods. Specifically, we consider the equilibrium rate of productivity growth (p-star), the level of potential output (gdp-star), the growth rate of potential output (g-star), the equilibrium level of the unemployment rate (u-star), the equilibrium level of the real rate of interest (r-star), the equilibrium rate of price inflation (pi-star), and the equilibrium rate of nominal wage inflation (w-star). For each star, we formulate a rich structure whose elements are guided by past research. In doing so, we contribute to the literature in six important ways.

First, we develop a multivariate UC model that provides direct estimates for all the stars of most relevance to macroeconomic policymakers. Previous research based on UC models to estimate time-varying stars has focused on smaller systems, often just two or three observables, and have relied on a minimal structure. In principle, proceeding with the joint estimation of a

framework that explicitly models the objects of interest and permits interactions among them (e.g., stars) has the potential to provide more reliable estimates of the objects than approaches that ignore them. Our results indicate that specification choices matter for all stars considered, and there are payoffs to modeling them jointly. For r-star and u-star, the choice of specification matters the most, and least for g-star. We find that joint modeling g-star and u-star is sufficient; however, to obtain the most precise estimates of g-star (and output gap), there are gains to using a larger model. Generally, the *contours* of our stars echo those documented elsewhere in the literature, but at times are different, and these differences can matter for policy. We also find that our model estimates the output gap similar to the CBO estimate based on a production function approach. This result provides evidence supporting the common practice of using the output gap estimates from the CBO as an exogenous variable in empirical macroeconomic models. We view this result as a valuable contribution to the applied macro literature.

Second, in our model, we allow for time-variation in the key macroeconomic relationships and the error variances (aka stochastic volatility). To the best of our knowledge, features such as time-varying parameters and stochastic volatility (SV) have not been implemented in a UC framework beyond a system consisting of at most four variables. In principle, these empirical features should help better distinguish between cyclical and idiosyncratic fluctuations and lower-frequency movements in the macroeconomic aggregates considered in chapter 4. Our results indicate economically and statistically significant evidence of time-variation in macroeconomic relationships and strong support for SV's inclusion in our model equations. It lends support to the popular narrative “(price) Phillips curve has weakened over time,” “wage Phillips curve is alive,” and “weakening in the procyclicality of labor productivity.”

Third, we extend the Chan, Clark, and Koop (2018) approach of using survey expectations to improve pi-star precision to other macroeconomic stars. Specifically, we explicitly model the link between the unobserved “star” and the expectations about the companion star in the BlueChip survey of economic forecasters for each macroeconomic variable of interest. By bringing in additional information from survey expectations, we find significant improvements in the precision of the stars' estimates. Fourth, we provide a model to obtain real-time estimates of w-star (the long-run equilibrium level of nominal wage inflation) and its determinants. Our choice of a particular specification for w-star permits a time-varying model-based decomposition of w-star into its determinants, p-star (long-run productivity) and pi-star (long-run inflation), as implied by economic theory. This specific decomposition is helpful to monetary policymakers, who often refer to developments in nominal wages to support their forecasts and in related discussions on price inflation and employment.

Fifth, our empirical model can generate competitive real-time point and density forecasts of macroeconomic variables of broader interest. Sixth, the estimates of the stars from our model can potentially be used to inform the steady state values for reduced-form VAR models (as in chapter 2) and or Dynamic Stochastic General Equilibrium (DSGE) models.

5.2 Further research

As is the case with most research, there are several aspects of my thesis that could be explored further and are left for future research. Regarding extensions related to chapter 2, it will be interesting to examine the efficacy of the “hybrid” proposal combining survey forecasts and model-based forecasts to other countries. In fact, recently, Ganics and Odendahl (2021) apply a slightly modified proposal (to the one specified in chapter 2 of this thesis) on euro area data and find support for the efficacy of the proposal. Another extension is to tilt the model forecasts to first, second, and third moments extracted from raw density forecasts of the US SPF. This extension is motivated by Galvão, Garratt, and Mitchell (2021), who apply such an approach to the UK data. Another extension worth exploring is whether our proposal, when applied to density combination, constructed by combining the ten individual forecast densities corresponding to the ten BVAR models (of chapter 2), yields more accurate density (and implied point) forecast accuracy compared to the individual hybrid forecasts. Both, Galvão, Garratt, and Mitchell (2021) and Banbura et al. (2021) include density combinations in their empirical exercises combining survey information with model-based forecasts via entropic tilting; the former focused on UK data, and the latter on euro area data. Another extension worth exploring is to expand the model space to include nonlinear VARs, such as threshold VAR, which produce non-normal density forecasts that the technique of relative entropy can conveniently tilt.

In the case of chapter 3, at least three extensions are worth considering. First, the model space can be expanded to include the mixed-frequency VAR, patterned along the lines of Schorfheide and Song (2015) but with the information set similar to Clark and Zaman (2011). At present, all three mixed-frequency model classes are based on single-equation models, and expanding the model space to include VAR specifications, which allow for feedback between inflation and its determinants and vice versa, would enrich the framework. Following a similar argument, expanding the model space to consider machine learning models such as those featured in Medeiros et al. (2021), including random forest, would be another fruitful avenue. Second, augmenting model information set with online price indices entertained in Aparicio and Bertolotto (2020) is another interesting extension. These online indices may prove helpful in providing more timely signals when the economy is undergoing unexpected shifts, such as the COVID-19 pandemic. Third, an alternate approach to obtaining a very flexible estimate of density nowcast is to use a single model (e.g., Deterministic Model Switching of Knotek and Zaman, 2017) and assume *Student-t* distributed errors, which is modeled as a mixture of normal. Monache and Petrella (2017) provide evidence supporting models featuring *Student-t* distributed errors for inflation density forecasting. One would expect this favorable evidence to carry over to inflation density nowcasting. Hence, it will be interesting to compare which of the two approaches deliver more accurate density nowcasts of inflation: chapter 3’s density combination based on combining evidence from many models versus a single model with *Student-t* distributed errors. This latter extension is the focus of the ongoing work by Knotek and Zaman (in-progress).

Regarding chapter 4, again, there are three extensions worth exploring. First, instead of taking the shadow federal funds rate as given, joint estimation of the shadow federal funds rate and other model components would be beneficial. For example, the MCMC algorithm of Johansson and Mertens (forthcoming), based on sampling with censored data, could be integrated within the MCMC sampler developed in chapter 4. Given the UC model’s complexity, this extension would be non-trivial. Second, the Federal Reserve has recently updated its statement on the longer-run goals and monetary policy strategy. And based on the revised strategy, the longer-run goal of maximum employment would now be assessed with broad-based progress in the labor markets. This development reduces the emphasis on the estimates of the natural rate of unemployment but requires estimating a long-run equilibrium employment rate (or maximum employment or employment “star”). The UC model developed in chapter 4 could be extended to include a range of labor market variables to estimate an (unobserved) employment *star*. However, to keep the estimation tractable, increasing the model size would require difficult decisions on removing or shutting time-variation in some of the model equations. Third, VAR models are an alternate approach to modeling jointly many different variables and estimating the long-run equilibrium values. The advantage of the VAR-based framework is the ability to handle larger amounts of information conveniently and flexibly than UC models (see Morley and Wong, 2020; Chan, 2019).

On the other hand, as discussed in Chan, Koop, and Potter (2016), the advantage of UC modeling is the availability of the direct estimates of stars, which in our case proves quite convenient to allow for explicit modeling of the relationships between various stars. In VARs, the long-run estimates are implied estimates; therefore, imposing long-run restrictions is not as straightforward. That said, it will be interesting to compare the estimates of the stars using a flexible VAR model as in Chan (2019), but with the same observables as used in chapter 4.

Bibliography

- [1] Aastveit, K. A., Carriero, A., Clark, T. E., & Marcellino, M. (2017). Have standard VARs remained stable since the crisis? *Journal of Applied Econometrics*, 32(5), 931-951.
- [2] Aastveit, K. A., Foroni, C., & Ravazzolo, F. (2017). Density forecasts with Midas models. *Journal of Applied Econometrics*, 32 (4): 783–801. <https://doi.org/10.1002/jae.2545>.
- [3] Aastveit, K. A., Gerdrup, K. R., Jore, S. A., & Thorsrud, L. A. (2014). Nowcasting GDP in real time: A density combination approach. *Journal of Business and Economic Statistics*, 32 (1): 48–68. <https://doi.org/10.1080/07350015.2013.844155>.
- [4] Aastveit, K. A., Mitchell, J., Ravazzolo, F., & Herman van Dijk (2018). The evolution of forecast density combinations in economics. 18-069/III. Tinbergen Institute Discussion Papers. Tinbergen Institute. <https://ideas.repec.org/p/tin/wpaper/20180069.html>.
- [5] Allan, G., Koop, G., McIntyre, S., & Smith, P. (2014). Nowcasting Scottish GDP growth. 1411. Working Papers. University of Strathclyde Business School, Department of Economics. <https://ideas.repec.org/p/str/wpaper/1411.html>.
- [6] Altavilla, C., Giacomini, R., & Ragusa, G. (2017). Anchoring the yield curve using survey expectations. *Journal of Applied Econometrics*, 32(6), 1055-1068.
- [7] Alessandri, P., & Mumtaz, H. (2017). Financial conditions and density forecasts for US output and inflation. *Review of Economic Dynamics*, 24, 66-78.
- [8] Amisano, G., & Geweke, J. (2011). Optimal prediction pools. *Journal of Econometrics*, 164 (1): 130–41. <https://doi.org/10.1016/j.jeconom.2011.02.017>.
- Amisano, G., & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics*, 25(2), 177-190
- [9] Andersson, M., Palmqvist, S., & Waggoner, D. (2010). *Density-conditional forecasts in dynamic multivariate models*. Sveriges Riksbank Working paper 247.
- [10] Ang, A., Bekaert, G., & Wei, M. (2016). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, 54(4), 1163-1212
- [11] Antolin-Diaz, J., Drechsel, T., & Petrella, I. (2017). Tracking the slowdown in long-run GDP growth. *Review of Economics and Statistics*, 99(2): 343–356
- [12] Antolin-Diaz, J., Petrella, I., & Rubio-Ramirez, J. F. (2019). *Structural scenario analysis with SVARs*. *Journal of Monetary Economics*.
- [13] Aparicio, D., & Bertolotto, M. I. (2020). Forecasting inflation with online prices. *International Journal of Forecasting*, 36(2): 232-247

- [14] Ashley, R., & Verbrugge, R. J. (2017). Frequency dependence in regression model coefficients: an alternative approach for modeling nonlinear dynamic relationships in time series. *Econometric Reviews*, 28(1–3): 4–20
- [15] Atkeson, A., & Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, Winter, 2-11.
- [16] Aucejo, E. M., French, J., Araya, M. P., & Zafar, B. (2020). The impact of COVID-19 on student experiences and expectations: evidence from a survey. *Journal of Public Economics*, 191: 104271. <https://doi.org/10.1016/j.jpubeco.2020.104271>.
- [17] Bache, W. I., Jore, S. A., Mitchell, J., & Vahey, S. P. (2011). Combining VAR and DSGE forecast densities. *Journal of Economic Dynamics and Control*, 35 (10): 1659–70. <https://doi.org/10.1016/j.jedc.2011.04.006>.
- [18] Bańbura, M., & Bobeica, E. (2020). Does the Phillips curve help to forecast euro area inflation? European Central Bank Working Paper Series No. 20202471
- [19] Bańbura, M., Brenna, F., Paredes, J., & Ravazzolo, F. (2021). Combining Bayesian VARs with survey density forecasts: does it pay off? European Central Bank Working Paper Series No. 2543
- [20] Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25, 71-92.
- [21] Bańbura, M., Giannone, D., & Lenza, M. (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting*, 31(3), 739-756.
- [22] Bańbura, M., & Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29 (1): 133–60.
- [23] Barbarino, A., Berge, T. J., Chen, H., & Stella, A. (2020). Which output gap estimates are stable in real time and why? Finance and Economics Discussion Series 2020-102. Washington: Board of Governors of the Federal Reserve System
- [24] Barnett, A., Mumtaz, H., & Theodoridis, K. (2014). Forecasting UK GDP growth and inflation under structural change: a comparison of models with time-varying parameters. *International Journal of Forecasting*, 30, 129-143.
- [25] Barnichon, R., & Mesters, G. (2018). On the demographic adjustment of unemployment. *Review of Economics and Statistics*, 100: 219-231.
- [26] Basistha, A., & Nelson, C. R. (2007). New measures of the output gap based on the forward-looking new Keynesian Phillips curve. *Journal of Monetary Economics*, 54:498-511.
- [27] Bates, J. M., & Granger, C. (1969). The combination of forecasts. *Operational Research Quarterly*, 20 (4): 451–68.
- [28] Bauer, M. D., & Rudebusch, G. D. (2020). Interest rates under falling stars. *American Economic Review*, 110(5): 1316:1354.
- [29] Bauwens, L., Lubrano, M., & Richard, J. F. (1999). *Bayesian inference in dynamic econo-*

- metric models*. Oxford University Press.
- [30] Beckmann, J., Koop, G., Korobilis, D., & Schüssler, R. A. (2020). Exchange rate predictability and dynamic Bayesian learning. *Journal of Applied Econometrics*, 35 (4): 410–21.
- [31] Benati, L. (2007). Drift and breaks in labor productivity. *Journal of Economic Dynamics & Control*, 31(8): 2847-2877.
- [32] Berger, T., Evaraert, G., & Vierke, H. (2016). Testing for time variation in an unobserved components model for the U.S. economy. *Journal of Economic Dynamics & Control*, 179-208.
- [33] Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics*, 19: 465-474.
- [34] Beveridge, S., & Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics*, 7:151-174
- [35] Beyer, R. C.M., & Wieland, V. (2019). Instability, imprecision and inconsistent use of equilibrium real interest rate estimates. *Journal of International Money and Finance*, 94:1–14
- [36] Bjørnland, H. C., Gerdrup, K., Jore, S. A., Smith, C. & Thorsrud, A. L. (2011). Weights and pools for a Norwegian density combination. *The North American Journal of Economics and Finance*, 22 (1): 61–76.
- [37] Bobeica, E., Ciccarelli, M., & Vansteenkiste, V. (2019). The link between labor cost and price inflation in the euro area. European Central Bank Working Paper 2235
- [38] Brand, C., & Mazelis, F. (2019). Taylor-rule consistent estimates of the natural rate of interest. European Central Bank Working Paper 2257
- [39] Breitung, J., & Roling, C. (2015). Forecasting inflation rates using daily data: A nonparametric MIDAS approach: Nonparametric MIDAS forecasting. *Journal of Forecasting*, 34 (7): 588–603.
- [40] Carlstrom, C. T., & Zaman, S. (2014). Using an Improved Taylor Rule to Predict When Policy Changes Will Occur. Federal Reserve Bank of Cleveland Economic Commentary, No. 2014-02
- [41] Carriero, A., Clark, T. E., & Marcellino, M. (2015a). Bayesian VARs: specification choices and forecast accuracy. *Journal of Applied Econometrics*, 30(1), 46-73.
- [42] Carriero, A., Clark, T. E., & Marcellino, M. (2015b). Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178 (4): 837–62.
- [43] Carriero, A., Clark, T. E., & Marcellino, M. (2016). *Large vector autoregressions with stochastic volatility and flexible priors*. Federal Reserve Bank of Cleveland Working Paper, no. 16-17.
- [44] Carriero, A., Clark, T. E., & Marcellino, M. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, 212: 137–154

- [45] Carriero, A., Clark, T. E., Marcellino, M., & Mertens, E. (2021). Addressing COVID-19 outliers in BVARs with stochastic volatility. Federal Reserve Bank of Cleveland, Working Paper No.21-02
- [46] Chan, J. (2013). Moving average stochastic models with application to inflation forecast. *Journal of Econometrics*, 176: 162-172
- [47] Chan, J. (2017). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics*, 35(1): 17-28
- [48] Chan, J. (2019). Large hybrid time-varying parameter VARs. CAMA paper 77
- [49] Chan, J., Clark, T. E., & Koop, G. (2018). A new model of inflation, trend inflation, and long-run inflation expectations. *Journal of Money, Credit, and Banking*, 50(1): 5-53
- [50] Chan, J., & Eisenstat, E. (2018a). Comparing hybrid time-varying parameter VARs. *Economics Letters*, 171:1-5
- [51] Chan, J., & Eisenstat, E. (2018b). Bayesian model comparison for time-varying parameter VARs with stochastic volatility. *Journal of Applied Econometrics*, 33(4): 509-532
- [52] Chan, J., & Grant, A. L. (2017). A Bayesian model comparison for trend-cycle decompositions of output. *Journal of Money, Credit and Banking* 49(2-3): 525-552
- [53] Chan, J., & Jeliaskov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modeling and Numerical Optimisation* 1: 101-120
- [54] Chan, J., & Koop, G. (2014). Modeling breaks and clusters in the steady states of macroeconomic variable. *Computational Statistics and Data Analysis*, 76, 186-193
- [55] Chan, J., Koop, G., & Potter, S. M. (2013). A new model of trend inflation. *Journal of Business and Economic Statistics* 31: 94-106
- [56] Chan, J., Koop, G., & Potter, S. M. (2016). A bounded model of time variation in trend inflation, NAIRU and the Phillips curve. *Journal of Applied Econometrics* 31: 551-565
- [57] Chan, J., Koop, G., Poirier, D. J., & Tobias J. L. (2019). Bayesian econometric methods 2nd edition. *Cambridge University Press*
- [58] Chan, J., & Strachan, R. (2012). Estimation in non-linear non-Gaussian state-space models with precision-based methods. CAMA working paper series 2012-13.
- [59] Chernick, M. R., & LaBudde, R. A. (2011). An introduction to bootstrap methods with applications to R. John Wiley and Sons.
- [60] Chernis, T., & Sekkel, R. (2018). Nowcasting Canadian economic activity in an uncertain environment. 2018–9. Staff Discussion Paper. Bank of Canada.
- [61] Christiano, L. J., & Fitzgerald, T. J. (2003). The band pass filter. *International Economic Review* 44:435-465
- [62] Clark, Peter K. (1987). The cyclical component of U.S. economic activity. *Quarterly Journal of Economics*, 10(4): 797-814
- [63] Clark, T. E. (2011). Real-Time density forecasts From Bayesian vector autoregressions

- with stochastic volatility. *Journal of Business and Economic Statistics*, 29 (3): 327–41
- [64] Clark, T. E. (2014). The importance of trend inflation in the search for missing disinflation. Federal Reserve Bank of Cleveland *Economic Commentary*, No. 2014-16
- [65] Clark, T. E., & Doh, T. (2014). Evaluating alternative models of trend inflation. *International Journal of Forecasting* 30: 426-448
- [66] Clark, T. E., & Kozicki, S. (2005). Estimating equilibrium real interest rates in real time. *North American Journal of Economics and Finance* 16: 395-413
- [67] Clark, T. E., & McCracken, M. W. (2008). Forecasting with small macroeconomic VARs in the presence of instabilities. In D. Rapach & M. Wohar (eds.), *Forecasting in the presence of structural breaks and model uncertainty* (pp. 93-147). Emerald Publishing.
- [68] Clark, T. E., & McCracken, M. W. (2010). Averaging forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics*, 25(1), 5-29.
- [69] Clark, T. E., & McCracken, M. W. (2013). Advances in forecast evaluation. In G. Elliott & A. Timmermann (eds.), *Handbook of economic forecasting* (vol. 2, pp. 1107-1201). Amsterdam, the Netherlands: Elsevier.
- [70] Clark, T. E., & McCracken, M. W. (2017). Tests of predictive ability for vector autoregressions used for conditional forecasting. *Journal of Applied Econometrics*, 32, 533-553.
- [71] Clark, T. E., McCracken, M. W. & Mertens, E. (2018). Modeling time-varying uncertainty of multiple-horizon forecast errors. *Review of Economics and Statistics*
- [72] Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30, 551-575
- [73] Clark, T. E., & Zaman, S. (2011). Food and energy price shocks: What other prices are affected? Federal Reserve Bank of Cleveland Economic Commentary 2011-14
- [74] Clements, M. P. (2014). Forecast uncertainty-ex ante and ex post: U.S. inflation and output growth. *Journal of Business and Economic Statistics*, 32, 206-216.
- [75] Clements, Michael P (2018). Are macroeconomic density forecasts informative? *International Journal of Forecasting*, 34(2): 181–98
- [76] Clements, M. P., & Galvão, B. A. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business and Economic Statistics*, 26 (4): 546–54.
- [77] Clements, M. P., & Hendry, D. F. (1999). *Forecasting non-stationary economic time series*. MIT Press: London.
- [78] Cogley, T., Morozov, S., & Sargent, T. J. (2005). Bayesian fan charts for U.K. inflation: forecasting and sources of uncertainty in an evolving monetary system. *Journal of Economic Dynamics and Control*, 29, 1893-1925.
- [79] Cogley, T., Primiceri, ., & Sargent, T. J. (2010). Inflation-gap persistence in the U.S. *American Economic Journal: Macroeconomics* 2: 43-69
- [80] Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes

- in the post World War II U.S. *Review of Economic Dynamics*, 8, 262-302.
- [81] Coibion, O., Gorodnichenko, Y., & Ulate, M. (2018). The cyclical sensitivity in estimates of potential output. *Brookings Papers on Economic Activity*, Fall 2018
- [82] Conflitti, C., De Mol, C., & Giannone, D. (2015). Optimal combination of survey forecasts. *International Journal of Forecasting*, 31 (4): 1096–1103
- [83] Cross, R., Darby, J., & Ireland, J. (2005). Uncertainties surrounding natural rate estimates in the G7. *Advances in Computational Economics*, Quantitative Economic Policy Vol 20: 337-361
- [84] Croushore, D. (2010). An evaluation of inflation forecasts from surveys using real-time data. *BE Journal of Macroeconomics*, 10(1), 1-32.
- [85] Crump, R. K., Eusepi, S., Giannoni, M., & Sahin, A. (2019). A unified approach to measuring u^* . *Brookings Papers on Economic Activity*
- [86] D’Agostino, A., Giannone, D., & Gambetti, L. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28(1), 81-101.
- [87] Daly, M., & Hobijn, B. (2014). Downward nominal wage rigidities bend the Phillips curve *Journal of Money, Credit and Banking* 46(2): 51-93
- [88] Daly, M., Hobijn, B., Sahin, A., & Valletta, R. G. (2012). A search and matching approach to labor markets: did the natural rate of unemployment rise?. *Journal of Economic Perspectives* 26: 3-26
- [89] Davidson, R., & Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, 146(1): 162-69
- [90] Davidson, R., & MacKinnon, J. G. (2006). Bootstrap methods in econometrics. In Kerry Patterson and Terence C. Mills, eds. *Palgrave Handbook of Econometrics: Vol 1, Econometric Theory*.
- [91] Davis, S. J., Faberman, J. R., & Haltiwanger, J. (2013). The establishment-level behavior of vacancies and hiring. *Quarterly Journal of Economics* 128(2): 581-622
- [92] Del Negro, M., Giannoni, M., & Schorfheide, F. (2015). Inflation in the great recession and New Keynesian models. *American Economic Journal: Macroeconomics*, 7:168-196
- [93] Del Negro, M., Giannone, D., Giannoni, M., & Tambalotti, A. (2017). Safety, liquidity, and the natural rate of interest. *Brookings Papers on Economic Activity*
- [94] Del Negro, M., Giannoni, M., Lenza, M., & Tambalotti, A. (2020). What’s up with the Phillips curve?. *Brookings Papers on Economic Activity*
- [95] Del Negro, M., & Primiceri, G. E. (2015). Time varying structural vector autoregressions and monetary policy: a corrigendum. *The review of economic studies*, 82(4), 1342-1345.
- [96] Del Negro, M., & Schorfheide, F. (2013). DSGE model-based forecasting. In *Handbook of Economic Forecasting*, Vol. 2, Part A, edited by Graham Elliott and Allan Timmermann, pp. 57–140.
- [97] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253-263.

- [98] Diebold, F. X., Gunther, T. A. & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39 (4): 863–83
- [99] Doan, T., Litterman, R., & Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3, 1-100.
- [100] Domit, S., Monti, F., & Sokol, A. (2019). Forecasting the UK economy with a medium-scale Bayesian VAR. *International Journal of Forecasting*, in press
- [101] Durlauf, S. N., and Vahey, S. P. (2010). Introduction: Model uncertainty and macroeconomics. *Journal of Applied Econometrics*, 25: 1-3
- [102] Edge, R. M., Laubach, T., & Williams, J.C. (2007). Learning and shifts in long-run productivity growth. *Journal of Monetary Economics* 54: 2421-2438
- [103] Eggertsson, G. B., Mehrotra, N. R., Singh, S. R., & Summers, L. H. (2016). A contagious malady? Open economy dimensions of secular stagnation. *IMF Economic Review* 64 (4): 581–634
- [104] Faust, J., & Wright, J. (2009). Comparing Greenbook and reduced form forecasts using a large realtime dataset. *Journal of Business and Economic Statistics*, 24(4): 468-479
- [105] Faust, J., & Wright, J. (2013). Forecasting inflation. In *Handbook of Economic Forecasting*, Vol. 2, edited by Graham Elliot and Allan Timmermann, pp: 3-56
- [106] Fernald, J. G. (2007). Trend breaks, long-run restrictions, and contractionary technology improvements. *Journal of Monetary Economics*, 54(8): 2467-2485
- [107] Fernald, J. G. & Wang, C. (2016). Why has the cyclical productivity changed? What does it mean? *Annual Review of Economics*, 8(1): 465–96
- [108] Fernandez-Villaverde, J., & Rubio-Ramirez, J. (2010). Macroeconomics and volatility: data, models, and estimation. NBER Working Papers, 16618, DOI 10.3386/w16618
- [109] Fleischman, C. A., & Roberts, J. M. (2011). From many Series, one cycle: improved estimates of the business cycle from a multivariate unobserved components model. Federal Reserve Board Working paper no 2011-46
- [110] Frey, C., & Mokiniski, F. (2016). Forecasting with Bayesian vector autoregressions estimated using professional forecasts. *Journal of Applied Econometrics*, 31:1083-1099
- [111] Gali, J. (2011). The return of the wage Phillips Curve. *Journal of the European Economic Association*, 9(3): 436-461
- [112] Gali, J., & Gambetti, L. (2019). Has the U.S. wage Phillips curve flattened? A semi-structural exploration. *NBER Working Paper series*, No. 25476
- [113] Gali, J., & van Rens, T. (2020). The vanishing procyclicality of labour productivity. *The Economic Journal*, pp 1-25
- [114] Galvão, A. B., Garratt, A., & Mitchell, J. (2021). Does judgment improve macroeconomic density forecasts? *International Journal of Forecasting*, 37(3): 1247-1260
- [115] Ganics, G. A. (2017). Optimal density forecast combinations. 1751. Banco de España Working Paper 1751. <https://ideas.repec.org/p/bde/wpaper/1751.html>.
- bibitemlatexcompanion Ganics, G. A., & Odendahl, F. (2021). Bayesian VAR forecasts,

- survey information, structural change in the euro area. *International Journal of Forecasting*, 37(2): 971-999
- [116] Ganics, G. A., Rossi, B. & Sekhposyan, T. (2019). From fixed-event to fixed-horizon density forecasts: professional forecasters' view on multi-horizon uncertainty. Banco de Espana Working Paper No. 1947.
- [117] Garcia, J. A., & Poon, A. (2018). Trend inflation and inflation compensation. IMF Working Paper series, WP/18/154
- [118] Garratt, A., Henckel, T. & Vahey, S. P. (2019). Empirically-Transformed linear opinion pools. 2019–47. CAMA Working Papers. Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.
- [119] Garratt, A., Mitchell, J. & Vahey, S. P. (2014). Measuring output gap nowcast uncertainty. *International Journal of Forecasting*, 30 (2): 268–79.
- [120] Garratt, A., Mitchell, J., Vahey, S. P. & Wakerly, E. C. (2011). Real-Time inflation forecast densities from ensemble Phillips curves. *The North American Journal of Economics and Finance*, 22 (1): 77–87.
- [121] Gerard, H., & Nimark, K. (2008). Combining multivariate density forecasts using predictive criteria. rdp2008-02. RBA Research Discussion Papers. Reserve Bank of Australia
- [122] Geweke, J. (1992). Evaluating the accuracy of sample-based approaches to the calculation of posterior moments, in *Bayesian Statistics 4* (ed. J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith), 641-649. Oxford: Clarendon Press.
- [123] Ghysels, E., Santa-Clara, P., & Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, 76 (3): 509–48. <https://doi.org/10.1016/j.jfineco.2004.03.008>.
- [124] Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131 (1–2): 59–95
- [125] Giacomini, R., & Ragusa, G. (2014). Theory-coherent forecasting. *Journal of Econometrics*, 182, 145-155.
- [126] Giannone, D., Lenza, M., & Primiceri, G. E. (2019). Priors for the long run. *Journal of the American Statistical Association*, 114(526), 565-580.
- [127] Giannone, D., Reichlin, L. & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55 (4): 665–76
- [128] Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, predictions, and estimations. *Journal of the American Statistical Association*, 102, 359-378.
- [129] Gonzalez-Astudillo, M., & Laforte, J. (2020). Estimates of r^* consistent with a supply-side structure and a monetary policy rule for the U.S. economy. Finance and Economics Discussion Series Paper 2020-085.
- [130] Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). A novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE proceedings F (Radar and signal*

- processing*) (Vol. 140, No. 2, pp. 107-113). IET Digital Library.
- [131] Grant, A. L., & Chan, J. (2017). Reconciling output gaps: unobserved components model and Hodrick-Prescott filter. *Journal of Economic Dynamics and Control* 114-121
- [132] Hall, S. G., & Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23 (1): 1–13.
- [133] Hamilton, J., Harris, E., Hatzius, J., & West, K. (2016). The equilibrium real funds rate: past, present and future. *IMF Economic Review*, 64(4): 660-707
- [134] Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13: 281-291.
- [135] Hendry, D. F., & Hubrich, K. (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business and Economic Statistics*, 29 (2): 216–27
- [136] Hodrick, R. J., & Prescott, E. C. (1997). Postwar US business cycles: an empirical investigation *Journal of Money, Credit and Banking*, 29:1-16.
- [137] Holston, K., Laubach, T. & Williams, J.. (2017). Measuring the natural rate of interest: international trends and determinants. *Journal of International Economics*, 108: 59-75.
- [138] Jacobs, J. P.A.M., & van Norden, S. (2016). Why are initial estimates of productivity growth so unreliable? *Journal of Macroeconomics*, 47:200-213
- [139] Johannsen, B. K., & Mertens, E. (2015). The shadow rate of interest, macroeconomic trends, and time-varying uncertainty. Unpublished manuscript
- [140] Johannsen, B. K., & Mertens, E. (forthcoming). A time series model of interest rates with the effective lower bound. *Journal of Money, Credit, and Banking*
- [141] Jorda, O., Singh, S. R., & Taylor, A. M. (2020). Longer-Run economic consequences of pandemics. San Francisco Fed Working paper 2020-09
- [142] Jore, A. S., Mitchell, J., & Vahey, S. P. (2010). Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics*, 25 (4): 621–34.
- [143] Kahn, J. A., & Rich, R. W. (2007). Tracking the new economy: Using growth theory to detect changes in trend productivity. *Journal of Monetary Economics* 54: 1670-1701
- [144] Kamber, G., Morley, J., & Wong, B. (2018). Intuitive and reliable estimates of the output gap from a Beveridge-Nelson filter. *Review of Economics and Statistics* 100(3): 550-566
- [145] Kascha, C., & Ravazzolo, F. (2010). Combining inflation density forecasts. *Journal of Forecasting*, 29 (1–2): 231–50.
- [146] Kiley, M. T. (2020). What can data tell us about the equilibrium real interest rate? *International Journal of Central Banking*, June 2020
- [147] Knotek, E. S. (2007). How useful is Okun’s Law. Federal Reserve Bank of Kansas City *Economic Review*, Fourth Quarter 2007
- [148] Knotek, E. S. (2015). Difficulties forecasting wage growth. Federal Reserve Bank of Cleveland *Economic Trends*
- [149] Knotek, E. S., & Zaman, S. (2014). On the relationships between wages, prices, and

- economic activity. Federal Reserve Bank of Cleveland *Economic Commentary*, No. 2014-14
- [150] Knotek, E. S., & Zaman, S. (2017). Nowcasting U.S. headline and core inflation. *Journal of Money, Credit and Banking*, 49 (5): 931–68
- [151] Knotek, E. S., & Zaman, S. (2019). Financial nowcasts and their usefulness in macroeconomic forecasting. *International Journal of Forecasting*, 35 (4): 1708–24
- [152] Knotek, E. S., Zaman, S., & Clark, T. E. (2015). Measuring inflation forecast uncertainty. *Economic Commentary* (Federal Reserve Bank of Cleveland), March, 1–6.
- [153] Knüppel, M. (2015). Evaluating the calibration of multi-step-ahead density forecasts using raw moments. *Journal of Business and Economic Statistics*, 33 (2): 270–81
- [154] Knüppel, M., & Krüger, F. (2019). Forecast uncertainty, disagreement, and the linear pool. 28/2019. Discussion Papers. Deutsche Bundesbank.
- [155] Koop, G. (2003). Bayesian econometrics. *Chichester: Wiley*.
- [156] Koop, G. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28, 177-203.
- [157] Koop, G., & Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3, 267-358
- [158] Koop, G., & Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, 53 (3): 867–86
- [159] Koop, G., & Korobilis, D. (2013). Large time-varying parameter VARS. *Journal of Econometrics*, 177, 185-198.
- [160] Koop, G., Leon-Gonzalez, R. & Strachan, R. W. (2010). Dynamic probabilities of restrictions in state space models: An application to the Phillips curve. *Journal of Business and Economic Statistics* 28(3): 370-379
- [161] Koop, G., McIntyre, S., & Mitchell, J. (2020). UK regional nowcasting using a mixed frequency vector auto-regressive model with entropic tilting. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183 (1): 91–119
- [162] Korobilis, D. (2017). Quantile regression forecasts of inflation under model uncertainty. *International Journal of Forecasting* 33(1): 11-20
- [163] Kozicki, S., & Tinsley, P. A. (1998). Moving endpoints and the internal consistency of agents' ex ante forecasts. *Computational Economics*, 11, 21-40.
- [164] Kozicki, S., & Tinsley, P. A. (2001a). Shifting endpoints in the term structure of interest rates. *Journal of Monetary Economics*, 47, 613-652.
- [165] Kozicki, S., & Tinsley, P. A. (2001b). Term structure views of monetary policy under alternative models of agent expectations. *Journal of Economic Dynamics and Control*, 25, 149-184.
- [166] Kozicki, S., & Tinsley, P. A. (2012). Effective use of survey information in estimating the evolution of expected inflation. *Journal of Money, Credit and Banking*, 44(1): 145-169
- [167] Krüger, F. (2015). Combining density forecasts under various scoring rules: An analysis of UK inflation. Manuscript.

- [168] Krüger, F., Clark, T. E., & Ravazzolo, F. (2017). Using entropic tilting to combine BVAR forecasts with external nowcasts. *Journal of Business and Economic Statistics*, 35 (3): 470–85
- [169] Kuttner, K. N. (1994). Estimating potential output as a latent variable. *Journal of Business and Economic Statistics*, 12(3): 361-368
- [170] Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions — five years of experience. *Journal of Business and Economic Statistics*, 4, 25-38.
- [171] Laubach, T. (2001). Measuring the NAIRU: Evidence from seven economies. *The Review of Economics and Statistics*, 83(2): 218-231
- [172] Laubach, T., & Williams, J. C. (2003). Measuring the natural rate of interest. *The Review of Economics and Statistics*, 85(4): 1063-1070
- [173] Laubach, T., & Williams, J. C. (2016). Measuring the natural rate of interest redux. *Business Economics*, 51(2): 57-67
- [174] Leduc, S., & Liu, Z. (2020). Can pandemic-induced job uncertainty stimulate automation? San Francisco Fed Working paper 2020-19
- [175] Lee, J. & Nelson, C. R. (2007). Expectation horizon and the Phillips curve: the solution to an empirical puzzle. *Journal of Applied Econometrics* 22(1): 161-178
- [176] Lewis, K. F. & Vazquez-Grande, F. (2019). Measuring the natural rate of interest: A note on transitory shocks. *Journal of Applied Econometrics* 34: 425-436
- [177] Lubik, T. A., & Matthes, C. (2015). Calculating the natural rate of interest: A comparison of two alternative approaches. *Richmond Fed Economic Brief*
- [178] Lunsford, K. (2017). Productivity growth and real interest rates in the long run. Federal Reserve Bank of Cleveland *Economic Commentary*, No. 2017-20
- [179] Lunsford, K., & West, K. D. (2019). Some evidence on secular drivers of US safe real rates. *American Economic Journal: Macroeconomics*, AEA, 11(4): 113-139
- [180] Marcelo, M C., Vasconcelos, G. F., Veiga, A., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business and Economic Statistics*, 39(1):98-119
- [181] Marsilli, C. (2017). Nowcasting US inflation using a MIDAS augmented Phillips curve. *International Journal of Computational Economics and Econometrics*, 7 (1/2): 64
- [182] Mazzi, G. L., Mitchell, J., & Montana, G. (2014). Density nowcasts and model combination: Nowcasting euro-area GDP growth over the 2008-09 recession. *Oxford Bulletin of Economics and Statistics*, 76 (2): 233–56. <https://doi.org/10.1111/obes.12015>.
- [183] Mertens, E. (2016). Measuring the level and uncertainty of trend inflation. *The Review of Economics and Statistics*, 98: 950-967
- [184] Mertens, E., & Nason, J. M. (2020). Inflation and professional forecast dynamics: An evaluation of stickiness, persistence and volatility. *Quantitative Economics*, 11:1485-1520
- [185] Mester, L. J. (2018). Demographics and their implications for the economy and policy. *Cato Journal*, 38(2): 399-413

- [186] Modugno, M. (2013). Now-Casting inflation using high frequency data. *International Journal of Forecasting*, 29 (4): 664–75.
- [187] Monache, D. D., & Petrella, I. (2017). Adaptive models and heavy tails with an application to inflation forecasting. *International Journal of Forecasting*, 33: 482-501
- [188] Monteforte, L., & Moretti, G. (2013). Real-Time forecasts of inflation: The role of financial variables. *Journal of Forecasting*, 32 (1): 51–61.
- [189] Morley, J. (2002). A state-space approach to calculating the Beveridge-Nelson decomposition. *Economics Letters*, 75: 123-127
- [190] Morley, J, Nelson, C. R. & Eric Zivot (2003). Why are the Beveridge-Nelson and unobserved components decompositions of GDP so different?. *Review of Economics and Statistics*, 85: 235-243
- [191] Morley, J., & Piger, J. (2012). The asymmetric business Cycle. *Review of Economics and Statistics*, 94: 208-221
- [192] Morley, J., & Wong, B. (2020). Estimating and accounting for the output gap with large Bayesian vector autoregressions. *Journal of Applied Econometrics*, 35: 1-18
- [193] Nason, J. M., & Smith, G. W. (2020). Measuring the slowly evolving trend in US inflation with professional forecasts. *Journal of Applied Econometrics*
- [194] Orphanides, A., & Williams, J. C. (2002). Robust monetary policy rules with unknown natural rates. Working paper 2003-01, Federal Reserve Bank of San Francisco
- [195] Pauwels, L. L., & Vasnev, A. L. (2016). A note on the estimation of optimal weights for density forecast combinations. *International Journal of Forecasting*, 32 (2): 391–97
- [196] Peneva, E. V., & Rudd, J. B. (2017). The passthrough of labor costs to price inflation. *Journal of Money, Credit, and Banking*, 49(8): 1777-1802
- [197] Pescatori, A., & Turunen, J. (2016). Lower for longer: Neutral rate in the U.S. *IMF Economic Review*, 64(4): 708-731
- [198] Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72, 821-852
- [199] Powell, J. H. (2018). Changing market structure and implication for monetary policy. Speech at the Jackson Hole Symposium organized by the Federal Reserve Bank of Kansas City, August 24, 2018.
- [200] Ravazzolo, F., & Vahey, S. P. (2014). Forecast densities for economic aggregates from disaggregate ensembles. *Studies in Nonlinear Dynamics and Econometrics*, 18 (4): 367–81
- [201] Reifschneider, D., & Tulip, P. (2019). Gauging the uncertainty of the economic outlook using historical forecasting errors: The Federal Reserve’s approach. *International Journal of Forecasting*, 35 (4): 1564–82
- [202] Roberts, J. (2001). Estimates of the productivity trend using time-varying parameter techniques. *Contributions to Macroeconomics*, 1(1):3
- [203] Robertson, J. C., & Tallman, E. W. (1999). Vector autoregressions: forecasting and reality. *Economic Review: Federal Reserve Bank of Atlanta*, 84(1), 4.

- [204] Robertson, J. C., Tallman, E. W., & Whiteman, C. H. (2005). Forecasting using relative entropy. *Journal of Money, Credit and Banking*, 37(3), 383-401.
- [205] Rossi, B. (2014). Density forecasts in economics, forecasting and policymaking. 37. Els Opuscles Del CREI. Centre de Recerca en Economia Internacional (CREI)
- [206] Rossi, B., & Sekhposyan, T. (2014). Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set. *International Journal of Forecasting*, 30(3): 662–82
- [207] Schorfheide, F., & Song, D. (2015). Real-Time forecasting with a mixed-frequency VAR. *Journal of Business and Economic Statistics*, 33(3):366-380
- [208] Schorfheide, F., & Song, D. (2020). Real-Time forecasting with a (standard) mixed-frequency VAR during a pandemic. Philadelphia Fed Working paper 20-26.
- [209] Shackleton, R. (2018). Estimating and projecting potential output using CBO’s forecasting growth model. Congressional Budget Office, Working Paper 2018-03
- [210] Sinclair, T. M. (2009). The relationships between permanent and transitory movements in U.S. output and the unemployment Rate. *Journal of Money, Credit and Banking* 41(2-3): 529-542
- [211] Staiger, D. O., Stock, J. H., & Watson, M. W. (1997). How precise are estimates of the natural rate of unemployment?. In Romer, C, Romer, D. (Eds.), *Reducing Inflation: Motivation and Strategy*, pp. 195-242, University of Chicago Press
- [212] Stella, A., & Stock, J. H. (2015). A state-dependent model for inflation forecasting. In S. J. Koopman, & N. Shephard (Eds.), *Unobserved components and time series econometrics*, (pp, 14-29), Oxford University Press
- [213] Stock, J. H., & Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39: 3-33
- [214] Stock, J. H., & Watson, M. W. (2016). Core inflation and trend inflation. *Review of Economics and Statistics*, 98(4): 770-784
- [215] Stock, J. H., & Watson, M. W. (2020). Slack and cyclically sensitive inflation. *Journal of Money, Credit and Banking*, 52(S2): 393-428
- [216] Summers, L. H. (2014). US economic prospects: Secular stagnation, hysteresis and the zero lower bound. *Business Economics*, 49(2): 65-73
- [217] Tallman, E. W., & Zaman, S. (2017). Forecasting inflation: Phillips curve effects on services price measures. *International Journal of Forecasting*, 33(2): 442-457
- [218] Tallman, E. W., & Zaman, S. (2018). *Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy*. Federal Reserve Bank of Cleveland working paper 18-09
- [219] Tallman, E. W., & Zaman, S. (2020). Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy. *International Journal of Forecasting*, 36(2): 373-398
- [220] Tay, A. S., & Wallis, K. F. (2000). Density forecasting: A survey. *Journal of Forecasting*,

- 19 (4): 235–54.
- [221] Taylor, J. B. (1999). A historical analysis of monetary policy rules. *University of Chicago Press*, Chapter 7: 319-348
- [222] Taylor, J. B. & Wieland, V. (2016). Finding the equilibrium real interest rate in a fog of policy deviations. *Business Economics*, 51(3): 147-154
- [223] Tasci, M. (2012). The ins and outs of unemployment in the long run: Unemployment flows and the natural rate. Federal Reserve Bank of Cleveland Working paper
- [224] Tasci, M. (2019). Challenges with estimating u^* in real time. Federal Reserve Bank of Cleveland *Economic Commentary*, 2019-18
- [225] Tulip, P. (2009). Has the economy become more predictable? Changes in Greenbook forecast accuracy. *Journal of Money, Credit and Banking*, 41 (6): 1217–31
- [226] Villani, M. (2009). Steady-state priors for vector autoregressions. *Journal of Applied Econometrics*, 24, 630-650.
- [227] Waggoner, D. F., & Zha, T. (1999). Conditional forecasts in dynamic multivariate models. *Review of Economics and Statistics*, 81, 639-651.
- [228] Weber, A. A., Lemke, W., & Worms, A. (2008). How useful is the concept of the natural real rate of interest for monetary policy? *Cambridge Journal of Economics*, 32:49-63
- [229] Weise, C. L. (2012). Political pressures on monetary policy during the US Great Inflation. *American Economic Journal: Macroeconomics*, 4:33-64
- [230] West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64, 1067-1084.
- [231] Williams, J. C. (2018). The future fortunes of r^* : Are they really rising? Federal Reserve Bank of San Francisco *Economic Letter*, No. 2018-13
- [232] Wright, J. H. (2013). Evaluating real-time VAR forecasts with an informative democratic prior. *Journal of Applied Econometrics*, 28: 762-776 doi:10.1002/jae.2268
- [233] Wright, J. H. (2019). Some observations on forecasting and policy. *International Journal of Forecasting*, 35 (3): 1186–92.
- [234] Wu, J. C., & Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the ZLB. *Journal of Money, Credit, and Banking*, 48(2-3): 253-291
- [235] Zaman, S. (2013). Improving inflation forecasts in the medium to long term. Federal Reserve Bank of Cleveland *Economic Commentary*, No. 2013-16.
- [236] Zaman, S. (2019). Cyclical versus Acyclical Inflation: A Deeper Dive. Federal Reserve Bank of Cleveland *Economic Commentary*, No. 2019-13
- [237] Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. Wiley: New York.

Appendix A

Chapter 2 Appendix

Chapter 2: Combining Survey Long-Run Forecasts and Nowcasts with BVAR Forecasts Using Relative Entropy.

A0: Technical Appendix

Specifically, we impose our prior beliefs about the coefficient estimates in A_1, \dots, A_p and Σ using normal inverse-Wishart (N-IW) conjugate priors.¹ The prior beliefs for the means and variances of the coefficient matrices are as follows:

$$E[A_l^{(i,j)}] = \begin{cases} \delta_i & \text{if } i = j, l = 1 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.1})$$

$$\text{Var}[A_l^{(i,j)}] = \lambda^2 \frac{1}{l^2} \frac{\sigma_i^2}{\sigma_j^2}, \quad l = 1, \dots, p. \quad (\text{A.2})$$

The scale factor $\frac{1}{l^2}$ helps to impose the prior belief that recent lags play a more influential role than distant ones by shrinking the variances on the more distant lags proportionally (centered on the prior mean of zero). The prior parameter σ_i equals the standard deviation of the residuals obtained from regressing the variable y_i on its own p lags and a constant over the sample period up to time t . We set δ_i equal to the sum of the autoregressive coefficients obtained from regressing the variable y_i on its own p lags and a constant over a pre-forecast evaluation sample.² The hyperparameter λ governs the tightness of our prior beliefs. As $\lambda \rightarrow 0$, the prior dominates, and so posterior equals prior; i.e., the data have no say. On the other hand,

¹Natural conjugate priors such as N-IW have computational advantages and at the same time competitive forecasting properties (see Koop, 2013).

²Since all of the variables that enter the VAR are stationary, we get qualitatively similar results if we instead set $\delta_i = 0$ for all i . Studies such as those of Clark (2011) and Carriero, Clark, and Marcellino (2015a) have used a value of 0.8 for variables that are known to exhibit persistence (e.g., the unemployment rate, inflation and the interest rate). Our results remain very similar if we instead set $\delta_i = 0.8$.

as $\lambda \rightarrow \infty$, the prior's influence diminishes, and so the posterior estimates converge to OLS estimates.

The VAR model in Eq. (2.1) can be rewritten in compact matrix notation as

$$Y = XA + E, \quad (\text{A.3})$$

where $Y = [y_1, \dots, y_T]'$ is a $T \times n$ matrix, $X = [x_1, \dots, x_T]'$ is a $T \times k$ matrix and $x_t = [1, y'_{t-1}, \dots, y'_{t-p}]$ is a $1 \times k$ vector. $A = [A_c \ A_1 \ \dots \ A_p]$ is a matrix of VAR coefficients of size $k \times n$, and $E = [\varepsilon_1 \ \dots \ \varepsilon_T]$ is a $T \times n$ matrix of innovation terms.

The conjugate normal inverse-Wishart (N-IW) prior is:

$$\text{vec}(A) | \Sigma \sim N(\text{vec}(A_0), \Sigma \otimes \Omega_0), \quad \Sigma \sim iW(S_0, v_0), \quad (\text{A.4})$$

where A_0 , Ω_0 , S_0 , and v_0 are the prior parameters of which the values are set based on the VAR model in Eq. (2.1) that satisfies the prior moment conditions specified in equations (A.1 and A.2).

Since our prior N-IW is conjugate, the resulting conditional posterior distribution (i.e., the product of the prior and the likelihood function) is also N-IW.

$$\text{vec}(A) | \Sigma, Y \sim N(\text{vec}(\bar{A}), \Sigma \otimes \bar{\Omega}), \quad \Sigma \sim iW(\bar{S}, \bar{v}), \quad (\text{A.5})$$

where

$$\bar{A} = (\Omega_0^{-1} + X'X)^{-1} (\Omega_0^{-1}A_0 + X'Y), \quad (\text{A.6})$$

$$\bar{\Omega} = (\Omega_0^{-1} + X'X)^{-1} \quad \text{and} \quad (\text{A.7})$$

$$\bar{v} = v_0 + T \quad (\text{A.8})$$

are the respective posterior mean estimates of the VAR model, and

$$\bar{S} = A_0 + \hat{\varepsilon}'\hat{\varepsilon} + \hat{A}'X'X\hat{A} + A_0'\Omega_0^{-1}A_0 - \hat{A}'\bar{\Omega}^{-1}\hat{A}, \quad (\text{A.9})$$

where $\hat{A} = (X'X)^{-1}X'Y$ is the OLS estimate of A and $\hat{\varepsilon} = Y - X\hat{A}$ are the OLS residuals (Zellner, 1971). We use the mixed estimation method of Litterman (1986) to implement the N-IW prior, which equates to appending the data matrices with dummy observations.

Previous research (e.g., Robertson & Tallman, 1999; Banbura et al., 2010) has documented further gains in forecast accuracy by imposing a 'sum of coefficients' (SOC) prior on the equations of the VAR. The hyperparameter μ governs the tightness of this prior.

A0.1 Optimal values of the hyperparameters using the marginal likelihood

The values of λ and μ are set by maximizing the marginal likelihood of the model over a predefined two-dimensional discrete grid of λ and μ . The optimization is performed over the

pre-forecast evaluation sample, and the optimal values obtained are kept fixed over the forecast evaluation sample.³

$$[\lambda^+, \mu^+] = \arg \max_{[\lambda^g, \mu^g]} \ln p(Y), \quad (\text{A.10})$$

where

$$p(Y) = \int p(Y|\theta)p(\theta)d(\theta) \quad (\text{A.11})$$

is the marginal likelihood, and θ is the set of all model coefficients. Given that we are using an N-IW prior, the marginal density $p(Y)$ can be solved in closed form as⁴

$$p(Y) = \left(\frac{1}{\pi}\right)^{\frac{nT}{2}} \times |(I + X\Omega_0X')^{-1}|^{\frac{n}{2}} \times |S_0|^{\frac{v_0}{2}} \\ \times \frac{\Gamma_n(\frac{v_0+T}{2})}{\Gamma_n(\frac{v_0}{2})} \times |S_0 + (Y - XA_0)'(I + X\Omega_0X')^{-1}(Y - XA_0)|^{-\frac{v_0+T}{2}}.$$

We evaluate the log marginal likelihood using the two-dimensional grid of discrete values that is defined as: $\lambda^g = [0.050, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50]$; and $\mu^g = [0.1, 0.15, 0.2, 0.25, 0.50, 1, 1.5, 2, 2.5, 3]$. For the small BVAR estimated using data starting in 1959, the optimal values obtained are $\lambda = 0.4$ and $\mu = 0.2$, while the optimal values for the medium VAR are $\lambda = 0.3$ and $\mu = 0.25$. The values that we obtain are close to those that other researchers have obtained through a grid search optimization.

Note that the prior specification for each equation is symmetric in its treatment of own lags of the dependent variable and lags of other variables. As such, we have a prior that is a natural conjugate (normal inverse-Wishart prior), which proves to be convenient when solving for the model, because these priors can be implemented easily by augmenting the data matrices with dummy variables, thus permitting OLS to estimate the model equation by equation (for details see Banbura et al., 2010, and Carriero et al., 2015a).

A0.2 Illustrating the influence of estimated steady state on the forecasts

This section illustrates two important concepts: the influence of a model's estimated mean (i.e., steady state) on the forecasts, and the direct role of persistence in determining the forecast horizon at which the convergence to the steady state takes place. To keep things simple, let us consider a covariance-stationary AR(1) model,

$$y_t = c + \rho y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2). \quad (\text{A.12})$$

³However, that being said, the results remain qualitatively similar if we instead fix $\lambda = 0.2$ and $\mu = 1$ (widely used values).

⁴Details of the derivation are provided by Bauwens, Lubrano, and Richard (1999), Carriero et al. (2015a) and Giannone et al. (2015).

Covariance-stationarity implies $|\rho| < 1$, and

$$\mathbb{E}(y_t) = \mathbb{E}(y_{t-1}) = \mu, \quad (\text{A.13})$$

where μ denotes the steady state of the model. In long samples, it matches the mean of estimation sample.

Next, taking the expectations of both sides of Eq. (A.12) gives us

$$\mathbb{E}(y_t) = c + \rho\mathbb{E}(y_{t-1}) + \mathbb{E}(\varepsilon_t). \quad (\text{A.14})$$

Plugging Eq. (A.13) into Eq. (A.14) yields

$$\mu = c + \rho\mu = \frac{c}{1 - \rho}. \quad (\text{A.15})$$

Rearranging the above equation gives

$$c = \mu(1 - \rho). \quad (\text{A.16})$$

Rewrite Eq. (A.12) by substituting Eq. (A.16) into Eq. (A.12):

$$\begin{aligned} y_t &= \mu(1 - \rho) + \rho y_{t-1} + \varepsilon_t \\ y_t - \mu &= \rho(y_{t-1} - \mu) + \varepsilon_t. \end{aligned} \quad (\text{A.17})$$

Here, ρ is the persistence of y , and determines the speed of convergence back to the steady state. $1 - \rho$ is the pace of adjustment towards steady state in each period. $\frac{1}{1 - \rho}$ is the number of periods that it takes for y to converge back to its steady state value. Higher values of ρ imply longer-lived departures of y from the steady state.

Denoting by $\hat{\rho}$ and $\hat{\mu}$ estimates obtained via estimation of available data through time t , the one-step-ahead forecast is

$$\hat{y}_{t|t+1} - \hat{\mu} = \hat{\rho}(\hat{y}_t - \hat{\mu}). \quad (\text{A.18})$$

Similarly, h -step-ahead forecast is

$$\begin{aligned} \hat{y}_{t|t+h} - \hat{\mu} &= \hat{\rho}^h(\hat{y}_t - \hat{\mu}) \\ \hat{y}_{t|t+h} &= \hat{\mu} + \hat{\rho}^h(\hat{y}_t - \hat{\mu}). \end{aligned} \quad (\text{A.19})$$

A closer inspection of the above equation reveals that the h -step-ahead forecast is the sum of two components: the estimated mean, and the gap between the estimated mean and y (as of the forecast origin t) multiplied by the h^{th} power of $\hat{\rho}$.

Note that $\hat{\rho}^h$ is a decreasing function of h ,

$$\hat{\rho} > \hat{\rho}^2 > \hat{\rho}^3 > \dots > \hat{\rho}^\infty,$$

which suggests that the influence of the component $\hat{\mu}$ becomes more important as h increases.

In the limit, as $h \rightarrow \infty$, $\hat{\rho}^h \rightarrow 0$, and $\lim_{h \rightarrow \infty} \hat{y}_{t|t+h} \rightarrow \hat{\mu}$.

The smaller the value of $\hat{\rho}$, the faster $\hat{\rho}^h$ goes to zero, and the more rapid the convergence to the model's implied steady state.

Please refer to Clements and Hendry (1999) for further details.

A1. Results for Medium VARs

The Medium VAR builds on Small VAR by five additional variables that include real personal consumption expenditures, nonfarm business productivity, the employment cost index-wage and salary of private workers (ECI), nonfarm payroll employment, and the core CPI (i.e., the CPI excluding food and energy). All these variables are transformed to quarterly annualized growth rates. To compute the growth rates, we use 400 times the log difference formula.

Tables A1 to A4 below report both point and density forecast accuracy comparisons for all 10 variables. The results indicate that the patterns of both point and density forecast accuracy improvements for hybrid forecasts are generally similar to that of small VAR. Specifically, results for medium VAR estimated with longer sample echo strongly the results reported for small VAR with longer sample. One difference is that the magnitude of improvements for hybrid forecasts is slightly smaller than those reported for small VAR. The additional variables in medium VAR are helping improve the accuracy of the core variables of interest (relative to small VAR) therefore with more accurate raw forecasts tilting is helping slightly less (see section A4 in this appendix). The finding that medium VAR generates more accurate forecasts compared to small VAR are in line with Banbura et al. (2010) and Koop (2013). Results for medium VAR estimated with shorter sample echo the results reported for small VAR with shorter sample. Unlike in the case of estimation with longer sample favoring medium-sized VAR over small-sized, we do not find this pattern for VAR models estimated with shorter sample.

The most useful aspect of the results for medium VAR is the strong positive spillover effects on the accuracy of the variables that are not directly tilted. Impressive and statistical significant gains in the accuracy of forecasts derived from hybrid approach are achieved for core CPI inflation, wage compensation, nonfarm payroll employment, and credit spread. It is also worth highlighting that adding core CPI inflation and SV in the medium VAR greatly helps improve the accuracy of the raw CPI forecasts. This implies tilting is less effective for CPI inflation for those specifications, but for core CPI inflation it is very effective in all medium VAR specifications.

Table A1: Out-of-Sample **Point** Forecasting Performance: **Medium BVAR est. 1960**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Medium VAR				Medium VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	3.88	6.89	6.58	6.18	3.69	6.55	6.80	6.87
Relative MSE								
Baseline/Raw	0.69***	0.98*	0.99	1.03	0.73**	0.98	1.01	1.01
Hybrid/Raw	0.69***	0.96	0.84	0.99	0.73**	0.91*	0.82**	0.89
Hybrid/Baseline	1.00	0.93*	0.87	0.97	1.00	0.92*	0.81**	0.89
CPI								
Raw	2.99	5.17	5.64	6.68	2.62	4.73	4.81	4.94
Relative MSE								
Baseline/Raw	0.35***	0.95	0.95	0.96*	0.39***	0.93**	0.95**	0.98
Hybrid/Raw	0.35***	0.89**	0.81***	0.73***	0.39***	0.94	0.95	1.00
Hybrid/Baseline	1.00	0.95***	0.87**	0.78***	1.00	1.00	1.00	1.02
UR								
Raw	0.05	0.61	2.43	4.10	0.05	0.58	2.34	3.99
Relative MSE								
Baseline/Raw	0.33***	0.81*	0.91**	0.96	0.32***	0.75**	0.92*	0.96
Hybrid/Raw	0.33***	0.82***	0.91	0.89*	0.32***	0.75**	0.88	0.89
Hybrid/Baseline	1.00	0.96	0.97	0.93	1.00	1.00	0.95	0.92
FFR								
Raw	0.20	1.84	4.83	8.33	0.09	1.65	5.64	9.43
Relative MSE								
Baseline/Raw	0.03***	0.75**	0.96	0.99	0.02***	0.73***	0.89***	0.96**
Hybrid/Raw	0.03***	0.85**	0.93	0.78***	0.05***	0.80***	0.85	0.75*
Hybrid/Baseline	1.00	1.14**	0.99	0.81*	1.00	1.09***	0.95	0.78*
Credit Spread								
Raw	0.09	0.63	1.06	1.21	0.08	0.66	1.19	1.40
Relative MSE								
Baseline/Raw	0.83	0.93**	0.99	1.01	0.90*	0.98*	0.99	0.97**
Hybrid/Raw	0.79	0.96*	0.89***	0.87***	0.90*	0.97*	0.86***	0.78**
Hybrid/Baseline	0.99	1.02	0.91***	0.87***	1.00	0.99	0.87**	0.80**

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; Credit Spread: in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts (both mean and variance) only. Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean squared error (MSE), the three rows immediately below report relative MSE: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table A1: Cont.: Out-of-Sample **Point** Forecasting Performance: **Medium BVAR est. 1960**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Medium VAR				Medium VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
Consumption								
Raw	2.57	4.26	4.35	4.14	2.37	4.08	4.54	4.61
Relative MSE								
Baseline/Raw	0.79***	0.94	1.04	1.06**	0.97	0.98	1.03	1.01
Hybrid/Raw	0.81**	0.94	1.03	1.07	0.99	0.95	0.93	0.98
Hybrid/Baseline	1.03	1.00	0.99	1.01	1.01	0.97	0.90	0.97
Core CPI								
Raw	0.33	0.77	1.50	2.88	0.31	0.50	0.84	1.38
Relative MSE								
Baseline/Raw	0.73**	0.73***	0.84***	0.91**	0.82**	0.80**	0.98	0.89**
Hybrid/Raw	0.68**	0.67***	0.53***	0.34***	0.81**	0.77	0.86	0.64*
Hybrid/Baseline	0.93***	0.91	0.64***	0.37***	0.99	0.96	0.88	0.72
Productivity								
Raw	4.99	5.57	5.22	5.99	4.89	5.34	5.02	5.96
Relative MSE								
Baseline/Raw	0.95	0.88**	0.97	1.02***	1.00	0.97	1.01	1.02
Hybrid/Raw	1.01	0.89*	0.97	1.07***	1.00	0.95	1.01	0.98
Hybrid/Baseline	1.06**	1.01	1.00	1.05***	0.99	0.98	1.00	0.96
Compensation								
Raw	0.77	1.22	2.09	2.91	0.75	1.06	1.33	1.43
Relative MSE								
Baseline/Raw	1.10	0.93**	0.91***	0.95**	1.03	0.95**	0.93***	0.96
Hybrid/Raw	0.95	0.83***	0.57***	0.40***	1.02	0.92**	0.84***	0.75**
Hybrid/Baseline	0.86*	0.89*	0.63***	0.42***	0.99	0.96	0.90*	0.78***
Payroll Employment								
Raw	0.55	3.49	4.76	4.68	0.52	3.04	4.30	4.33
Relative MSE								
Baseline/Raw	0.64***	0.91*	0.98*	1.07	0.63***	0.88**	0.97***	1.01
Hybrid/Raw	0.62***	0.81**	0.70**	0.82**	0.61***	0.77**	0.69***	0.73***
Hybrid/Baseline	0.98	0.89**	0.72**	0.76***	0.97**	0.88*	0.71***	0.72***

Notes for Table: For all the variables in this panel, the forecasts and the associated accuracy correspond to the quarterly annualized rate. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts (both mean and variance) only. Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean squared error (MSE), the three rows immediately below report relative MSE: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table A2: Out-of-Sample **Density** Forecasting Performance: **Medium BVAR est. 1960**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Medium VAR				Medium VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	1.15	1.46	1.42	1.40	1.08	1.40	1.44	1.44
Relative CRPS								
Baseline - Raw	-0.22***	-0.01*	0.01	0.01	-0.16***	-0.03	0.00	0.01
Hybrid - Raw	-0.22***	-0.04	-0.07	0.00	-0.16***	-0.07*	-0.11*	-0.04
Hybrid - Baseline	0.00	-0.03	-0.08*	-0.01	0.00	-0.05	-0.10*	-0.05
CPI								
Raw	0.96	1.17	1.28	1.38	0.90	1.11	1.18	1.22
Relative CRPS								
Baseline - Raw	-0.38***	-0.03	-0.02	-0.02	-0.32***	-0.05**	-0.03**	-0.02**
Hybrid - Raw	-0.38***	-0.06**	-0.10**	-0.16***	-0.32***	-0.04	-0.04	-0.02
Hybrid - Baseline	0.00	-0.03**	-0.08**	-0.14**	0.00	0.01	-0.02	0.00
UR								
Raw	0.12	0.39	0.81	1.11	0.12	0.37	0.78	1.08
Relative CRPS								
Baseline - Raw	-0.05***	-0.04*	-0.05**	-0.03	-0.04***	-0.05**	-0.03	-0.01
Hybrid - Raw	-0.05***	-0.05**	-0.06	-0.06	-0.05***	-0.05***	-0.05	-0.04
Hybrid - Baseline	0.00	-0.01**	-0.01	-0.03	0.00	-0.01	-0.02	-0.03
FFR								
Raw	0.27	0.78	1.26	1.68	0.17	0.71	1.35	1.82
Relative CRPS								
Baseline - Raw	-0.23***	-0.10***	-0.02	0.00	-0.14***	-0.11***	-0.09***	-0.05**
Hybrid - Raw	-0.23***	-0.04**	0.00	-0.14	-0.14***	-0.08***	-0.11*	-0.25*
Hybrid - Baseline	0.00	0.06***	0.01	-0.14*	0.00	0.03**	-0.02	-0.20
Credit Spread								
Raw	0.15	0.42	0.56	0.63	0.14	0.43	0.61	0.69
Relative CRPS								
Baseline - Raw	-0.01	-0.01**	0.00	0.01*	-0.01	-0.01*	-0.01*	0.00
Hybrid - Raw	-0.01	-0.01*	-0.04***	-0.05***	0.00	-0.01*	-0.06***	-0.08***
Hybrid - Baseline	0.00	0.00	-0.04***	-0.06***	0.00	0.00	-0.05***	-0.08***

Notes for Table: Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The numbers reported in the row labeled Raw are the mean cumulative ranked probability score (CRPS), all other numbers are relative CRPS; relative to Raw or relative to Baseline. A negative value for Relative CRPS suggests that on average the Baseline forecast is more accurate compared to the Raw forecast (in the case of Baseline - Raw); Hybrid more accurate compared to Raw (in the case of Hybrid - Raw); Hybrid more accurate compared to Baseline (in the case of Hybrid - Baseline). The left panel reports results for time-invariant Medium VAR and right panel reports results corresponding to Medium VAR with stochastic volatility. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table A2: Cont. Out-of-Sample **Density** Forecasting Performance: **Medium BVAR est. 1960**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Medium VAR				Medium VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
Consumption								
Raw	0.93	1.15	1.16	1.16	0.87	1.10	1.18	1.20
Relative CRPS								
Baseline - Raw	-0.10***	-0.01	0.02	0.04**	-0.02	-0.02	0.00	-0.01
Hybrid - Raw	-0.09***	-0.01	0.04	0.04**	-0.01	-0.04	-0.02	0.00
Hybrid - Baseline	0.01	0.00	0.02	0.00	0.01	-0.02	-0.02	0.01
Core CPI								
Raw	0.37	0.56	0.77	1.02	0.31	0.41	0.54	0.70
Relative CRPS								
Baseline - Raw	-0.05***	-0.05***	-0.05***	-0.04**	-0.02**	-0.04***	-0.01	-0.03***
Hybrid - Raw	-0.06***	-0.05**	-0.12**	-0.25***	-0.02**	-0.04*	-0.03	-0.08
Hybrid - Baseline	-0.01	0.00	-0.08***	-0.22***	0.00	0.00	-0.01	-0.05
Productivity								
Raw	1.28	1.35	1.33	1.40	1.26	1.31	1.30	1.39
Relative CRPS								
Baseline - Raw	-0.04	-0.06**	0.00	0.03***	0.01	-0.01	0.00	0.02
Hybrid - Raw	-0.05***	-0.05**	-0.06	-0.06	0.02	-0.01	0.00	0.00
Hybrid - Baseline	0.03**	-0.01	0.00	0.03**	0.01	0.00	0.00	-0.02
Compensation								
Raw	0.53	0.65	0.85	1.02	0.48	0.58	0.67	0.72
Relative CRPS								
Baseline - Raw	0.01	-0.03***	-0.04***	-0.01	0.00	-0.02***	-0.02***	-0.02
Hybrid - Raw	-0.01	-0.05***	-0.14***	-0.25***	0.00	-0.03***	-0.04***	-0.05*
Hybrid - Baseline	-0.02*	-0.02*	-0.10***	-0.23***	0.00	-0.01	-0.02*	-0.04*
Payroll Employment								
Raw	0.42	0.98	1.16	1.15	0.40	0.90	1.10	1.13
Relative CRPS								
Baseline - Raw	-0.08***	-0.04	0.00	0.05	-0.07***	-0.07***	-0.01**	0.01
Hybrid - Raw	-0.08***	-0.09*	-0.19**	-0.11*	-0.08	-0.13***	-0.19**	-0.15**
Hybrid - Baseline	0.00	-0.05**	-0.19**	-0.16**	-0.01*	-0.06*	-0.17**	-0.16**

Notes for Table: Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The numbers reported in the row labeled Raw are the mean cumulative ranked probability score (CRPS), all other numbers are relative CRPS; relative to Raw or relative to Baseline. A negative value for Relative CRPS suggests that on average the Baseline forecast is more accurate compared to the Raw forecast (in the case of Baseline - Raw); Hybrid more accurate compared to Raw (in the case of Hybrid - Raw); Hybrid more accurate compared to Baseline (in the case of Hybrid - Baseline). The left panel reports results for time-invariant Medium VAR and right panel reports results corresponding to Medium VAR with stochastic volatility. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table A3: Out-of-Sample **Point** Forecasting Performance: **Medium BVAR est. 1985**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Medium VAR				Medium VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	3.72	7.10	7.75	7.36	3.99	6.72	8.00	8.04
Relative MSE								
Baseline/Raw	0.72**	0.93***	0.96	1.03***	0.68***	0.97	1.01	1.02
Hybrid/Raw	0.72**	0.86*	0.77***	0.91**	0.68***	0.88	0.75**	0.78**
Hybrid/Baseline	1.00	0.92	0.80**	0.89***	1.00	0.92	0.74*	0.77**
CPI								
Raw	3.65	5.08	5.39	6.51	2.70	4.83	4.88	5.44
Relative MSE								
Baseline/Raw	0.28***	0.99	1.001	1.00	0.38***	0.98	0.97**	0.99
Hybrid/Raw	0.28***	0.92	0.84**	0.83***	0.38***	0.95	0.92*	0.90**
Hybrid/Baseline	1.00	0.93*	0.84***	0.83***	1.00	0.97	0.95	0.91**
UR								
Raw	0.05	0.53	2.21	4.11	0.04	0.49	2.16	4.15
Relative MSE								
Baseline/Raw	0.31***	0.72*	0.85	0.93	0.39***	0.80***	0.90*	0.95
Hybrid/Raw	0.31***	0.73*	0.88	0.91	0.39***	0.83***	0.93	0.89
Hybrid/Baseline	1.00	1.02	1.04	0.98	1.00	1.04	1.03	0.94
FFR								
Raw	0.10	2.03	7.39	12.95	0.08	1.94	6.81	12.83
Relative MSE								
Baseline/Raw	0.05***	0.66***	0.86***	0.96*	0.06***	0.66***	0.85***	0.93**
Hybrid/Raw	0.05***	0.60***	0.65***	0.66**	0.06***	0.64***	0.65**	0.50
Hybrid/Baseline	1.00	0.90	0.75**	0.69**	1.00	0.97	0.76	0.54
Credit Spread								
Raw	0.11	0.66	1.09	1.21	0.09	0.69	1.16	1.25
Relative MSE								
Baseline/Raw	0.68**	0.92*	0.98**	1.04	0.87*	1.01	1.03	1.04
Hybrid/Raw	0.68**	0.89**	0.91	0.94	0.84*	1.00	0.86	0.82
Hybrid/Baseline	1.01	0.96	0.93	0.90*	0.97**	0.99	0.84	0.79

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; Credit Spread: in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts (both mean and variance) only. Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean squared error (MSE), the three rows immediately below report relative MSE: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table A3: Cont.: Out-of-Sample **Point** Forecasting Performance: **Medium BVAR est. 1985**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Medium VAR				Medium VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
Consumption								
Raw	2.55	4.79	5.91	6.45	2.70	4.64	6.27	6.71
Relative MSE								
Baseline/Raw	0.94	0.95	0.97	1.01	0.97	0.98	0.98	1.06
Hybrid/Raw	0.93	0.96	0.86***	0.81**	0.98	0.93	0.79	0.82***
Hybrid/Baseline	0.99	1.01	0.88	0.80	1.01	0.94	0.80	0.77***
Core CPI								
Raw	0.36	0.59	0.95	1.54	0.35	0.55	0.74	1.12
Relative MSE								
Baseline/Raw	0.87**	0.96	0.91**	0.95***	0.82**	0.91**	0.92**	0.92**
Hybrid/Raw	0.88**	0.82*	0.59**	0.54***	0.81**	0.77**	0.74	0.60**
Hybrid/Baseline	1.01	0.86	0.65**	0.57**	0.99	0.85*	0.81	0.65*
Productivity								
Raw	4.81	5.89	5.06	6.45	4.78	5.58	5.41	6.55
Relative MSE								
Baseline/Raw	0.98	0.90**	1.01	1.02**	0.98	0.98**	1.00	1.02***
Hybrid/Raw	0.95	0.88***	1.03	0.95	0.96	0.98	1.02	0.97
Hybrid/Baseline	0.97*	0.97	1.02	0.93**	0.98*	1.01	1.02	0.95
Compensation								
Raw	0.83	1.00	1.19	1.36	0.77	0.92	1.05	1.23
Relative MSE								
Baseline/Raw	1.03	1.05	0.96	1.00	1.06*	1.04	1.00	0.99
Hybrid/Raw	1.05	1.04	0.96	0.96	1.04	1.07	1.01	0.95
Hybrid/Baseline	1.02	0.99	1.00	0.96	0.98	1.03	1.00	0.96
Payroll Employment								
Raw	0.56	2.76	4.14	3.97	0.49	2.88	4.64	4.57
Relative MSE								
Baseline/Raw	0.67***	0.83**	0.98	1.08	0.70***	0.85**	0.99	1.02
Hybrid/Raw	0.64***	0.72**	0.79**	0.90	0.67***	0.80*	0.71	0.70**
Hybrid/Baseline	0.95*	0.87	0.81*	0.83***	0.96**	0.94	0.72	0.69*

Notes for Table: for all the variables in this panel, the forecasts and the associated accuracy correspond to the quarterly annualized rate. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts (both mean and variance) only. Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean squared error (MSE), the three rows immediately below report relative MSE: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table A4: Out-of-Sample **Density** Forecasting Performance: **Medium BVAR est. 1985**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Medium VAR				Medium VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	1.08	1.47	1.53	1.49	1.13	1.44	1.57	1.61
Relative CRPS								
Baseline - Raw	-0.16**	-0.06***	-0.02	0.03	-0.20***	-0.02	0.00	-0.01
Hybrid - Raw	-0.16**	-0.11	-0.19**	-0.07**	-0.20***	-0.07	-0.15**	-0.13***
Hybrid - Baseline	0.00	-0.04	-0.17**	-0.10***	0.00	-0.05	-0.15*	-0.12***
CPI								
Raw	1.02	1.23	1.26	1.40	0.93	1.16	1.21	1.31
Relative CRPS								
Baseline - Raw	-0.43***	-0.03	0.01	0.02	-0.35***	-0.03**	-0.02***	-0.02*
Hybrid - Raw	-0.43***	-0.11*	-0.15**	-0.17***	-0.35***	-0.05*	-0.05	-0.07**
Hybrid - Baseline	0.00	-0.08**	-0.17***	-0.19***	0.00	-0.02	-0.03	-0.05**
UR								
Raw	0.12	0.36	0.77	1.15	0.11	0.35	0.75	1.11
Relative CRPS								
Baseline - Raw	-0.05***	-0.05**	-0.05	-0.03*	-0.04***	-0.04***	-0.04**	-0.03*
Hybrid - Raw	-0.05***	-0.05	-0.02	-0.03	-0.04***	-0.04***	-0.02	-0.03
Hybrid - Baseline	0.00	0.01	0.03	0.00	0.00	0.01	0.02	0.00
FFR								
Raw	0.18	0.84	1.63	2.15	0.16	0.80	1.53	2.05
Relative CRPS								
Baseline - Raw	-0.14***	-0.17***	-0.13***	-0.06***	-0.13***	-0.15***	-0.13***	-0.08***
Hybrid - Raw	-0.14***	-0.20***	-0.35***	-0.46***	-0.13***	-0.15***	-0.25**	-0.43
Hybrid - Baseline	0.00	-0.04	-0.22**	-0.40**	0.00	0.00	-0.13	-0.35
Credit Spread								
Raw	0.17	0.44	0.59	0.66	0.16	0.44	0.59	0.63
Relative CRPS								
Baseline - Raw	-0.02***	-0.01	-0.01*	0.02	-0.01*	0.00	0.00	0.00
Hybrid - Raw	-0.02**	-0.03*	-0.04*	-0.03	-0.01*	-0.01	-0.04*	-0.05*
Hybrid - Baseline	0.00	-0.02*	-0.03*	-0.05**	0.00	0.00	-0.04	-0.05

Notes for Table: Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts and long-horizon forecasts. The numbers reported in the row labeled Raw are the mean cumulative ranked probability score (CRPS), all other numbers are relative CRPS; relative to Raw or relative to Baseline. A negative value for Relative CRPS suggests that on average the Baseline forecast is more accurate compared to the Raw forecast (in the case of Baseline - Raw); Hybrid more accurate compared to Raw (in the case of Hybrid - Raw); Hybrid more accurate compared to Baseline (in the case of Hybrid - Baseline). The left panel reports results for time-invariant Medium VAR and right panel reports results corresponding to Medium VAR with stochastic volatility. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

A2. Density Forecast Calibration Diagnostics

As indicated in the main body of the chapter, the density calibration is assessed using:

- a. Interval forecasts (i.e. 70% prediction intervals)
- b. Formal assessment of probability integral transforms (PITS) using
 - i). Berkowitz (2001) test: tests for the uniformity of the PITS by testing whether the inverse normal transformation of the PITS is distributed $iidN(0, 1)$. Specifically, for the case of one-step ahead density forecast, the test involves testing whether the transformed PIT distribution is $iidN(0, 1)$. The test statistic (likelihood ratio test) is asymptotically $\chi^2(3)$ distributed. For horizons beyond one-step, we test whether transformed PIT distribution is identically distributed $N(0, 1)$ and the test statistic is asymptotically $\chi^2(2)$ distributed.
 - ii). Kolomogorov-Smirnov test: tests for the closeness between the CDF of a $U(0, 1)$ and the empirical distribution of PITS. The critical values for the tests are obtained using the procedure discussed in Rossi and Sekhposyan (2014).
 - iii). Knüppel (2015) test: is considered a more powerful test because it is a raw moments (i.e. nonstandardized, noncentral moments) based test and is robust to the presence of serial correlation in the PITS corresponding to the multi-step ahead density forecasts.

Generally, all three tests point to similar inference. Therefore, in the interest of brevity, we report inference based on Knüppel (2015) test only.

Tables A5 through A8 report the density calibration for all the VAR models considered in the chapter.

The entries reported in the rows labeled COV are the 70% empirical coverage rates. And the associated sharpness of the prediction intervals (defined as the average difference between the 15th and 85th percentiles of the predictive densities) are reported in the rows labeled LEN. Smaller values of LEN indicate more concentrated densities (higher sharpness). Sharper forecasts are preferred conditional on correctly calibrated densities.

Table A5: **Calibration** Assessment of Density Forecasts: **Small VAR est. 1960**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)											
Small VAR			Small VAR with SV				Small VAR with SV				
			h=1Q	h=4Q	h=8Q	h=12Q		h=1Q	h=4Q	h=8Q	h=12Q
GDP	Raw	PIT	0.00	0.00	0.00	0.00	PIT	0.07	0.00	0.01	0.02
		COV	86.4	85.9	92.6	90.9	COV	75.0	70.6	74.1	79.2
		LEN	6.09	6.87	7.21	7.29	LEN	4.28	5.06	5.68	6.08
	Baseline	PIT	0.49	0.00	0.00	0.00	PIT	0.37	0.00	0.01	0.03
		COV	71.6	84.7	88.9	89.6	COV	72.7	68.2	76.5	80.5
		LEN	3.54	6.71	6.97	7.07	LEN	3.37	4.90	5.46	5.92
	Hybrid	PIT	0.49	0.00	0.00	0.00	PIT	0.38	0.07	0.12	0.02
		COV	72.7	88.2	90.1	87.0	COV	70.5	71.8	76.5	81.8
		LEN	3.57	6.65	7.06	7.03	LEN	3.38	5.08	5.82	6.40
CPI Inflation	Raw	PIT	0.92	0.00	0.00	0.00	PIT	0.61	0.03	0.00	0.00
		COV	67.1	80.0	80.3	81.8	COV	65.9	78.8	76.5	83.1
		LEN	3.16	4.80	5.99	6.48	LEN	3.37	4.94	6.26	7.07
	Baseline	PIT	0.60	0.01	0.00	0.00	PIT	0.88	0.01	0.00	0.00
		COV	69.3	82.4	82.7	79.2	COV	70.5	77.7	79.0	83.1
		LEN	2.24	4.54	5.82	6.44	LEN	2.18	4.52	5.91	6.73
	Hybrid	PIT	0.74	0.00	0.00	0.00	PIT	0.90	0.00	0.00	0.00
		COV	70.5	83.5	90.1	87.0	COV	70.5	80.0	81.5	92.2
		LEN	2.19	4.52	6.02	6.56	LEN	2.17	4.69	6.22	7.22
UR	Raw	PIT	0.57	0.00	0.00	0.00	PIT	0.95	0.01	0.00	0.00
		COV	79.6	84.7	74.1	62.3	COV	71.6	77.7	70.4	62.3
		LEN	0.47	1.47	2.14	2.55	LEN	0.40	1.19	1.89	2.49
	Baseline	PIT	0.00	0.00	0.00	0.00	PIT	0.01	0.01	0.00	0.00
		COV	70.5	80.0	75.3	63.6	COV	72.7	77.7	72.8	62.3
		LEN	0.27	1.29	2.06	2.47	LEN	0.26	1.06	1.78	2.34
	Hybrid	PIT	0.01	0.00	0.00	0.00	PIT	0.01	0.01	0.00	0.00
		COV	70.5	83.5	75.3	58.4	COV	72.7	75.3	65.4	50.7
		LEN	0.27	1.32	2.09	2.47	LEN	0.26	1.09	1.87	2.51
FFR	Raw	PIT	0.00	0.00	0.00	0.00	PIT	0.38	0.00	0.00	0.00
		COV	93.2	85.9	72.8	57.1	COV	76.1	71.8	63.0	49.4
		LEN	1.70	4.09	5.77	6.55	LEN	0.73	2.62	4.44	5.38
	Baseline	PIT	0.01	0.00	0.00	0.00	PIT	0.01	0.00	0.00	0.00
		COV	75.0	88.2	72.8	58.4	COV	76.1	75.3	61.7	50.7
		LEN	0.15	3.51	5.55	6.39	LEN	0.13	2.14	4.19	5.21
	Hybrid	PIT	0.00	0.00	0.00	0.00	PIT	0.01	0.00	0.00	0.00
		COV	79.6	87.1	76.5	72.7	COV	79.6	74.1	72.8	70.1
		LEN	0.16	3.49	5.60	6.62	LEN	0.13	2.23	4.41	5.61

Notes for Table: Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. PIT are the p-values from the statistical test of Knüppel (2015). The values less than 0.10 suggest instances where the hypothesis of correctly calibrated density forecasts is rejected at the 10% significance level. COV corresponds to the 70% empirical coverage rate; values closer to nominal value of 70% is preferred. LEN corresponds to the width of the 70% prediction intervals.

Table A6: **Calibration** Assessment of Density Forecasts: **Small VAR est. 1985**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)											
Small VAR							Small VAR with SV				
			h=1Q	h=4Q	h=8Q	h=12Q					
			h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q	
GDP	Raw	PIT	0.83	0.27	0.22	0.27	PIT	0.31	0.11	0.17	0.17
		COV	69.3	64.7	69.1	71.4	COV	73.9	71.8	71.6	76.6
		LEN	3.96	4.74	4.82	4.81	LEN	4.15	5.06	5.41	5.67
	Baseline	PIT	0.49	0.43	0.21	0.16	PIT	0.38	0.20	0.18	0.24
		COV	71.6	69.4	69.1	70.1	COV	70.5	72.9	74.1	76.6
		LEN	3.49	4.62	4.82	4.70	LEN	3.38	4.87	5.27	5.54
	Hybrid	PIT	0.44	0.50	0.54	0.21	PIT	0.36	0.50	0.19	0.13
		COV	71.6	68.2	75.3	74.0	COV	70.5	74.1	76.5	77.9
		LEN	3.51	4.67	4.86	4.76	LEN	3.38	5.01	5.51	5.83
CPI Inflation	Raw	PIT	0.08	0.10	0.01	0.00	PIT	0.84	0.21	0.22	0.38
		COV	59.1	60.0	54.3	54.6	COV	60.2	69.4	64.2	66.2
		LEN	2.80	3.11	3.19	3.23	LEN	3.26	3.81	4.07	4.40
	Baseline	PIT	0.76	0.10	0.01	0.00	PIT	0.81	0.12	0.23	0.42
		COV	70.5	58.8	51.9	49.4	COV	68.2	64.7	64.2	64.9
		LEN	2.20	3.02	3.05	3.21	LEN	2.16	3.61	3.98	4.27
	Hybrid	PIT	0.96	0.21	0.28	0.13	PIT	0.79	0.21	0.46	0.53
		COV	69.3	65.9	61.7	62.3	COV	69.3	72.9	71.6	71.4
		LEN	2.19	3.05	3.08	3.25	LEN	2.15	3.70	4.13	4.42
UR	Raw	PIT	0.14	0.00	0.00	0.00	PIT	0.75	0.05	0.04	0.00
		COV	64.8	68.2	63.0	54.6	COV	71.6	80.0	70.4	61.0
		LEN	0.32	1.09	1.98	2.37	LEN	0.40	1.26	2.26	2.86
	Baseline	PIT	0.00	0.00	0.00	0.00	PIT	0.01	0.02	0.02	0.00
		COV	69.3	74.1	61.7	49.4	COV	71.6	76.5	70.4	59.7
		LEN	0.26	0.98	1.85	2.30	LEN	0.26	1.11	2.11	2.78
	Hybrid	PIT	0.00	0.00	0.00	0.00	PIT	0.01	0.00	0.01	0.05
		COV	69.3	69.4	59.3	53.3	COV	70.5	75.3	65.4	66.2
		LEN	0.26	1.00	1.86	2.30	LEN	0.26	1.13	2.19	2.93
FFR	Raw	PIT	0.12	0.00	0.00	0.00	PIT	0.13	0.00	0.00	0.00
		COV	80.7	60.0	60.5	55.8	COV	80.7	70.6	67.9	61.0
		LEN	0.66	2.32	4.07	4.96	LEN	0.66	2.55	4.69	5.98
	Baseline	PIT	0.01	0.00	0.00	0.00	PIT	0.01	0.00	0.00	0.00
		COV	77.3	64.7	58.0	54.6	COV	76.1	76.5	67.9	59.7
		LEN	0.15	1.89	3.73	4.76	LEN	0.13	2.05	4.31	5.73
	Hybrid	PIT	0.01	0.00	0.00	0.00	PIT	0.01	0.00	0.00	0.00
		COV	78.4	67.1	61.7	63.6	COV	77.3	77.7	75.3	70.1
		LEN	0.15	1.93	3.77	4.79	LEN	0.13	2.11	4.45	5.95

Notes for Table: Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. PIT are the p-values from the statistical test of Knüppel (2015). The values less than 0.10 suggest instances where the hypothesis of correctly calibrated density forecasts is rejected at the 10% significance level. COV corresponds to the 70% empirical coverage rate; values closer to nominal value of 70% is preferred. LEN corresponds to the width of the 70% prediction intervals.

Table A7: **Calibration** Assessment of Density Forecasts: **Small TVP-VAR**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)											
		Small TVP VAR				Small TVP VAR with SV					
			h=1Q	h=4Q	h=8Q	h=12Q		h=1Q	h=4Q	h=8Q	h=12Q
GDP	Raw	PIT	0.00	0.00	0.00	0.00	PIT	0.16	0.03	0.23	0.60
		COV	88.6	84.7	90.1	92.2	COV	60.2	60.0	65.4	68.8
		LEN	6.55	7.12	7.57	7.78	LEN	3.28	3.84	4.12	4.20
	Baseline	PIT	0.37	0.00	0.00	0.00	PIT	0.45	0.03	0.22	0.49
		COV	72.7	87.1	90.1	90.9	COV	70.5	63.5	67.9	70.1
		LEN	3.50	7.03	7.53	7.72	LEN	3.25	3.80	4.17	4.30
	Hybrid	PIT	0.41	0.00	0.00	0.00	PIT	0.41	0.11	0.73	0.37
		COV	72.7	88.2	92.6	90.9	COV	71.6	67.1	70.4	70.1
		LEN	3.51	7.18	7.61	7.82	LEN	3.27	3.87	4.37	4.52
CPI Inflation	Raw	PIT	0.44	0.00	0.00	0.00	PIT	0.05	0.01	0.00	0.00
		COV	70.5	87.1	87.7	89.6	COV	50.0	72.9	74.1	76.6
		LEN	3.80	5.88	6.98	7.34	LEN	2.68	4.10	4.76	5.10
	Baseline	PIT	0.75	0.00	0.00	0.00	PIT	0.97	0.03	0.00	0.00
		COV	69.3	85.9	87.7	88.3	COV	68.2	75.3	75.3	72.7
		LEN	2.22	5.52	6.83	7.27	LEN	2.11	4.01	4.80	5.18
	Hybrid	PIT	0.81	0.00	0.00	0.00	PIT	0.94	0.10	0.02	0.01
		COV	70.5	89.4	92.6	92.2	COV	68.2	80.0	76.5	83.1
		LEN	2.23	5.74	7.22	7.70	LEN	2.10	4.19	5.08	5.58
UR	Raw	PIT	0.02	0.00	0.00	0.00	PIT	0.00	0.00	0.00	0.00
		COV	77.3	92.9	82.7	72.7	COV	46.6	48.2	40.7	44.2
		LEN	0.54	1.76	2.68	3.13	LEN	0.28	0.90	1.60	2.08
	Baseline	PIT	0.01	0.00	0.00	0.00	PIT	0.01	0.00	0.00	0.00
		COV	73.9	89.4	85.2	71.4	COV	72.7	60.0	39.5	37.7
		LEN	0.27	1.49	2.56	3.06	LEN	0.26	0.87	1.57	2.08
	Hybrid	PIT	0.02	0.00	0.00	0.00	PIT	0.01	0.00	0.00	0.00
		COV	71.6	90.6	85.2	80.5	COV	72.7	63.5	49.4	46.8
		LEN	0.27	1.51	2.61	3.22	LEN	0.26	0.88	1.63	2.23

Notes for Table: Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. PIT are the p-values from the statistical test of Knüppel (2015). The values less than 0.10 suggest instances where the hypothesis of correctly calibrated density forecasts is rejected at the 10% significance level. COV corresponds to the 70% empirical coverage rate; values closer to nominal value of 70% is preferred. LEN corresponds to the width of the 70% prediction intervals.

Table A8: **Calibration** Assessment of Density Forecasts: **Medium VAR est. 1960**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)																						
Medium VAR			h=1Q				h=4Q				h=8Q				h=12Q							
			Medium VAR with SV				h=1Q				h=4Q				h=8Q				h=12Q			
GDP	Raw	PIT	0.00	0.00	0.00	0.00	PIT	0.36	0.00	0.00	0.00	PIT	0.37	0.01	0.00	0.00	PIT	0.36	0.00	0.00	0.00	
		COV	87.5	87.1	91.4	90.9	COV	69.3	72.9	81.5	83.1	COV	71.6	70.6	79.0	81.8	COV	62.5	75.3	79.0	81.8	
		LEN	5.72	6.70	7.10	7.31	LEN	4.08	5.24	5.80	6.11	LEN	3.39	5.15	5.65	6.04	LEN	3.15	4.14	4.92	5.59	
	Baseline	PIT	0.49	0.00	0.00	0.00	PIT	0.37	0.01	0.00	0.00	PIT	0.91	0.30	0.10	0.00	PIT	0.91	0.30	0.10	0.00	
		COV	72.7	87.1	92.6	89.6	COV	71.6	70.6	79.0	81.8	COV	68.2	74.1	75.3	81.8	COV	72.7	77.7	70.4	66.2	
		LEN	3.54	6.53	7.11	7.20	LEN	3.39	5.15	5.65	6.04	LEN	2.16	3.95	4.73	5.36	LEN	0.26	1.10	1.90	2.53	
	Hybrid	PIT	0.60	0.00	0.00	0.00	PIT	0.46	0.10	0.03	0.01	PIT	0.91	0.11	0.02	0.00	PIT	0.01	0.00	0.00	0.00	
		COV	76.1	87.1	91.4	88.3	COV	71.6	75.3	77.8	81.8	COV	68.2	77.7	79.0	85.7	COV	70.5	78.8	72.8	62.3	
		LEN	3.57	6.66	6.97	7.28	LEN	3.39	5.29	5.95	6.31	LEN	2.15	4.02	4.92	5.57	LEN	0.26	1.11	1.99	2.64	
CPI Inflation	Raw	PIT	0.76	0.05	0.00	0.00	PIT	0.58	0.32	0.05	0.00	PIT	0.52	0.02	0.00	0.00	PIT	0.52	0.02	0.00	0.00	
		COV	63.6	76.5	81.5	83.1	COV	62.5	75.3	79.0	81.8	COV	77.3	72.9	64.2	54.6	COV	94.3	87.1	81.5	70.1	
		LEN	3.05	4.50	5.44	6.01	LEN	3.15	4.14	4.92	5.59	LEN	0.76	2.41	3.97	4.90	LEN	1.57	3.74	5.32	6.14	
	Baseline	PIT	0.82	0.11	0.00	0.00	PIT	0.91	0.30	0.10	0.00	PIT	0.01	0.01	0.00	0.00	PIT	0.01	0.00	0.00	0.00	
		COV	71.6	77.7	82.7	84.4	COV	68.2	74.1	75.3	81.8	COV	72.7	77.7	70.4	66.2	COV	78.4	84.7	80.3	68.8	
		LEN	2.22	4.28	5.36	5.89	LEN	2.16	3.95	4.73	5.36	LEN	0.13	2.04	3.74	4.75	LEN	0.15	3.33	5.08	5.94	
	Hybrid	PIT	0.88	0.07	0.00	0.00	PIT	0.91	0.11	0.02	0.00	PIT	0.01	0.00	0.00	0.00	PIT	0.00	0.00	0.00	0.00	
		COV	72.7	77.7	86.4	89.6	COV	68.2	77.7	79.0	85.7	COV	71.6	82.4	69.1	59.7	COV	79.6	85.9	79.0	68.8	
		LEN	2.23	4.31	5.52	6.20	LEN	2.15	4.02	4.92	5.57	LEN	0.26	1.11	1.99	2.64	LEN	0.16	3.33	5.22	6.20	
UR	Raw	PIT	0.42	0.00	0.00	0.00	PIT	0.33	0.01	0.00	0.00	PIT	0.33	0.01	0.00	0.00	PIT	0.33	0.01	0.00	0.00	
		COV	72.7	81.2	71.6	62.3	COV	67.1	71.8	70.4	68.8	COV	67.1	71.8	70.4	68.8	COV	67.1	71.8	70.4	68.8	
		LEN	0.40	1.37	2.15	2.61	LEN	0.39	1.19	2.01	2.61	LEN	0.39	1.19	2.01	2.61	LEN	0.39	1.19	2.01	2.61	
	Baseline	PIT	0.01	0.00	0.00	0.00	PIT	0.01	0.01	0.00	0.00	PIT	0.01	0.01	0.00	0.00	PIT	0.01	0.01	0.00	0.00	
		COV	68.2	76.5	72.8	63.6	COV	72.7	77.7	70.4	66.2	COV	72.7	77.7	70.4	66.2	COV	72.7	77.7	70.4	66.2	
		LEN	0.27	1.24	2.05	2.49	LEN	0.26	1.10	1.90	2.53	LEN	0.26	1.10	1.90	2.53	LEN	0.26	1.10	1.90	2.53	
	Hybrid	PIT	0.00	0.00	0.00	0.00	PIT	0.01	0.00	0.00	0.00	PIT	0.01	0.00	0.00	0.00	PIT	0.01	0.00	0.00	0.00	
		COV	70.5	78.8	72.8	62.3	COV	71.6	82.4	69.1	59.7	COV	71.6	82.4	69.1	59.7	COV	71.6	82.4	69.1	59.7	
		LEN	0.26	1.24	2.07	2.48	LEN	0.26	1.11	1.99	2.64	LEN	0.26	1.11	1.99	2.64	LEN	0.26	1.11	1.99	2.64	
FFR	Raw	PIT	0.00	0.00	0.00	0.00	PIT	0.52	0.02	0.00	0.00	PIT	0.52	0.02	0.00	0.00	PIT	0.52	0.02	0.00	0.00	
		COV	94.3	87.1	81.5	70.1	COV	77.3	72.9	64.2	54.6	COV	77.3	72.9	64.2	54.6	COV	77.3	72.9	64.2	54.6	
		LEN	1.57	3.74	5.32	6.14	LEN	0.76	2.41	3.97	4.90	LEN	0.76	2.41	3.97	4.90	LEN	0.76	2.41	3.97	4.90	
	Baseline	PIT	0.01	0.00	0.00	0.00	PIT	0.01	0.01	0.00	0.00	PIT	0.01	0.01	0.00	0.00	PIT	0.01	0.01	0.00	0.00	
		COV	78.4	84.7	80.3	68.8	COV	76.1	75.3	66.7	52.0	COV	76.1	75.3	66.7	52.0	COV	76.1	75.3	66.7	52.0	
		LEN	0.15	3.33	5.08	5.94	LEN	0.13	2.04	3.74	4.75	LEN	0.13	2.04	3.74	4.75	LEN	0.13	2.04	3.74	4.75	
	Hybrid	PIT	0.00	0.00	0.00	0.00	PIT	0.00	0.00	0.00	0.00	PIT	0.00	0.00	0.00	0.00	PIT	0.00	0.00	0.00	0.00	
		COV	79.6	85.9	79.0	68.8	COV	78.4	67.1	67.9	68.8	COV	78.4	67.1	67.9	68.8	COV	78.4	67.1	67.9	68.8	
		LEN	0.16	3.33	5.22	6.20	LEN	0.13	2.09	3.92	4.98	LEN	0.13	2.09	3.92	4.98	LEN	0.13	2.09	3.92	4.98	

Notes for Table: Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. PIT are the p-values from the statistical test of Knüppel (2015). The values less than 0.10 suggest instances where the hypothesis of correctly calibrated density forecasts is rejected at the 10% significance level. COV corresponds to the 70% empirical coverage rate; values closer to nominal value of 70% is preferred. LEN corresponds to the width of the 70% prediction intervals.

A3. Density Forecast Assessment based on Log-Score metric

The two widely used metrics for assessing the relative accuracy for density forecasts are the CRPS and Log-score. In the main body of the chapter we report the density forecast accuracy based on the CRPS metric. In this section we report the accuracy comparison results based on log-score. Log-score is the logarithm of the predictive density evaluated at the actual realization. Higher values are preferred to lower. As a reminder, in the case of CRPS, lower values are preferred to higher. The log-scores are computed based on normal kernel approximation method (for untilted densities, i.e. RAW forecasts, both parametric normal approximation and normal kernel approximation methods gives similar scores).

Tables A9 to A11 report the density forecast accuracy results based on the log-score. The numbers reported in the row labeled Raw are the mean logarithmic score of the Raw forecasts, all other numbers are relative log-Score. The relative log-score is the difference between average log-scores of the two forecasts. A positive value for Relative Log-Score suggests that on average the Baseline forecast is more accurate compared to the Raw forecast (in the case of Baseline - Raw); Hybrid more accurate compared to Raw (in the case of Hybrid - Raw); Hybrid more accurate compared to Baseline (in the case of Hybrid - Baseline). The left panel reports results for time-invariant Small VAR and right panel reports results corresponding to Small VAR with stochastic volatility.

The assessment based on log-score metric generally confirms the inference based on CRPS metric. In that Baseline density forecasts are more accurate compared to Raw forecasts, and density forecasts derived from the Hybrid forecasts are more accurate compared to both the Raw and Baseline forecasts. However, assessment based on log-score point to slightly fewer statistical significant gains which we suspect is partly due to the use of likelihood-ratio statistical test of Amisano and Giacomini (2007) that is more suitable for the rolling estimation scheme.

The Amisano and Giacomini (2007) test consists of regressing the differences in the log-scores (between two forecasts) on a constant and testing whether the constant is different from zero. The test statistic is based on a two-sided t-test, and are computed using the Newey-West variance estimator (i.e. serial-correlation robust variance of the regression constant). This test is suitable for forecasts generated from models that are estimated with rolling estimation scheme. This is not the case here because forecasts are generated using recursive scheme therefore the statistical significance should only be viewed as a rough approximation.

Table A9: **Density** Forecasting using **Log-Score** metric : **Small BVAR est. 1960**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Small VAR				Small VAR with SV			
	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
GDP								
Raw	-2.25	-2.46	-2.46	-2.44	-2.09	-2.33	-2.40	-2.42
Relative Log-Score								
Baseline - Raw	0.32***	-0.06	0.02	0.01	0.11**	-0.33	-0.04	-0.06
Hybrid - Raw	0.33**	0.04	0.00	0.01	0.12**	-0.03	-0.01	-0.06
Hybrid - Baseline	0.01	0.09	-0.02	-0.01	0.01	0.30	0.03	0.00
CPI								
Raw	-1.93	-2.21	-2.38	-2.47	-1.93	-2.17	-2.32	-2.39
Relative Log-Score								
Baseline - Raw	0.49***	0.06	0.05	0.00	0.50***	0.07	0.00	0.02
Hybrid - Raw	0.48***	0.11***	0.17***	0.10	0.50***	0.05	0.03	0.06*
Hybrid - Baseline	-0.01	0.06	0.12***	0.10*	0.00	-0.01	0.03	0.05
UR								
Raw	0.08	-1.08	-1.87	-2.25	0.18	-0.99	-2.00	-2.28
Relative Log-Score								
Baseline - Raw	0.50***	0.13***	0.22*	0.20	0.41***	0.17**	0.26	0.16
Hybrid - Raw	0.49***	0.21***	0.25	0.17	0.41***	0.20**	0.37	0.16
Hybrid - Baseline	-0.01	0.08	0.03	-0.03	0.00	0.03	0.11	0.00
FFR								
Raw	-0.86	-1.88	-2.37	-2.62	-0.15	-1.71	-2.47	-2.82
Relative Log-Score								
Baseline - Raw	2.19***	0.15***	0.03	-0.04	1.54***	0.22***	0.13***	0.05*
Hybrid - Raw	2.12***	0.15***	0.05	0.11**	1.56***	0.23***	0.28***	0.32***
Hybrid - Baseline	-0.06	0.00	0.03	0.14***	0.02	0.01	0.15**	0.27***
Credit Spread								
Raw	-0.10	-1.18	-1.62	-1.78	-0.02	-1.49	-1.95	-2.22
Relative Log-Score								
Baseline - Raw	0.05	0.00	0.04	0.01	0.02	0.02	0.11	-0.29*
Hybrid - Raw	0.07	0.06	0.13	0.19**	0.05	0.21	0.35	0.08
Hybrid - Baseline	0.02	0.06	0.09	0.18***	0.03	0.19*	0.24***	0.37***

Notes for Table: Raw forecast is defined as the unconditional forecast from the BVAR. Baseline forecast is defined as the raw BVAR forecast tilted towards survey nowcasts only. Hybrid forecast is defined as the raw BVAR forecast tilted towards both survey nowcasts and long-horizon forecasts. The numbers reported in the row labeled Raw are the mean logarithmic score (Log-Score), all other numbers are relative Log-Score; relative to Raw or relative to Baseline. A positive value for Relative Log-Score suggests that on average the Baseline forecast is more accurate compared to the Raw forecast (in the case of Baseline - Raw); Hybrid more accurate compared to Raw (in the case of Hybrid - Raw); Hybrid more accurate compared to Baseline (in the case of Hybrid - Baseline). The left panel reports results for time-invariant Small VAR and right panel reports results corresponding to Small VAR with stochastic volatility. The table reports statistical significance of the equal predictive accuracy based on testing whether the constant term in the regression of the differences in the log-score on the constant is statistically different from zero. The test statistics are based on a two-sided t-test, and are computed using the Newey-West variance estimator (i.e. serial-correlation robust variance of the regression constant); *10 %, **5 %, ***1 % significance levels.

Table A10: **Density** Forecasting using **Log-Score** metric : **Small BVAR est. 1985**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Small VAR				Small VAR with SV			
	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
GDP								
Raw	-2.04	-2.33	-2.37	-2.37	-2.07	-2.32	-2.40	-2.41
Relative Log-Score								
Baseline - Raw	0.14**	0.05	-0.07***	-0.08	0.14**	-0.16	-0.03	-0.01
Hybrid - Raw	0.12*	-0.06	0.08	-0.04	0.15**	-0.18	0.02	-0.01
Hybrid - Baseline	-0.02	-0.10	0.15	0.04	0.01	-0.02	0.05*	0.00
CPI								
Raw	-1.91	-2.16	-2.21	-2.30	-1.89	-2.12	-2.15	-2.22
Relative Log-Score								
Baseline - Raw	0.40***	0.00	-0.03	0.01	0.47***	-0.01	0.00	-0.03
Hybrid - Raw	0.42***	0.04	0.14***	0.09*	0.47***	0.02	0.05	0.00
Hybrid - Baseline	0.02	0.03	0.17***	0.08**	0.00	0.03	0.05	0.03
UR								
Raw	0.13	-0.97	-1.92	-2.42	0.19	-0.98	-1.82	-2.30
Relative Log-Score								
Baseline - Raw	0.32***	0.15	0.20	0.04	0.41***	0.24***	0.03	0.03
Hybrid - Raw	0.35***	0.07	0.09	-0.06	0.40***	0.10**	0.00	0.01
Hybrid - Baseline	0.03	-0.08	-0.12	-0.10	-0.01	-0.14*	-0.03	-0.02
FFR								
Raw	-0.20	-1.79	-2.43	-2.64	-0.07	-1.69	-2.42	-2.70
Relative Log-Score								
Baseline - Raw	1.48***	0.27***	0.05	-0.09*	1.46***	0.26***	0.10**	0.05
Hybrid - Raw	1.39***	0.29***	0.12	0.07	1.51***	0.23***	0.17***	0.08
Hybrid - Baseline	-0.09	0.02	0.07	0.17*	0.05	-0.03	0.07	0.03
Credit Spread								
Raw	-0.11	-1.33	-1.75	-2.00	-0.10	-1.28	-1.52	-1.75
Relative Log-Score								
Baseline - Raw	0.14*	0.15	-0.04	-0.25*	0.08**	0.11	-0.04	-0.13*
Hybrid - Raw	0.17**	0.20	0.19	-0.12	0.07**	0.10	-0.03	-0.11
Hybrid - Baseline	0.03	0.05	0.23	0.13	-0.01	-0.01	0.01	0.02

Notes for Table: Raw forecast is defined as the unconditional forecast from the BVAR. Baseline forecast is defined as the raw BVAR forecast tilted towards survey nowcasts only. Hybrid forecast is defined as the raw BVAR forecast tilted towards both survey nowcasts and long-horizon forecasts. The numbers reported in the row labeled Raw are the mean logarithmic score (Log-Score), all other numbers are relative Log-Score; relative to Raw or relative to Baseline. A positive value for Relative Log-Score suggests that on average the Baseline forecast is more accurate compared to the Raw forecast (in the case of Baseline - Raw); Hybrid more accurate compared to Raw (in the case of Hybrid - Raw); Hybrid more accurate compared to Baseline (in the case of Hybrid - Baseline). The left panel reports results for time-invariant Small VAR and right panel reports results corresponding to Small VAR with stochastic volatility. The table reports statistical significance of the equal predictive accuracy based on testing whether the constant term in the regression of the differences in the log-score on the constant is statistically different from zero. The test statistics are based on a two-sided t-test, and are computed using the Newey-West variance estimator (i.e. serial-correlation robust variance of the regression constant); *10 %, **5 %, ***1 % significance levels.

Table A11: **Density** Forecasting using **Log-Score** metric : **Small TVP-VAR**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Small TVP VAR				Small TVP VAR with SV			
	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
GDP								
Raw	-2.31	-2.46	-2.46	-2.46	-2.16	-2.41	-2.35	-2.35
Relative Log-Score								
Baseline - Raw	0.31***	0.02**	-0.01	-0.01	0.20***	0.05	-0.07	-0.08
Hybrid - Raw	0.32***	0.00	0.00	-0.03**	0.18***	0.06	-0.07	-0.23
Hybrid - Baseline	0.01	-0.02	0.01	0.02**	-0.01	0.02	0.00	-0.15
CPI								
Raw	-2.02	-2.34	-2.43	-2.50	-2.13	-2.18	-2.25	-2.30
Relative Log-Score								
Baseline - Raw	0.62***	0.08***	0.02	0.03	0.35***	0.11***	0.03	-0.02
Hybrid - Raw	0.62***	0.11***	0.05***	0.07**	0.34***	0.16***	0.07	0.03
Hybrid - Baseline	0.00	0.03*	0.03**	0.04	-0.01	0.05*	0.04	0.05
UR								
Raw	0.08	-1.12	-1.81	-2.09	-0.02	-1.15	-2.00	-2.31
Relative Log-Score								
Baseline - Raw	0.52***	0.14***	0.03	-0.03	0.62***	0.18***	-0.02	0.07
Hybrid - Raw	0.50***	0.10**	-0.03	-0.17	0.61***	0.12**	0.00	0.01
Hybrid - Baseline	-0.01	-0.04	-0.05	-0.14	-0.01	-0.06	0.03	-0.06

Raw forecast is defined as the unconditional forecast from the BVAR. Baseline forecast is defined as the raw BVAR forecast tilted towards survey nowcasts only. Hybrid forecast is defined as the raw BVAR forecast tilted towards both survey nowcasts and long-horizon forecasts. The numbers reported in the row labeled Raw are the mean logarithmic score (Log-Score), all other numbers are relative Log-Score; relative to Raw or relative to Baseline. A positive value for Relative Log-Score suggests that on average the Baseline forecast is more accurate compared to the Raw forecast (in the case of Baseline - Raw); Hybrid more accurate compared to Raw (in the case of Hybrid - Raw); Hybrid more accurate compared to Baseline (in the case of Hybrid - Baseline). The left panel reports results for Small time-varying VAR (with constant variance) and right panel reports results corresponding to Small time-varying VAR with stochastic volatility. The table reports statistical significance of the equal predictive accuracy based on testing whether the constant term in the regression of the differences in the log-score on the constant is statistically different from zero. The test statistics are based on a two-sided t-test, and are computed using the Newey-West variance estimator (i.e. serial-correlation robust variance of the regression constant); *10 %, **5 %, ***1 % significance levels.

A4. Ranking the Models: Before tilting vs. Post-tilting

The focus of our results is to establish that our proposal is working on a range of model specifications. Given we are considering several model specifications, that gives rise to additional curiosities such as the usefulness of additional variables in improving forecasts of core variables, importance of stochastic volatility, the role of estimation sample, importance of time-varying parameters etc. Is there any one model that dominates others in forecasting core variables of interest? And relatedly, is tilting more beneficial for models that are relatively inferior in Raw form?

These curiosities can be answered by collectively looking at all the tables included in the

main body of the chapter, and tables included in section A1 of this appendix. To provide a more convenient way to help answer some of these curiosities, we have summarized in table A12 the point forecast accuracy metrics (using RMSE) for all the models we consider.

The results reported in this table can be used to directly answer some of the above questions. For example, if a forecaster is interested in obtaining accurate forecasts for real GDP growth, CPI inflation, and unemployment rate using a single multivariate model, then an optimal choice would be to pick a Small VAR estimated using post-1985 data (if interest is only on point forecasts) or Small VAR with SV estimated using post-1985 data (if interest is both on point and density forecasts; evidence in rest of the chapter supports the use of SV for density forecasts and calibration).

The results in the table indicate that additional predictors help improve the forecast accuracy for core variables but only when VAR is estimated with longer sample of data (confirming results in Banbura et al, 2010; Koop, 2013; and Carriero et al, 2015a; all these studies focused on estimation with longer sample). But in contrast, for VAR models estimated with post-85 sample these additional predictors do not appear to be helpful (e.g. Small VAR competitive to Medium VAR).

The information reported in columns ‘% change in RMSE (indicating percentage gains in forecast accuracy via tilting’, ‘Rank Raw’ and ‘Rank Hybrid’ of the table could be used to answer the specific question, ‘is tilting more beneficial for models that are relatively inferior in Raw form?’ For medium-horizon ($h=8Q$, and $h=12Q$ -not shown), we find that our (Hybrid) proposal helps more models that perform worse in Raw form. For short-horizon, rankings are generally maintained post-tilting, i.e. models ranked low prior to tilting continues to rank low post-tilting but the differences across accuracy are much smaller. As discussed in the main body of the chapter (and Clark and McCracken, 2008, Kozicki and Tinsley, 2001; Clements and Hendry, 1999), beyond 4 or 5 quarters, the forecasts are increasingly influenced by the model’s implied steady-state/trend, and given our proposal is targeting the trend, the most gains from proposal would be expected for horizons beyond 4 quarters. As a further check on robustness of our finding, we estimate a Small VAR with loose priors and assess to what extent our proposal is helping improve its accuracy. It is well-known that such a specification of a VAR performs poorly (due to overparameterization) in forecasting. Then we would expect our proposal to help this particular VAR specification significantly more compared to the set of VAR models we consider that are estimated with relatively tighter Bayesian priors. Indeed, we find that this VAR specification (with loose priors) ranks the lowest before tilting, gains the most in terms of accuracy improvement from tilting, and remains ranked at the bottom post-tilting but with significantly reduced margin when compared to next best model. The results are reported in table A13

Table A12: Forecast Accuracy Assessment: Before Tilting (Raw) and Post-Tilting (Hybrid)

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
Real GDP								
Models	h=4Q				h=8Q			
	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE
Small VAR est 1960	3	2.7	4	-0.1	2	2.6	1	-8.2
Small VAR SV est 1960	2	2.6	1	-5.2	2	2.6	1	-7.2
Small VAR est 1985	1	2.5	2	+1.7	2	2.6	1	-8.9
Small VAR SV est 1985	1	2.5	1	-1.2	2	2.6	1	-7.7
Medium VAR est 1960	2	2.6	3	-1.5	2	2.6	1	-8.3
Medium VAR SV est 1960	2	2.6	1	-4.8	2	2.6	1	-9.4
Medium VAR est 1985	3	2.7	4	-0.1	3	2.8	1	-15.5
Medium VAR SV est 1985	2	2.6	1	-5.9	3	2.8	1	-13.4
TVP VAR	2	2.6	3	-0.9	1	2.5	1	-3.8
TVP SV VAR	2	2.6	1	-7.1	1	2.5	1	-5.1
CPI Inflation								
Models	h=4Q				h=8Q			
	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE
Small VAR est 1960	3	2.4	2	-8.4	4	2.6	2	-15.3
Small VAR SV est 1960	2	2.3	1	-7.7	2	2.3	2	-6.2
Small VAR est 1985	1	2.2	1	-2.1	2	2.3	1	-6.0
Small VAR SV est 1985	1	2.2	1	-4.5	1	2.2	1	-1.5
Medium VAR est 1960	2	2.3	1	-5.9	3	2.4	1	-10.2
Medium VAR SV est 1960	1	2.2	1	-2.8	1	2.2	1	-2.5
Medium VAR est 1985	2	2.3	3	+0.5	2	2.3	1	-8.8
Medium VAR SV est 1985	1	2.2	1	-2.5	1	2.2	1	-3.9
TVP VAR	4	2.5	2	-9.2	4	2.6	1	-16.8
TVP SV VAR	2	2.3	1	-6.3	2	2.3	1	-9.4
Unemployment rate								
Models	h=4Q				h=8Q			
	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE
Small VAR est 1960	2	0.8	2	-11.1	3	1.6	2	-5.2
Small VAR SV est 1960	2	0.8	2	-14.2	2	1.5	1	-6.7
Small VAR est 1985	1	0.7	2	-5.5	2	1.5	2	+0.9
Small VAR SV est 1985	1	0.7	1	-6.7	1	1.4	1	-1.8
Medium VAR est 1960	2	0.8	2	-8.9	3	1.6	2	-4.6
Medium VAR SV est 1960	2	0.8	2	-13.3	2	1.5	1	-6.4
Medium VAR est 1985	1	0.7	2	-4.1	2	1.5	2	+1.7
Medium VAR SV est 1985	1	0.7	1	-8.8	2	1.5	1	-3.5
TVP VAR	1	0.7	2	-4.9	1	1.4	1	+1.9
TVP SV VAR	1	0.7	1	-8.2	1	1.4	1	+0.1

Notes for Table: Rank Raw: rank of the raw (i.e. unconditional) forecast based on RMSE value (using one-decimal point); similarly Rank Hybrid is the rank of the hybrid forecast.

A5. Effect of Tilting on Small VAR with loose priors

Overall, our point and density forecasting results using real-time data provide compelling evidence that tilting VAR forecasts to match the long-run forecasts from the Survey of Professional Forecasters systematically leads to improved forecast accuracy for most variables over the forecast horizon of interest to monetary policymakers. Generally, it is the case that our proposal is helping more models that are performing worse in Raw form (i.e. raw VAR forecasts). Interestingly, rankings are generally maintained post-tilting for shorter horizons (e.g. $h=4Q$), i.e. models ranked low prior to tilting continues to rank low post-tilting but the differences across accuracy are much smaller. As an additional check on robustness of this statement, we estimate a Small VAR with loose priors⁵ and assess to what extent our proposal is helping improve its accuracy. We know that a VAR model estimated with post-1985 sample consisting of five variables and four lags would be severely over parametrized and so would produce significantly inferior forecasts. Accordingly, we would expect it to gain the most from tilting and yet we expect the accuracy of the hybrid forecast from this model to continue to rank at the bottom.

We confirm that indeed this is the case. We find that this VAR specification ranks the lowest (as would be expected) prior to tilting, gains the most in terms of accuracy improvement from tilting, and remains ranked at the bottom post-tilting but with significantly reduced margin when compared to next best model. The results of this exercise are reported in table A13.

⁵This is achieved by setting the hyper parameters $\mu=200$ and $\lambda=200$; compare this with the optimal values of $\mu=1$ and $\lambda=0.4$ used for Small VAR est 1985.

Table A13: Forecast Accuracy Assessment: Before Tilting (Raw) and Post-Tilting (Hybrid):
Added Small VAR with loose priors

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
Real GDP								
Models	h=4Q				h=8Q			
	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE
Small VAR est 1960	3	2.7	4	-0.1	2	2.6	1	-8.2
Small VAR SV est 1960	2	2.6	1	-5.2	2	2.6	1	-7.2
Small VAR est 1985	1	2.5	2	+1.7	2	2.6	1	-8.9
Small VAR SV est 1985	1	2.5	1	-1.2	2	2.6	1	-7.7
Medium VAR est 1960	2	2.6	3	-1.5	2	2.6	1	-8.3
Medium VAR SV est 1960	2	2.6	1	-4.8	2	2.6	1	-9.4
Medium VAR est 1985	3	2.7	4	-0.1	3	2.8	1	-15.5
Medium VAR SV est 1985	2	2.6	1	-5.9	3	2.8	1	-13.4
TVP VAR	2	2.6	3	-0.9	1	2.5	1	-3.8
TVP SV VAR	2	2.6	1	-7.1	1	2.5	1	-5.1
Small VAR 1985, loose priors	4	3.1	5	-6.4	4	3.6	1	-32.6
CPI Inflation								
Models	h=4Q				h=8Q			
	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE
Small VAR est 1960	3	2.4	2	-8.4	4	2.6	2	-15.3
Small VAR SV est 1960	2	2.3	1	-7.7	2	2.3	2	-6.2
Small VAR est 1985	1	2.2	1	-2.1	2	2.3	1	-6.0
Small VAR SV est 1985	1	2.2	1	-4.5	1	2.2	1	-1.5
Medium VAR est 1960	2	2.3	1	-5.9	3	2.4	1	-10.2
Medium VAR SV est 1960	1	2.2	1	-2.8	1	2.2	1	-2.5
Medium VAR est 1985	2	2.3	3	+0.5	2	2.3	1	-8.8
Medium VAR SV est 1985	1	2.2	1	-2.5	1	2.2	1	-3.9
TVP VAR	4	2.5	2	-9.2	4	2.6	1	-16.8
TVP SV VAR	2	2.3	1	-6.3	2	2.3	1	-9.4
Small VAR 1985, loose priors	3	2.4	4	-0.7	3	2.4	1	-12.5
Unemployment rate								
Models	h=4Q				h=8Q			
	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE	Rank Raw	RMSE Raw	Rank Hybrid	% Δ RMSE
Small VAR est 1960	2	0.8	2	-11.1	3	1.6	2	-5.2
Small VAR SV est 1960	2	0.8	2	-14.2	2	1.5	1	-6.7
Small VAR est 1985	1	0.7	2	-5.5	2	1.5	2	+0.9
Small VAR SV est 1985	1	0.7	1	-6.7	1	1.4	1	-1.8
Medium VAR est 1960	2	0.8	2	-8.9	3	1.6	2	-4.6
Medium VAR SV est 1960	2	0.8	2	-13.3	2	1.5	1	-6.4
Medium VAR est 1985	1	0.7	2	-4.1	2	1.5	2	+1.7
Medium VAR SV est 1985	1	0.7	1	-8.8	2	1.5	1	-3.5
TVP VAR	1	0.7	2	-4.9	1	1.4	1	+1.9
TVP SV VAR	1	0.7	1	-8.2	1	1.4	1	+0.1
Small VAR 1985, loose priors	3	0.9	3	-16.1	4	1.9	3	-16.6

Notes for Table: Rank Raw: rank of the raw (i.e. unconditional) forecast based on RMSE value (using one-decimal point); similarly Rank Hybrid is the rank of the hybrid forecast.

A6. Are there benefits to utilizing survey information for additional horizons?

The focus of our chapter is tilting the VAR model's trend forecasts towards the long-horizon survey forecasts and its implications on the forecast accuracy over the forecast horizon of interest to monetary policymakers (current quarter to three years out). In light of widely used practice of conditioning or tilting current quarter VAR forecasts (of quarterly models) towards external nowcasts (i.e. nowcasts informed from mixed-frequency models or surveys) we also tilt the VAR one-quarter ahead forecast to survey nowcasts. However, surveys such as SPF and BC also report forecasts for subsequent four quarters beyond the nowcast quarter. So this gives rise to a natural curiosity is there benefit in using these additional survey forecasts?

One can certainly proceed to impose these survey forecasts as additional restrictions on the VAR model forecasts, and the technique of relative entropy is well-suited to achieving such a proposal.

To assess the usefulness of these additional survey forecasts as conditions for the VAR is to compare their real-time accuracy with the hybrid forecast generated from our approach. Table A14 reports the Root Mean Squared Errors (RMSE) for the hybrid forecasts derived from various models considered in our empirical exercises along with the accuracy of the SPF forecast. Interestingly, we find that once the VAR models are tilted towards the survey nowcast and long-horizon forecasts, the resulting hybrid forecasts are generally competitive to the survey forecasts for the remaining horizons covered by the survey (i.e. $h=2Q$ through $h=5Q$). In other words, tilting VAR at the jumping-off horizon and at terminal horizon is sufficient to produce forecasts whose accuracy rival the survey forecasts for the remaining horizons for which survey forecasts are available. There are some exceptions for real GDP growth that we detail below.

Specifically, the forecast accuracy of hybrid forecasts for inflation and unemployment rate produced from most VAR models under consideration matches the accuracy of the survey forecasts. However, we note that in the case of real GDP growth, survey forecast outperform hybrid forecasts (from all VAR models) for forecast horizon immediately preceding the nowcast quarter (i.e. $h=2Q$), and hybrid forecasts from all VAR models except one (obtained from TVP-SV VAR) for $h=3Q$. This suggests that in the case of real GDP growth, there may be gains from tilting towards survey forecasts for additional forecast quarters (i.e. $h=2Q$ and $h=3Q$).

Based on the results reported in table A14, utilizing all these additional survey forecasts as conditions to generate the hybrid forecast is certainly a viable option. As doing so leads to a net benefit because all models experience accuracy gains in the two quarters following the nowcast quarter for real GDP growth. However, benefits for the remaining horizons (i.e. forecast horizons not covered by the survey, $h=6Q$ through $h=12Q$) are likely to be small especially

given that most models perform competitively beyond three quarters. In addition, for real GDP growth, the spillover effects of improved accuracy to future horizons would be limited due to the rapid mean-reverting nature of real GDP growth.

Our findings that survey forecasts are among the most accurate is consistent with a large number of papers who have demonstrated that survey forecasts are hard to outperform (e.g. Ang et al, 2007; Clark and McCracken, 2008; Croushore, 2010; Wright, 2013). As stressed in Wright (2013), the ability of an econometric approach to match the forecast accuracy of surveys is a huge advantage as the approach could then be used to generate forecasts for forecast horizons and variables not covered by the survey; and be used to produce forecasts at any point in time (compared to survey forecasts that are available once a month in the case of Blue Chip, and quarterly in the case of SPF). Another important advantage of the modeling approach is the ability to obtain higher moment forecasts, specifically density forecasts which have gained importance over the past few years coinciding with increasing interest in macroeconomic risks assessment.

Therefore, we view the ability of the hybrid forecasts (in this chapter) to rival the forecast accuracy of the survey forecasts (for the forecast horizons surveys are available) as a success. As one, we would be more confident in our hybrid forecast one to three years out (for forecast horizons covered by surveys) for our core variables; secondly, we can rely on model's forecasts for additional variables such as consumption or wage inflation which have gained increased prominence as of late (e.g. in recent years policy makers have been more directly asked about their forecasts for productivity and wages); thirdly, we will have more confidence in our risk assessment based on density forecasts obtained from our models (e.g. risk of deflation etc.).

It is fair to characterize results in the main body of the chapter as the lower bound of improvements from hybrid approach. One could certainly improve on our results (in terms of improved accuracy for hybrid forecasts), for example, by tilting the variance around the long-run means. These variance conditions (that far out) could be constructed using univariate unobserved components model that allow stochastic volatility or through a recently developed method of Clark, McCracken and Mertens (2018). Our framework is quite flexible and can easily handle any number of restrictions on the forecasts over the forecast horizon. To facilitate the use of our approach across a range of practitioners we will make available the code for tilting (in the form of a Matlab function) that could easily be adapted for use with any econometric model that has the ability to construct predictive densities.

Finally, we highlight a practical consideration that may play a role in deciding whether to use the SPF forecasts as conditions (in VARs) for additional forecast quarters such as real GDP. In the chapter, we use nowcasts informed from the survey so to be systematic in using survey information in our empirical exercises and to follow Krüger, Clark and Ravazzolo (2017) (KCR)

as this allows us to characterize our exercises as truly building on KCR (KCR perform two sets of exercises one using survey nowcasts and another using model-based nowcasts). However, in practice a modeler or forecaster who relies on quarterly VARs to produce forecasts would likely use other models (built specifically to do nowcasting) to produce nowcasts and use those as the starting conditions for the quarterly VAR. It is reasonable to assume that as information accumulates over the course of quarter, the nowcasts from external models to become more accurate (due to timing advantage) compared to SPF nowcast estimates that were formed based on dated information available through the beginning of the second month of the quarter. All this suggests, that it is a strong possibility that tilting the VAR forecasts towards an updated (and more accurate) nowcast informed from a nowcasting model (e.g. GDP) and survey long-horizon forecast results in a hybrid forecast whose accuracy for $h=2Q$ and $h=3Q$ matches or is more accurate than the survey forecast (made possible through the spillover effects from more accurate nowcast). If indeed this is the case then the benefit of using the additional survey conditions (e.g. $h=2Q$ and $h=3Q$ for real GDP growth) is reduced. Therefore, the usefulness of additional survey forecasts is a model-specific empirical matter.

Table A14: Real-Time Forecasting Performance: Hybrid vs. Survey of Professional Forecasters

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)					
Root Mean Squared Error (RMSE)					
Real GDP					
Model	h=1Q	h=2Q	h=3Q	h=4Q	h=5Q
Small VAR est 1960	1.6	2.2	2.5	2.7	2.3
Small VAR SV est 1960	1.6	2.1	2.3	2.4	2.4
Small VAR est 1985	1.6	2.0	2.3	2.5	2.3
Small VAR SV est 1985	1.6	2.1	2.3	2.4	2.3
Medium VAR est 1960	1.6	2.2	2.4	2.6	2.3
Medium VAR SV est 1960	1.6	2.1	2.3	2.4	2.4
Medium VAR est 1985	1.6	2.3	2.5	2.7	2.3
Medium VAR SV est 1985	1.6	2.2	2.3	2.4	2.4
TVP VAR	1.6	2.3	2.4	2.6	2.4
TVP SV VAR	1.6	2.1	2.2	2.4	2.4
Survey of Professional Forecasters	1.6	1.9	2.2	2.3	2.4
CPI Inflation					
Model	h=1Q	h=2Q	h=3Q	h=4Q	h=5Q
Small VAR est 1960	1.0	2.1	2.2	2.2	2.2
Small VAR SV est 1960	1.0	2.1	2.1	2.1	2.2
Small VAR est 1985	1.0	2.1	2.1	2.1	2.1
Small VAR SV est 1985	1.0	2.1	2.1	2.1	2.1
Medium VAR est 1960	1.0	2.1	2.1	2.1	2.1
Medium VAR SV est 1960	1.0	2.2	2.1	2.1	2.1
Medium VAR est 1985	1.0	2.1	2.2	2.3	2.1
Medium VAR SV est 1985	1.0	2.1	2.1	2.1	2.1
TVP VAR	1.0	2.2	2.2	2.2	2.2
TVP SV VAR	1.0	2.2	2.2	2.1	2.1
Survey of Professional Forecasters	1.0	2.0	2.1	2.1	2.1
Unemployment rate					
Model	h=1Q	h=2Q	h=3Q	h=4Q	h=5Q
Small VAR est 1960	0.1	0.3	0.5	0.7	1.0
Small VAR SV est 1960	0.1	0.3	0.4	0.7	0.9
Small VAR est 1985	0.1	0.3	0.4	0.7	0.9
Small VAR SV est 1985	0.1	0.3	0.4	0.6	0.8
Medium VAR est 1960	0.1	0.3	0.5	0.7	0.9
Medium VAR SV est 1960	0.1	0.3	0.4	0.7	0.9
Medium VAR est 1985	0.1	0.3	0.5	0.7	0.9
Medium VAR SV est 1985	0.1	0.3	0.4	0.6	0.9
TVP VAR	0.1	0.3	0.4	0.7	0.9
TVP SV VAR	0.1	0.3	0.4	0.6	0.9
Survey of Professional Forecasters	0.1	0.3	0.4	0.7	0.9

Notes for Table: Entries in bold indicate the lowest RMSE for a given forecast horizon.

A7. Hybrid vs. Federal Reserve's GreenBook

The staff at the Federal Reserve Board of Governors produces Greenbook (GB) forecasts roughly a week and a half prior to each FOMC meeting. The FOMC meetings are held eight times a year, with roughly two meetings scheduled per quarter. These meetings are irregularly spaced and therefore any forecasting exercise comparing GB to alternative methods needs to account meticulously for comparability of the information sets at forecast origins. GB forecasts are released to the public with a 5-year delay. The historical forecasts could be downloaded either from the website of the Federal Reserve Board of Governors or from the website of the Federal Reserve Bank of Philadelphia.

GB forecast can be thought of as a combination of model and judgement, because judgment adjusts the base model forecasts. Therefore, it will be expected to handle structural change better than standard VARs. Put differently, we would expect the pattern of forecast accuracy improvements to somewhat echo the pattern of forecast accuracy seen for hybrid forecasts (illustrated in the main body of the chapter). We confirm that indeed this is the case, and strikingly, the hybrid forecasts for real GDP growth and CPI inflation from our simple VARs are competitive to the GB forecasts. In the case of unemployment rate, hybrid forecast under-performs GB during the Great Recession but on average is competitive to GB.

To perform a fair comparison, we collect GB forecasts corresponding to the FOMC meeting days closer to the SPF survey dates. Specifically, we collect GB forecasts either prepared in the second or third month of the quarter. GB forecasts prepared closer to the end of second month or anytime in the third month will have an informational advantage over the SPF. That is, GB is likely to have more accurate nowcast estimates (i.e. jumping-off points) compared to the SPF and by extension our VAR models (which uses SPF nowcasts to produce both baseline and hybrid forecasts). We could fix this by further restricting the set of GB forecasts for evaluation to only those that are prepared (for FOMC) around a narrow window of the SPF release date. Instead, we let the GB have more accurate starting points. This is done for two main reasons: (1) our focus is on the medium-term forecast horizons (2) doing so would have significantly reduced the set of available forecasts for comparison.

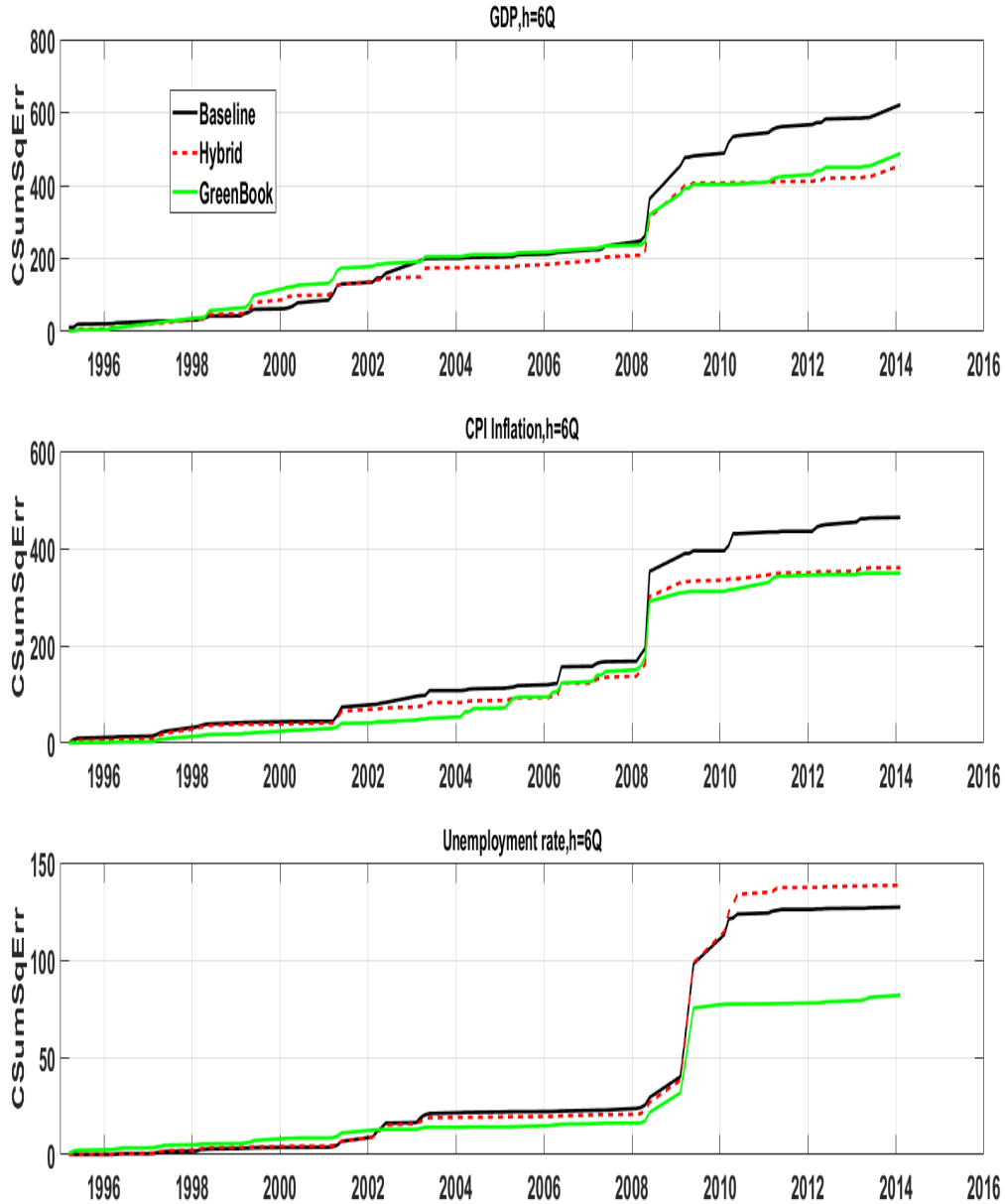
Another issue worth mentioning is the maximum horizon of the forecast reported in the GB varies across meetings, but all the forecasts (over our forecast sample period of interest) at least goes through $h=6Q$. Therefore, in our forecast comparison exercises with GB we consider up $h=1Q$ through $h=6Q$.

Figure A1 plots the cumulative sum of squared errors for GB forecast, baseline forecast, and the hybrid forecast respectively for forecast horizon $h=6Q$. Both the baseline and hybrid forecasts are obtained from the Small VAR (1960). We deliberately feature this model as it ranks at the bottom of the list of the 10 model specifications we consider. The figure illustrates the ability of both hybrid and GB forecasts to adjust in the face of structural change of 2008-2009. It is quite interesting that both hybrid and GB forecasts perform comparably for real GDP growth and CPI inflation and are more accurate than the baseline VAR forecast. In the case

of unemployment rate, for most part all three perform comparably as evidenced by a stable (or negligible) gap across the three. The differences in the accuracy arise around the Great Recession period, where the GB is substantially more accurate compared to hybrid. The substantially more accurate GB forecast for the few quarters (around the crisis period) contributed to the gap between GB and VAR forecasts. Those improvements (for just few quarters) were large enough to lower the average RMSE for the GB by two to three tenths lower compared to the hybrid forecast (1.0 vs. 1.2; for $h=6Q$), see table A15. As the recovery from the Great Recession progressed, all three performed comparably.

Based on the RMSE reported in the table, and ignoring the nowcast quarter, for real GDP growth the accuracy is competitive to VAR models across all horizons shown. This result is consistent with Faust and Wright (2009) who showed that beyond the current quarter, GB forecasts of real GDP performed comparably to other forecasts obtained from atheoretic models. In contrast, Faust and Wright document that inflation forecasts (in their case GDP deflator) were the most accurate and hard to outperform. In contrast, we find that beyond two quarters, the hybrid forecasts from all models are competitive (and in some cases slightly more accurate) compared to the GB. For the unemployment rate, GB is among the most accurate through $h=3Q$, and just slightly more accurate for the remaining forecast horizons. For the unemployment rate, we are unable to relate our findings to Faust and Wright since they only focused on real GDP and inflation.

Figure A1: Cumulative Squared Errors, GreenBook vs. Baseline and Hybrid (Small VAR 1960)



Notes for the figure: Cumulative sum of squared errors at $h=6Q$, for Greenbook forecast (solid green); baseline (solid black), and hybrid forecasts (dotted red) both obtained from Small VAR with constant variance estimated with longer sample. The x-axis dates reflect the forecast evaluation quarters.

Table A15: Real-Time Forecasting Performance: Hybrid vs. Green Book

Full Sample (Recursive evaluation: 1994.Q1-2014.Q1)						
Root Mean Squared Error (RMSE)						
Real GDP						
Model	h=1Q	h=2Q	h=3Q	h=4Q	h=5Q	h=6Q
Small VAR est 1960	1.6	2.2	2.6	2.7	2.4	2.5
Small VAR SV est 1960	1.6	2.2	2.4	2.5	2.4	2.5
Small VAR est 1985	1.6	2.1	2.4	2.6	2.4	2.5
Small VAR SV est 1985	1.6	2.1	2.3	2.5	2.3	2.5
Medium VAR est 1960	1.6	2.2	2.5	2.6	2.4	2.5
Medium VAR SV est 1960	1.6	2.1	2.3	2.5	2.4	2.5
Medium VAR est 1985	1.6	2.4	2.6	2.7	2.4	2.5
Medium VAR SV est 1985	1.6	2.3	2.3	2.5	2.4	2.6
TVP VAR	1.6	2.3	2.5	2.6	2.5	2.6
TVP SV VAR	1.6	2.1	2.3	2.4	2.4	2.5
Survey of Professional Forecasters	1.6	1.9	2.3	2.4	2.4	–
Green Book	1.4	2.0	2.3	2.4	2.5	2.5
CPI Inflation						
Model	h=1Q	h=2Q	h=3Q	h=4Q	h=5Q	h=6Q
Small VAR est 1960	1.0	2.2	2.2	2.2	2.2	2.2
Small VAR SV est 1960	1.0	2.2	2.2	2.2	2.3	2.2
Small VAR est 1985	1.0	2.2	2.2	2.2	2.1	2.1
Small VAR SV est 1985	1.0	2.2	2.2	2.1	2.1	2.1
Medium VAR est 1960	1.0	2.2	2.2	2.2	2.2	2.1
Medium VAR SV est 1960	1.0	2.2	2.2	2.1	2.1	2.2
Medium VAR est 1985	1.0	2.2	2.2	2.3	2.1	2.1
Medium VAR SV est 1985	1.0	2.1	2.2	2.2	2.1	2.1
TVP VAR	1.0	2.2	2.2	2.3	2.2	2.1
TVP SV VAR	1.0	2.2	2.2	2.2	2.1	2.1
Survey of Professional Forecasters	1.0	2.0	2.1	2.1	2.1	–
Green Book	0.5	1.7	2.3	2.2	2.1	2.1
Unemployment rate						
Model	h=1Q	h=2Q	h=3Q	h=4Q	h=5Q	h=6Q
Small VAR est 1960	0.1	0.3	0.5	0.8	1.0	1.2
Small VAR SV est 1960	0.1	0.3	0.4	0.7	0.9	1.2
Small VAR est 1985	0.1	0.3	0.4	0.7	0.9	1.2
Small VAR SV est 1985	0.1	0.3	0.4	0.7	0.9	1.1
Medium VAR est 1960	0.1	0.3	0.5	0.7	1.0	1.2
Medium VAR SV est 1960	0.1	0.3	0.5	0.7	0.9	1.2
Medium VAR est 1985	0.1	0.3	0.5	0.7	1.0	1.2
Medium VAR SV est 1985	0.1	0.3	0.4	0.7	0.9	1.1
TVP VAR	0.1	0.3	0.4	0.7	0.9	1.1
TVP SV VAR	0.1	0.3	0.4	0.7	0.9	1.1
Survey of Professional Forecasters	0.1	0.3	0.4	0.7	0.9	–
Green Book	0.1	0.3	0.4	0.6	0.8	1.0

A8. Hybrid vs. Time-Varying VAR (Baseline)

In the body of this chapter and in the previous sections of this appendix we document the good forecasting properties of TVP VAR SV. Our results for TVP VAR SV confirm the findings from a long list of papers documenting the importance of allowing stochastic volatility and time-varying parameters in improving the forecast accuracy of real GDP growth and inflation (e.g. D'Agostino et al, 2013; Clark and Ravazzolo, 2015; for the US, and Barnett et al, 2014 for the UK).

Until recently these models can feasibly be estimated only with three or four variables mainly because of computational constraints. The availability of greater computing power and the introduction of newer methods requiring reduced computational needs are helping increase the use of these models. The latter advantage has allowed the possibility of estimating these models with larger information set (e.g. Koop and Korobilis, 2013). However, these models are complex (as there are so many moving parts involved) and as a result operating them requires a certain level of sophistication which somewhat limits their wider use.

There are situations where using a time-varying VAR may not be feasible. These situations include shorter estimation sample size, an issue for many developing and some emerging countries. Practical considerations such as ease of communicating forecasts (and the narrative) to the principal (or intended audience) along with higher comfort level in operating simpler models would suggest preference for simpler VARs (e.g. many of the economic commentary and associated economic analysis performed by economic consulting firms, banks, etc.). For some, the time varying VARS may be undesirable due to the uncomfortable choices for prior elicitation necessary when using these models especially with a shorter sample.

Practitioners who are unwilling or unable to estimate a time-varying VAR may ask, how well hybrid forecasts obtained using simple VARs do against forecasts from these complex models?

Our results indicate that forecast accuracy of the hybrid forecasts from simple VARs perform comparably or better (for inflation) than the time-varying VARs (baseline forecasts). Of course a forecaster using a time-varying VAR can gain in terms of accuracy by adopting our proposal. However, if that is not an option the forecaster should take some solace that by using a simple VAR (and adopting our proposal) he or she may not be losing much in regards to the forecast accuracy.

To illustrate the competitive forecast accuracy, we compare the forecast accuracy between the hybrid approach using the Small VAR SV and the TVP-VAR SV that uses the three macroeconomic variables (real GDP, CPI inflation and unemployment rate). For a fair horserace, we also require the TVP-VAR SV one-step-ahead forecast to be tilted to match the more accurate

survey nowcast (i.e. the baseline forecast). This will ensure that competing models start from the same jumping-off point.

Table A16 reports the forecast accuracy comparison results between the hybrid forecasts (from Small VAR SV (1960) and Small VAR SV (1985) and baseline forecast from TVP-VAR SV. The top panel (i.e., Panel A) reports the results corresponding to point forecast accuracy and the bottom panel (i.e., Panel B) reports the results corresponding to density forecast accuracy. In both Panels A and B, the top portion reports the accuracy comparison between hybrid forecast from Small VAR SV (1960) and baseline forecast from TVP-VAR SV; the bottom portion reports the accuracy comparison between hybrid forecast from Small VAR SV (1985) and baseline from TVP-VAR SV. In the case of the point forecast comparison, a ratio of less than one suggests that the hybrid point forecast accuracy is on average more accurate compared to the TVP-VAR SV. In the case of the density forecast comparison, a negative number suggests that the hybrid density forecast accuracy is on average more accurate compared to the TVP-VAR SV.

Beginning with the point forecast accuracy, the hybrid forecasts corresponding to real GDP growth and CPI inflation are generally more accurate in an absolute sense, and for inflation, the gains are statistically significant. The ability of the survey expectations to adapt more quickly compared to the TVP-VAR SV partly explains the more accurate inflation forecasts from the hybrid approach. In the case of the unemployment rate, statistically speaking the forecast accuracy is competitive but in an absolute sense hybrid forecasts are inferior. A closer inspection of the errors reveals relatively big misses around the Great Recession and in the early part of the post-crisis period; this explains why on average the inferiority of the hybrid unemployment forecasts is statistically not significant.

For the density forecast accuracy, the hybrid forecasts generally rival the density forecast accuracy of the TVP-VAR SV with an exception that in the case of CPI inflation, the Small VAR SV (1985) is significantly more accurate.

Table A16: Real-Time Forecasting Accuracy Hybrid (from Small VAR) vs. TVP-VAR SV

Panel A: Point Forecast Accuracy (Recursive evaluation: 1994.Q1-2016.Q4)						
	h=1Q	h=4Q	h=6Q	h=8Q	h=10Q	h=12Q
Relative MSE: MSE Hybrid from Small VAR SV (1960) / MSE Baseline from TVP-VAR SV						
Real GDP	1.00	0.99	0.91	0.93	1.02	1.00
CPI Inflation	1.00	0.92***	0.95***	0.95	0.82**	0.85**
Unemployment rate	1.00	1.08	1.10	1.10	1.08	1.10
Relative MSE: MSE Hybrid from Small VAR SV (1985) / MSE Baseline from TVP-VAR SV						
Real GDP	1.00	0.98	0.91	0.91	0.98	0.99
CPI Inflation	1.00	0.90***	0.83***	0.89*	0.77***	0.82***
Unemployment rate	1.00	0.98	1.03	1.06	1.10	1.11
Panel B: Density Forecast Accuracy (Recursive evaluation: 1994.Q1-2016.Q4)						
	h=1Q	h=4Q	h=6Q	h=8Q	h=10Q	h=12Q
Relative CRPS: CRPS Hybrid from Small VAR SV (1960) - CRPS Baseline from TVP-VAR SV						
Real GDP	0.00	-0.02	-0.05	0.02	0.05	0.08*
CPI Inflation	0.00	-0.02	0.04	0.03	0.00	0.02
Unemployment rate	0.00	-0.02	-0.04	-0.05	-0.05	-0.03
Relative CRPS: CRPS Hybrid SV from Small VAR SV (1985) - CRPS Baseline from TVP-VAR SV						
Real GDP	0.00	-0.02	-0.05	-0.02	0.02	0.06
CPI Inflation	0.00	-0.05**	-0.09**	-0.10*	-0.15**	-0.11*
Unemployment rate	0.00	-0.02	-0.03	-0.03	-0.02	-0.01

Notes for Table: The top panel compares the forecast accuracy of the time varying parameter VAR with SV (TVP-VAR SV) to hybrid from Small VAR SV (1960) and Small VAR SV (1985) respectively. The first half of the top panel reports the Mean Square Error (MSE) corresponding to the Small VAR SV (1960) tilted to match the SPF nowcast (mean and variance) and SPF long-horizon forecast (Hybrid) relative to the TVP-VAR SV tilted to match the SPF nowcast (mean and variance). The second half of the top panel report the relative MSE: MSE from the Small VAR with SV (1985) tilted to match the SPF nowcast (mean and variance) and SPF long-horizon forecast (Hybrid SV) / MSE from TVP-VAR SV tilted to match the SPF nowcast (mean and variance) only. So a ratio of less than 1 indicates that tilted VAR hybrid is on average more accurate. The bottom panel reports the corresponding density forecast accuracy performance. A negative value suggests that hybrid density forecast is on average more accurate. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values. All models use the SPF nowcast for the one-step-ahead forecast; as a result, the relative MSE is equal to one and is not reported.

A9. Steady-State BVAR vs. Hybrid (Small BVAR est. 1960)

Villani (2009) proposed a fully Bayesian approach to adjust the long-run forecast of VAR models. His methodology for both stationary and co-integrated Bayesian VARs permits the forecaster to specify prior beliefs on the unconditional mean of the variable, including variance around that prior mean (commonly denoted in the literature as a steady-state BVAR). Using a seven-variable steady-state BVAR model estimated with data for the Swedish economy, he illustrated the improved forecast accuracy compared to an unrestricted BVAR.

Since then the steady-state VAR model has become a widely used tool to incorporate off-model information including the information from the surveys. Wright (2013) uses the steady-state BVAR technology of Villani (2009) to show that prior beliefs on the unconditional mean of the variable informed by a survey's (Blue Chip) long-run forecasts lead to systematic improvements in forecast accuracy for a range of U.S. macroeconomic variables, especially for inflation. Frey and Mokinski (2016) extend the methodology of Wright (2013) by augmenting the vector of dependent variables of a steady-state VAR with their corresponding survey nowcasts and the relationship of the survey nowcasts with all the lagged dependent variables of the model is allowed to deviate from the coefficients capturing the relationship between dependent variables and the lagged dependent variables. The extent of the deviation is operationalized through the prior and is pinned down through the joint estimation of the data and the survey nowcasts. They show that doing so leads to more accurate forecasts compared to Wright (2013), as their approach acts as a form of Bayesian shrinkage technique that helps sharpen the parameter estimates of the unaugmented (i.e., original) steady-state VAR.

Chan and Koop (2014) extend the steady-state BVAR technology by developing a methodology that allows for detection of changes in the steady states of included variables in an automatic fashion. However, their econometric approach relies on past data and therefore, it is likely to be slower to detect any shifts in variables' mean compared to professional forecasters (e.g., knowledge of forward guidance and extended zero lower bound). In their empirical exercise, they consider five variables (output growth, hours worked, labor share, inflation, and interest rates), and their methodology detects changes in the steady states for inflation and the interest rate when estimated over the sample 1954.Q3 through 2012.Q3. This finding coincides with our results on tilting VAR forecasts to long-term survey conditions. Specifically, our results indicate a notable improvement in forecast accuracy for inflation and the interest rate, largely because these two macroeconomic variables have exhibited sizable shifts in the mean since the 1950s.

Our approach has practical advantages over the steady-state BVAR: (1) our approach does not require the modeler to re-specify and re-parameterize the model, and (2) our approach does not require the VAR to be stationary or co-integrated. Therefore, modelers can continue to use

their preferred BVAR/VAR specification and seamlessly integrate the relative entropy approach by simply calling a particular routine that takes as an input the output from the BVAR and the modeler's desired short-term and long-term forecasts informed by judgment, other models, or surveys. A related advantage is that the forecaster can get a sense of the differences between the forecast that imposes conditions compared to the one that does not without the need to re-specify the model. The forecaster can test the implications on the forecasts of a wide spectrum of restrictions without re-estimating the model each time. The technique of relative entropy generates several diagnostics that help the modeler assess the severity of tilting, that is, reflecting the presumed reliability of inferences drawn from the tilted distributions – and the role of the various moment conditions contributing to the severity.

We next show the forecast accuracy comparison between the steady-state VAR and our approach based on a Small VAR estimated with longer sample (and for a specification without stochastic volatility). To perform a fair horse-race both these two approaches uses the same data and survey information. A priori we would expect the two approaches to be competitive with each other. Indeed the results reported below (Table A17) confirm our expectations.

The steady-state BVAR model includes the same set of variables as used in our Small BVAR. The steady states are informed by the long-term forecasts of the Survey of Professional Forecasters (i.e., same values to which the Small BVAR is tilted). To run a fair horserace, we tilt both the steady-state BVAR and Small BVAR in the near term on the same nowcasts to ensure that both models start with the same jumping-off point. For this exercise, we do not tilt towards the nowcast variance, only the nowcast mean.

The estimation procedure and the prior settings are the same as in Clark (2011) with the exception that prior variances around the steady-state priors are set very tight (as proposed in Wright, 2013). This will ensure that variables converge to the modeler's specified steady-states which in this case are the long-term forecasts from the SPF. (The prior variances are set at a value of 0.001).

Table A17 reports the forecast accuracy comparison between the steady-state BVAR and the hybrid approach from the Small BVAR. Panel A reports the point forecast accuracy comparison using relative MSE: $\text{MSE Small BVAR} / \text{MSE steady-state BVAR}$. A ratio of less than one suggests that the hybrid forecast from the Small BVAR is on average more accurate compared to the forecast from the steady-state BVAR. Panel B reports the density forecast accuracy comparison in the form of the mean relative CRPS, where negative numbers indicate that density forecasts from the hybrid approach (of Small BVAR) are on average more accurate compared to the steady-state BVAR.

Table A17: Real-Time Out-of-Sample Forecasting Performance: **Small BVAR (Hybrid)**
vs. **Steady-State BVAR**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
Panel A: Point Forecast Accuracy of Small BVAR (Hybrid) vs. Steady-State BVAR (Baseline)								
Series	h=1Q	h=4Q	h=6Q	h=8Q	h=10Q	h=12Q	h=20Q	h=32Q
Relative MSE: Small BVAR (Hybrid) / Steady-State BVAR (Baseline)								
Real GDP	1.00	0.90	0.84*	0.87*	0.88	0.95*	0.98	1.00
CPI Inflation	1.00	1.02	1.00	1.04	0.94	1.03	0.90*	1.00
Unemployment	1.00	0.98	0.96	0.95	0.96	1.00	1.07	0.93
Federal Funds	1.00	1.15*	1.06	0.98	0.92	0.88	0.87	1.02
Credit Spread	1.04	1.07	1.05	1.03	1.03	1.07	1.06***	1.14**
Panel B: Density Forecast Accuracy of Small BVAR (Hybrid) vs. Steady-State BVAR (Baseline)								
	h=1Q	h=4Q	h=5Q	h=8Q	h=10Q	h=12Q	h=20Q	h=32Q
Mean (Relative CRPS Score: Small BVAR (Hybrid) - Steady-State BVAR (Baseline))								
Real GDP	-0.02	-0.07	-0.11	-0.08*	-0.07	-0.04**	-0.02*	-0.01
CPI Inflation	-0.01	0.00	0.02	0.02	0.00	0.03**	-0.02	0.04
Unemployment	0.00	-0.01	-0.02	-0.04	-0.04	-0.02	0.04	-0.04
Federal Funds	-0.00	0.04*	0.03	0.02	-0.01	-0.05	-0.03	0.25**
Credit Spread	0.00	0.02	0.03	0.03	0.03	0.03	0.04**	0.06*

Notes for Table: The numbers reported in the top panel of the table are relative mean squared errors: mean squared error conditional on nowcasts and long-horizon survey forecasts from Small BVAR (Hybrid) / mean squared error from the steady-state BVAR conditional on nowcasts only (Baseline). So a ratio of less than 1 indicates that point forecasts from the hybrid approach corresponding to fixed-coefficient Small VAR are on average more accurate compared to the forecasts from the steady-state BVAR. The numbers reported in the bottom panel of the table are mean relative CRPS: CRPS from the Hybrid - CRPS from steady-state BVAR conditional on nowcasts only. So a negative number indicates that density forecasts from the hybrid approach corresponding to fixed-coefficient Small VAR are on average more accurate compared to the forecasts from the steady-state BVAR. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

A10. BVAR in Gaps vs. Hybrid (Small BVAR est. 1960)

Another popular approach to anchor model forecasts to survey expectations is to model variables by first transforming them into a gap form (i.e., deviation from the respective long-run survey expectations) and then estimating them using a VAR (or a univariate regression for the single variable of interest). The forecasts of the gap coming out of the VAR are then transformed back to the units of interest by adding the latest estimate of the survey expectations available as of the forecast origin to construct the corresponding implied forecasts (see Faust and Wright 2013 in the context of the univariate inflation case; Clark and McCracken 2010 in the case of the VAR). The trend estimate (proxied by the survey expectations measure) is assumed to follow a random walk over the forecast horizon. By construction, the implied long-run forecasts from this approach would be close to the latest available estimate of the survey expectations plugged in as of the time forecast is generated. The advantage of this approach is its simplicity, and therefore, it has gained traction over the past few years. A key drawback is that it requires a time series of survey expectations as long as the estimation sample (necessary for constructing the transformed gap variable). This issue may be more likely to bind for regions outside the United States and Europe for which publicly available survey forecasts have a shorter history.

We construct a BVAR model in gaps that includes the same set of variables as used in a Small VAR estimated with longer sample (and for a specification without stochastic volatility). The variables (with the exception of the credit spread) are transformed to the gap by taking a deviation from their respective time series of the survey expectations. To run a fair horserace, we condition both the BVAR in gaps and Small BVAR in the near-term on the same nowcasts. This will ensure that both models start with the same jumping-off points. For this exercise, we do not tilt towards the nowcast variance, only the nowcast mean.

We construct the expectation series for real GDP, CPI inflation, the unemployment rate, and the federal funds rate going back to 1959.Q4 as follows:

From **1959.Q4 to 1993.Q4**:

1. Real GDP growth trend=constant 3%
2. Unemployment rate trend is computed using an exponential smoother with a smoothing parameter of 0.02 (as in Clark, 2011):

$$u_t^* = u_{t-1}^* + 0.02(u_t - u_{t-1}^*)$$

3. CPI inflation trend is the PTR (long-term inflation expectations series used in the Federal Reserve Board's FRB/US econometric model)

4. Nominal federal funds rate trend is assumed to be the same as the CPI inflation trend

From **1994.Q1 to 2016.Q4**:

The respective trend estimates are the long-run forecasts from the SPF.

The prior settings are the same as those of the Small BVAR. Table A6 reports the forecast accuracy comparison between the Small BVAR in gaps and the hybrid approach from the Small BVAR. Panel A reports the point forecast accuracy comparison using relative MSE: $\text{MSE Small BVAR} / \text{MSE Small BVAR in Gaps}$. A ratio of less than one suggests that the hybrid forecast from the Small BVAR is on average more accurate compared to the forecast from the BVAR in Gaps. Panel B reports the density forecast accuracy comparison in the form of mean relative CRPS where negative numbers indicate density forecasts from the hybrid approach (of Small BVAR) are on average more accurate compared to the BVAR in Gaps. A priori we would expect the two econometric approaches to perform comparably. Based on the results reported in Table A18, for real GDP growth, CPI Inflation, and the federal funds rate, the hybrid approach generates more accurate forecasts and the gains are statistically significant.

It is worth noting that even though an attempt is made to anchor the forecasts closer to the survey expectations (through modeling the variables in gap transformation), there is no guarantee that the medium- to long-term forecast would converge to the survey expectations. It may even settle far from the assumed underlying trend. This is due to the presence of an intercept term in the gap equation (e.g., inflation gap) that captures the long-run historical deviation of the gap from zero within the estimation sample. The estimate of the intercept term will be positive if the variable (e.g., inflation) has exceeded its trend (informed by the survey) on average during the sample, while it will be negative if the variable has been below trend on average. So an inflation forecast three years out may settle at a level that is lower than the trend estimate informed by the survey expectations (and the modeler's desired level).

Table A18: Real-Time Out-of-Sample Forecasting Performance: **Small BVAR (Hybrid)** vs. **Small BVAR in Gaps (Baseline)**

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)							
Panel A: Point Forecast Accuracy of Small BVAR (Now and LR) vs. BVAR in Gaps (Now Only)							
Series	h=1Q	h=4Q	h=6Q	h=8Q	h=10Q	h=12Q	h=20Q
Relative MSE: Small BVAR (Hybrid) Forecast / BVAR in Gaps (Baseline) Forecast							
Real GDP	1.00	0.91	0.89*	0.87***	0.92**	0.94***	0.97*
CPI Inflation	1.00	0.94***	0.87***	0.82***	0.78***	0.83**	0.80***
Unemployment rate	1.00	0.95	0.99	1.00	1.01	1.03	1.13
Federal Funds rate	1.00	0.95	0.89	0.80*	0.73**	0.69***	0.67***
Credit Spread	0.91	0.97	0.96	0.95	0.94	0.98	1.08

Panel B: Density Forecast Accuracy of BVAR (Now and LR) vs. BVAR in Gaps (Now Only)							
	h=1Q	h=4Q	h=5Q	h=8Q	h=10Q	h=12Q	h=20Q
Mean (Relative CRPS Score: Small BVAR (Hybrid) Forecast - BVAR in Gaps (Baseline) Forecast)							
Real GDP	0.01	-0.07	-0.07*	-0.08***	-0.05**	-0.03	-0.01
CPI Inflation	0.00	-0.02	-0.03	-0.06*	-0.06*	0.00	0.04
Unemployment rate	0.00	-0.01	-0.01	-0.01	0.00	0.00	0.04
Federal Funds rate	0.00	0.00	-0.03	-0.09	-0.17	-0.26*	-0.40***
Credit Spread	-0.01	0.00	-0.01	-0.01	-0.01	0.00	0.04

Notes for Table: The numbers reported in the top panel of the table are relative mean squared errors: mean squared error conditional on nowcasts and long-horizon survey forecasts from Small BVAR (Hybrid) / mean squared error from the BVAR in Gaps conditional on nowcasts only. So a ratio of less than 1 indicates that point forecasts from the hybrid approach corresponding to fixed-coefficient Small VAR are on average more accurate compared to the forecasts from the BVAR in gaps. The numbers reported in the bottom panel of the table are mean relative CRPS: CRPS from the Hybrid - CRPS from BVAR in Gaps conditional on nowcasts only. So a negative number indicates that density forecasts from the hybrid approach corresponding to fixed-coefficient Small VAR are on average more accurate compared to the forecasts from the BVAR in gaps. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

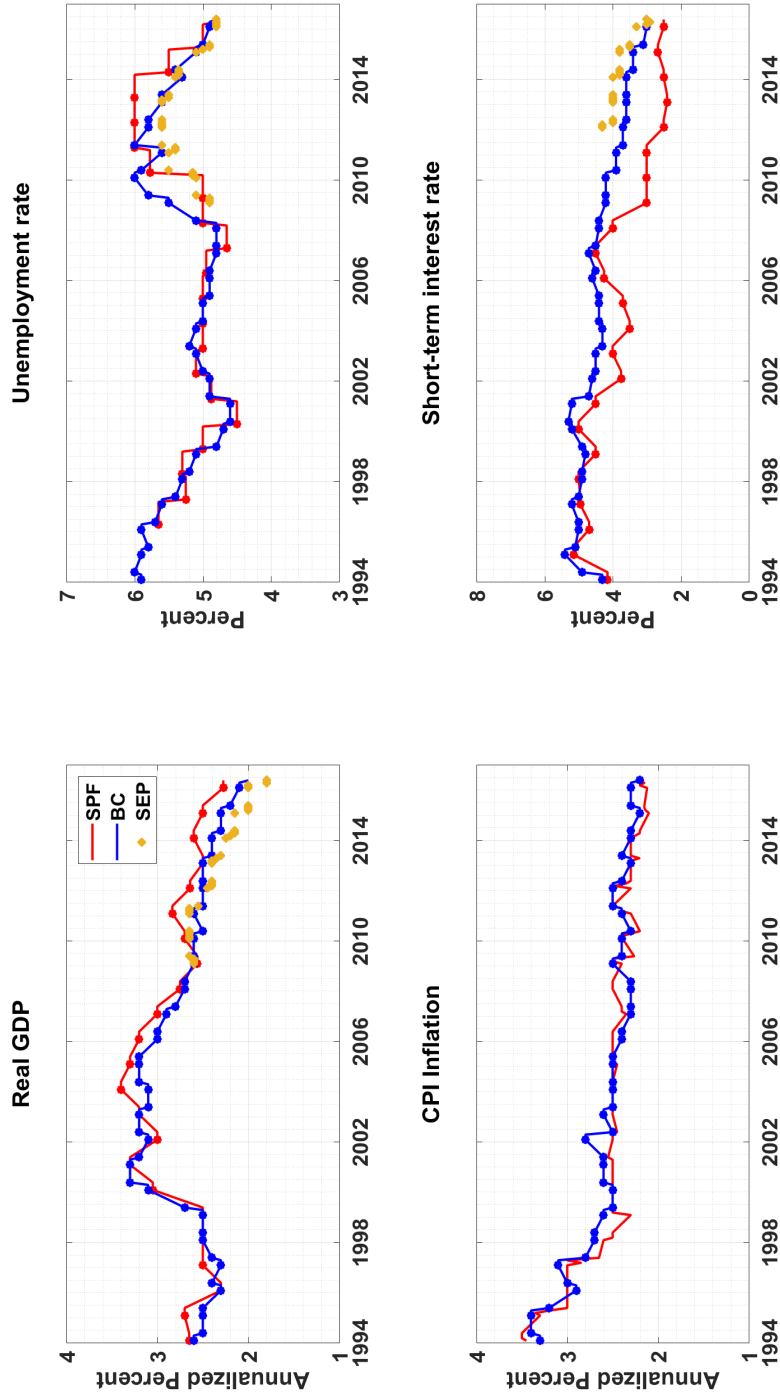
A11. Evolution of Long-Run Forecasts from Other Surveys

In the United States, the SPF and the Blue Chip Economic Indicators (BC) are the two most widely known and easily available forecast surveys routinely published. All the forecast evaluation exercises in this chapter uses SPF, and Figure 1 in the chapter plots the evolution of the long-run forecasts from the SPF. In this section, Figure A2 plots the evolution of Blue Chip (consensus estimates) alongside the SPF. Also plotted to facilitate comparison are the median projections from the Federal Open Market Committee's (FOMC) Summary of Economic Projections (SEP).

The BC is a monthly survey released at the end of the first week of the month. It is a monthly survey therefore the forecasts and nowcasts collected reflect the developments in the intra-quarterly flow of information. Two times a year, i.e., each March and October, the BC also reports the respondents' long run projections (defined as an average for the 7 to 11 years ahead). The BC reports long-run forecasts for all of our core variables of interest: CPI inflation, real GDP growth, the unemployment rate, and the short-term interest rate.

SEP projections are made available roughly once per quarter beginning 2009.Q1. SEP reports projections for real GDP growth, unemployment rate, headline inflation based on Personal Consumption Expenditures (PCE) price index, and core PCE inflation. It also provides the long-run estimate for the federal funds rate. The projections for federal funds rate were added beginning in 2012.Q1.

Figure A2: Real-Time Long Run Forecasts



Notes for the figure: plots the real-time long-run forecasts from the Survey of Professional Forecasters (SPF), Blue Chip (BC), and FOMC's Summary of Economic Projections (SEP). In the case of SPF, for inflation, real GDP growth, and short-term interest rate it is the 10-year-ahead average forecast and for the unemployment rate it is the natural rate estimate; for BC it is the average for 7 to 11 years ahead for all variables. The short-term interest rate in the case of both SPF and BC refers to the 3-month T-bill rate, but in the case of SEP, it refers to the federal funds rate.

A12. Gaussian example: Illustrating Spillover Effects of Tilting

Restricting the elements of the forecast matrix by imposing conditions on some future horizon will influence the forecast starting from the jumping-off point all the way to the tilted forecast horizon. For example, if we tilt real GDP growth at forecast horizon $h=6Q$, then tilting it will potentially impact the forecast trajectory from forecast horizons $h=1Q$ to $h=5Q$ and from $h=7Q$ and beyond for all the variables. The extent and degree of the spillover effects will be determined importantly by the BVAR's implied estimates of the covariances and autocorrelations among the variables and across forecast horizons.

To provide an intuition of the mechanics behind the spillover effects below, we illustrate using an example of a multivariate normal density (as would be obtained from a constant coefficient VAR model). Our example below generalizes the examples provided in Robertson, Tallman, and Whiteman (2005) and KCR. For convenience, we keep the same notation where possible.

We begin with a multivariate normal density $f(Y) = N(\theta, \Sigma)$ corresponding to the H -variate vector Y of forecast, $Y = [y_1, y_2, \dots, y_H]$; Σ is positive definite and $\theta = [\theta_1, \dots, \theta_H]'$.

We obtain a KLIC-closest density $f(Y)^* = N(\mu, \Omega)$ such that it satisfies the restriction that the mean and the variance of the first element of vector Y , y^1 equals μ_1 and $\Omega_{1,1}$, respectively (e.g., a nowcast informed by the survey expectations).

The parameters of the tilted density f^* are defined as follows,

$$\mu_{2:H} = \theta_{2:H} + \Sigma_{1,1}^{-1} \Sigma_{1,2:H} (\mu_1 - \theta_1) \quad (\text{A.20})$$

$$\Omega_{2:H,2:H} = \Sigma_{2:H,2:H} - \Sigma_{2:H,1} \Sigma_{1,1}^{-1} \Sigma_{1,2:H} \times \left(\frac{\Sigma_{1,1} - \Omega_{1,1}}{\Sigma_{1,1}} \right) \quad (\text{A.21})$$

$$\Omega_{2:H,1} = \Sigma_{2:H,1} \Sigma_{1,1}^{-1} \Omega_{1,1} \quad (\text{A.22})$$

The matrices indexed by $i : j, a : b$ represents a matrix containing rows from i to j , and columns a to b . Accordingly, matrices indexed by $i : j, a$ correspond to column vector and those indexed by $i, a : b$ correspond to row vector. The elements of column vector $\Sigma_{2:H,1}$ reflect the correlation between the nowcast horizon and the forecast horizons beyond the nowcast.

From the above definitions of the parameters, it can be easily seen that imposing the moment restriction $\Omega_{1,1} = 0$ is equivalent to the standard conditional forecasting. Also, by imposing the

mean moment condition only, it can be seen that the tilted variance is the same as variance of the untilted density (i.e. the original variance).

A13. Sampling from Tilted Predictive Density: Multinomial Algorithm

To sample from the modified (i.e., tilted) predictive density $g(\cdot)$, we follow the approach suggested in Cogley et al. (2005). Specifically, they suggest using the multinomial resampling algorithm of Gordon et al. (1993) to redraw from the original predictive density $p(\cdot)$ using the modified weights, ω^* , to obtain a sample corresponding to the tilted density $g(\cdot)$.

Algorithm

Given a sample $Y_i^{T+1, T+H}$, $i = 1, \dots, D$ from the predictive density $p(\cdot)$ along with the weights, ω_i^* corresponding to the tilted density $g(\cdot)$ the steps listed below are used to obtain a sample from $g(\cdot)$

Step 1: Define NC as a $D \times 1$ vector representing the number of offspring corresponding to each draw obtained from the original density $p(\cdot)$

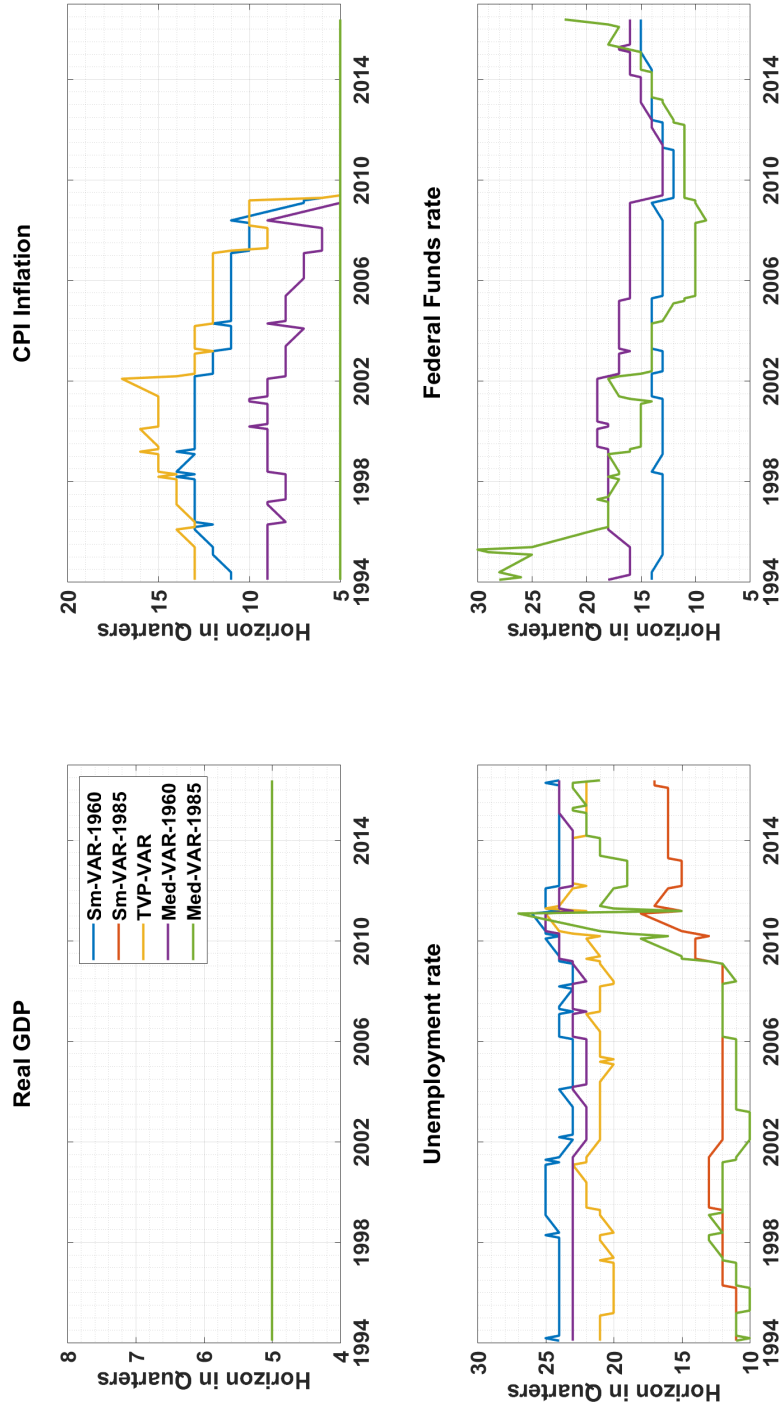
Step 2: Define a value for D^* such that $D^* > D$. D^* represents the number of draws for the tilted predictive density $g(\cdot)$, our object of interest.
The above two steps ensure that $\sum_{i=1}^D NC_i = D^*$.

Step 3: Draw D number of draws for NC from a multinomial distribution $NC \sim MN(D^*; w_1^*, w_2^*, \dots, w_D^*)$. (Matlab function `mnrnd` is used to draw from the MN distribution.)

Step 4: Given a sample for NC obtained in the previous step, construct a density $g(\cdot)$ by replicating $Y_i^{T+1, T+H}$ NC_i times for $i = 1, \dots, D$

A14. Evolution of Forecast Horizons for Tilting

Figure A3: Forecast Horizons at which model combined with long-run survey



Notes for the figure: plots correspond to the models, Small VAR estimated with longer sample (1960), Small VAR estimated over shorter sample (1985), Time-varying parameters VAR (TVP-VAR), Medium VAR (1960), and Medium VAR (1985). The x-axis values represent forecast origins spanning 1994.Q1 through 2016.Q4. The horizons plotted are determined based on the approach laid out in the section 4.2 ('Determining the Forecast Horizon for Tilting')

A15. Sensitivity to Getting the Horizon Wrong

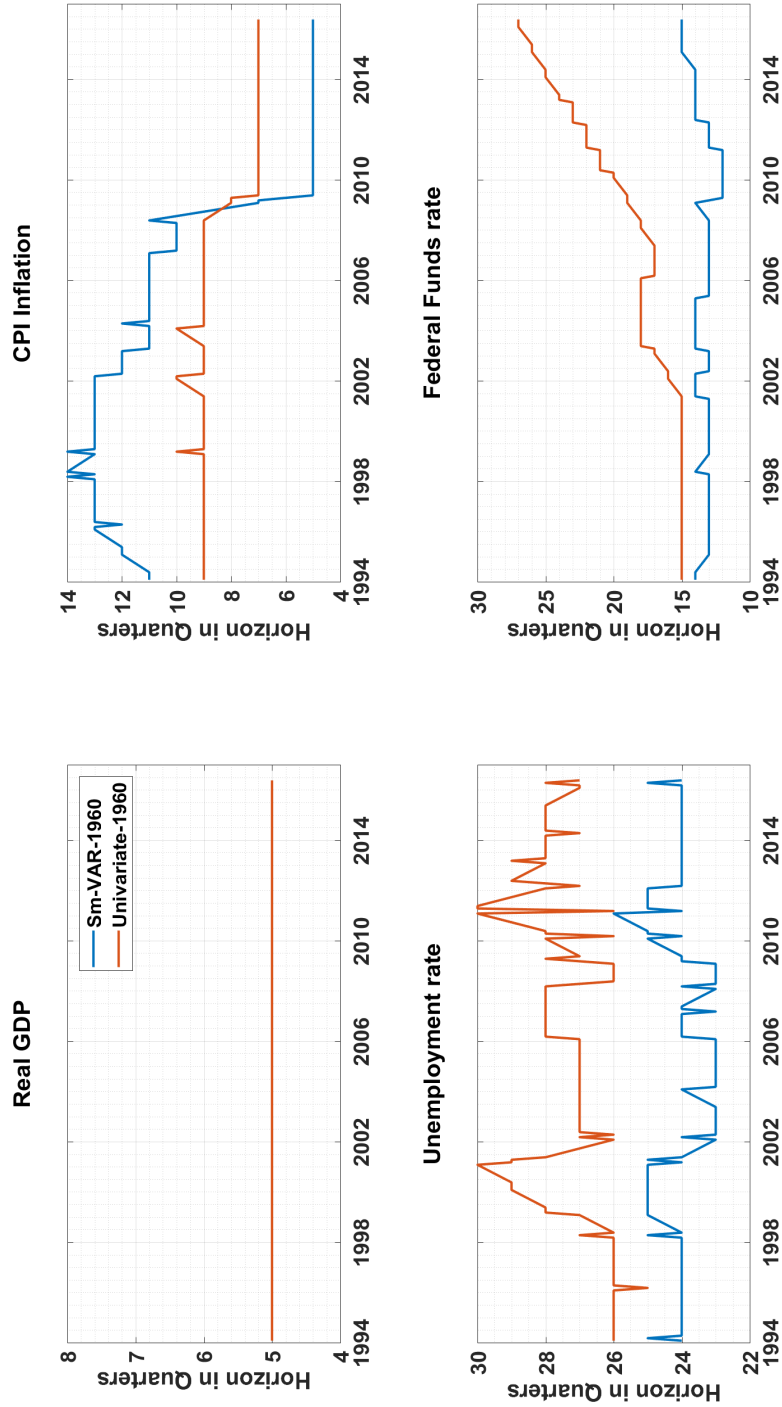
Thus far, all reported results corresponding to hybrid forecasts are constructed by combining the model and long-term survey expectations at the forecast horizon based on the algorithm detailed in section 2.4.2 (of chapter 2).

In this section, we explore the sensitivity of our results to two alternative approaches. The first approach is based on an alternative algorithm to compute an estimate for the persistence parameter. The second approach ignores the persistence (or dynamic behavior of variables) and dogmatically set the combination horizon 7 years out ($h=25Q$); we also show results for dogmatically setting the horizon 10 years out ($h=40Q$).

The alternative algorithm to estimate the persistence is based on estimating a univariate AR(4) process to determine the persistence estimate and in turn the forecast horizon at which the variable would be expected to converge. Specifically, at each forecast origin, the forecaster decides to estimate separately univariate autoregressions for each variable of interest to get an indication of the persistence of each variable. And then use these estimates of the persistence (i.e. proxy for the speed of convergence) to compute the respective values for the combination forecast horizon. In contrast to the approach discussed in section 2.4.2, this approach fully ignores the possible effects on the persistence from other variables included in the VAR. Figure A4 plots the values of the forecast horizon for combination obtained from the univariate approach alongside the values for Small VAR (1960) obtained from the approach used in the chapter. There are some notable differences except for real GDP growth, for which both approaches give values less than or equal to $h=5$. However, we find that results reported in Tables A19 (for Small VAR 1960) and A20 (for Small VAR 1985) are robust to obtaining forecast horizon for combination from this alternative approach (see entries in the rows Hybrid/Baseline).

In the second approach, the forecast horizon for combination is dogmatically set at $h=25Q$ (i.e. 7 years out). This approach ignores the dynamic behavior of the variables and so will be expected to perform less well compared to approaches that in some way account for the persistence of the variable. The results of this exercise are reported in Table A21, and it confirms our prior beliefs. As can be seen looking at the entries in the rows labeled Hybrid/Baseline, the gains in accuracy are notably reduced for the forecast horizons of interest (i.e. 1 to 3 years out). It is still the case that statistical significant gains are achieved for CPI inflation and the federal funds rate but the magnitude of accuracy gains are meaningfully reduced. For the VAR specification with SV, statistical significant gains are only seen for the federal funds rate. The results for other VAR models (not shown) largely echo the results reported for the VAR specification with SV. Table A22 reports the results if we set the forecast horizon for combination even further out (i.e. 10 years out; 40 quarters). As expected, the accuracy gains for the forecast horizons of interest are reduced further. For the specification with SV, the relative ratios for Hybrid/Baseline are close to one, and none are statistically significant.

Figure A4: Forecast Horizons at which model combined with long-run survey



Notes for the figure: plots correspond to the models, Small VAR estimated with longer sample (1960), and Univariate AR for respective variables estimated with longer sample (1960). The x-axis values represent forecast origins spanning 1994.Q1 through 2016.Q4. The horizon values corresponding to Small VAR are determined based on the approach laid out in the section 4.2 ('Determining the Forecast Horizon for Tilting')

Table A19: Out-of-Sample **Point** Forecasting Performance: **Small BVAR est. 1960**
Horizon for combination determined by Univariate process

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)

	Small VAR				Small VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	4.35	7.19	6.55	6.20	3.67	6.68	6.61	6.26
Relative MSE								
Baseline/Raw	0.62***	1.01	1.05**	1.02	0.74**	1.00	1.01	1.00
Hybrid/Raw	0.62***	0.91	0.91	0.99	0.74**	0.89**	0.86*	0.98
Hybrid/Baseline	1.00	0.90**	0.86*	0.98*	1.00	0.88**	0.85**	0.98
CPI								
Raw	3.11	5.61	6.70	8.52	2.94	5.26	5.50	6.50
Relative MSE								
Baseline/Raw	0.33***	0.90**	0.95	0.92*	0.35***	0.86***	0.99	0.94***
Hybrid/Raw	0.33***	0.81***	0.70***	0.61***	0.35***	0.82***	0.87*	0.78***
Hybrid/Baseline	1.00	0.90**	0.74***	0.66***	1.00	0.96**	0.87**	0.82***
UR								
Raw	0.05	0.64	2.41	3.78	0.05	0.60	2.39	3.75
Relative MSE								
Baseline/Raw	0.29***	0.73*	0.92	0.95	0.32***	0.73**	0.90*	0.96
Hybrid/Raw	0.29***	0.74*	0.89	0.96	0.32***	0.72**	0.84	0.95
Hybrid/Baseline	1.00	1.01	0.97	1.00	1.00	0.99	0.94	0.98
FFR								
Raw	0.19	2.11	6.12	10.16	0.10	2.12	7.17	11.66
Relative MSE								
Baseline/Raw	0.03***	0.68***	0.87**	0.93***	0.05***	0.63***	0.85***	0.92**
Hybrid/Raw	0.03***	0.71**	0.72*	0.63**	0.05***	0.65***	0.68***	0.61***
Hybrid/Baseline	1.00	1.04	0.82*	0.68**	1.00	1.03*	0.80**	0.66***
Credit Spread								
Raw	0.09	0.67	1.14	1.28	0.08	0.73	1.30	1.53
Relative MSE								
Baseline/Raw	0.77*	0.95***	0.97***	0.99	0.89**	0.99	0.98**	0.99*
Hybrid/Raw	0.78	0.90**	0.84***	0.85***	0.87**	0.92***	0.79***	0.74***
Hybrid/Baseline	1.01	0.95**	0.86***	0.86**	0.98	0.93***	0.81***	0.75***

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; Credit Spread: in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean squared error (MSE), the three rows immediately below report relative MSE: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table A20: Out-of-Sample **Point** Forecasting Performance: **Small BVAR est. 1985**
Horizon for combination determined by Univariate process

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)								
	Small VAR				Small VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	3.43	6.23	6.65	6.46	3.42	6.07	6.51	6.62
Relative MSE								
Baseline/Raw	0.79**	0.93	1.00	1.03***	0.79**	0.96	1.01	1.00
Hybrid/Raw	0.79**	0.88*	0.85*	0.95	0.79**	0.93	0.85***	0.92**
Hybrid/Baseline	1.00	0.94	0.86**	0.92*	1.00	0.97	0.84***	0.92*
CPI								
Raw	2.95	4.71	5.11	6.12	2.77	4.79	4.67	5.50
Relative MSE								
Baseline/Raw	0.35***	1.00	1.00	1.00	0.37***	0.95***	1.05	1.02
Hybrid/Raw	0.35***	0.96	0.89***	0.85***	0.37***	0.91***	0.97	0.89***
Hybrid/Baseline	1.00	0.96*	0.89***	0.85**	1.00	0.95	0.93**	0.88***
UR								
Raw	0.04	0.48	2.22	3.94	0.04	0.46	2.09	3.84
Relative MSE								
Baseline/Raw	0.35***	0.75*	0.84	0.93	0.35***	0.76**	0.86	0.92*
Hybrid/Raw	0.35***	0.76*	0.82	0.89	0.35***	0.77**	0.85	0.88*
Hybrid/Baseline	1.00	1.01	0.98	0.95*	1.00	1.02	0.99	0.96
FFR								
Raw	0.08	1.95	6.79	10.58	0.08	1.95	6.85	11.33
Relative MSE								
Baseline/Raw	0.06***	0.57***	0.81**	0.94	0.06***	0.57***	0.81***	0.93*
Hybrid/Raw	0.06***	0.52***	0.59**	0.63	0.06***	0.53***	0.63***	0.66**
Hybrid/Baseline	1.00	0.92*	0.73**	0.68*	1.00	0.93***	0.78**	0.71**
Credit Spread								
Raw	0.09	0.63	1.06	1.12	0.09	0.66	1.09	1.19
Relative MSE								
Baseline/Raw	0.72**	0.91**	1.00	1.07	0.84**	0.91**	0.99	1.00
Hybrid/Raw	0.70**	0.89*	0.88	0.93	0.83**	0.89**	0.91	0.89
Hybrid/Baseline	0.97	0.97	0.88	0.87	0.99	0.97	0.91	0.88

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; Credit Spread: in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean squared error (MSE), the three rows immediately below report relative MSE: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table A21: Out-of-Sample **Point** Forecasting Performance: **Small BVAR est. 1960**
Horizon for combination dogmatically set at h=25Q (i.e. 7 years out)

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)

	Small VAR				Small VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	4.35	7.19	6.55	6.20	3.67	6.68	6.61	6.26
Relative MSE								
Baseline/Raw	0.62***	1.01	1.05**	1.02	0.74**	1.00	1.01	1.00
Hybrid/Raw	0.62***	1.00	1.05*	1.09**	0.74**	0.96*	1.04*	1.05
Hybrid/Baseline	1.00	0.99	1.00	1.07	1.00	0.95*	1.03	1.05
CPI								
Raw	3.11	5.61	6.70	8.52	2.94	5.26	5.50	6.50
Relative MSE								
Baseline/Raw	0.33***	0.90**	0.95	0.92*	0.35***	0.86***	0.99	0.94***
Hybrid/Raw	0.33***	0.85***	0.77**	0.69**	0.35***	0.88***	0.95	0.88***
Hybrid/Baseline	1.00	0.94	0.81**	0.75**	1.00	1.03	0.96	0.93
UR								
Raw	0.05	0.64	2.41	3.78	0.05	0.60	2.39	3.75
Relative MSE								
Baseline/Raw	0.29***	0.73*	0.92	0.95	0.32***	0.73**	0.90*	0.96
Hybrid/Raw	0.29***	0.71*	0.92*	1.00	0.32***	0.73**	0.89*	0.98
Hybrid/Baseline	1.00	0.97*	1.00	1.05	1.00	1.00	1.00	1.02
FFR								
Raw	0.19	2.11	6.12	10.16	0.10	2.12	7.17	11.66
Relative MSE								
Baseline/Raw	0.03***	0.68***	0.87**	0.93***	0.05***	0.63***	0.85***	0.92**
Hybrid/Raw	0.03***	0.65**	0.77*	0.76	0.05***	0.62***	0.77***	0.84**
Hybrid/Baseline	1.00	0.95**	0.89	0.82	1.00	0.99	0.91*	0.91**
Credit Spread								
Raw	0.09	0.67	1.14	1.28	0.08	0.73	1.30	1.53
Relative MSE								
Baseline/Raw	0.77*	0.95***	0.97***	0.99	0.89**	0.99	0.98**	0.99*
Hybrid/Raw	0.80	0.93**	0.95**	1.01	0.89**	0.94***	1.02	0.99
Hybrid/Baseline	1.03*	0.98	0.98	1.02	0.99	0.95*	1.03	1.00

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; Credit Spread: in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean squared error (MSE), the three rows immediately below report relative MSE: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Table A22: Out-of-Sample **Point** Forecasting Performance: **Small BVAR est. 1960**
Horizon for combination dogmatically set at h=40Q (i.e. 10 years out)

Full Sample (Recursive evaluation: 1994.Q1-2016.Q4)

	Small VAR				Small VAR with SV			
	h=1Q	h=4Q	h=8Q	h=12Q	h=1Q	h=4Q	h=8Q	h=12Q
GDP								
Raw	4.35	7.19	6.55	6.20	3.67	6.68	6.61	6.26
Relative MSE								
Baseline/Raw	0.62***	1.01	1.05**	1.02	0.74**	1.00	1.01	1.00
Hybrid/Raw	0.62***	0.99	1.08**	1.07**	0.74**	1.03	1.01	1.02
Hybrid/Baseline	1.00	0.98	1.02	1.05	1.00	1.02	1.01	1.03
CPI								
Raw	3.11	5.61	6.70	8.52	2.94	5.26	5.50	6.50
Relative MSE								
Baseline/Raw	0.33***	0.90**	0.95	0.92*	0.35***	0.86***	0.99	0.94***
Hybrid/Raw	0.33***	0.86***	0.79**	0.73**	0.35***	0.88***	0.94**	0.97
Hybrid/Baseline	1.00	0.95**	0.84***	0.80**	1.00	1.03	0.94	1.03
UR								
Raw	0.05	0.64	2.41	3.78	0.05	0.60	2.39	3.75
Relative MSE								
Baseline/Raw	0.29***	0.73*	0.92	0.95	0.32***	0.73**	0.90*	0.96
Hybrid/Raw	0.29***	0.72*	0.93	0.99	0.32***	0.74**	0.91*	0.97*
Hybrid/Baseline	1.00	0.99	1.01	1.03	1.00	1.01	1.01	1.00
FFR								
Raw	0.19	2.11	6.12	10.16	0.10	2.12	7.17	11.66
Relative MSE								
Baseline/Raw	0.03***	0.68***	0.87**	0.93***	0.05***	0.63***	0.85***	0.92**
Hybrid/Raw	0.03***	0.65***	0.74**	0.75*	0.05***	0.64***	0.84***	0.94**
Hybrid/Baseline	1.00	0.95	0.85*	0.80*	1.00	1.01	0.99	1.02
Credit Spread								
Raw	0.09	0.67	1.14	1.28	0.08	0.73	1.30	1.53
Relative MSE								
Baseline/Raw	0.77*	0.95***	0.97***	0.99	0.89**	0.99	0.98**	0.99*
Hybrid/Raw	0.82*	0.92***	0.96***	0.99	0.89**	0.97**	1.01	1.03
Hybrid/Baseline	1.06	0.97**	0.99	1.01	1.00	0.98	1.02	1.05

Notes for Table: GDP: real GDP growth quarterly annualized rate; CPI: inflation quarterly annualized rate; UR: unemployment rate in levels; FFR: effective federal funds rate in levels; Credit Spread: in levels. Raw forecast is defined as the unconditional forecast from the VAR. Baseline forecast is defined as the raw VAR forecast tilted towards survey nowcasts only (both mean and variance). Hybrid forecast is defined as the raw VAR forecast tilted towards both survey nowcasts (both mean and variance) and long-horizon forecasts. The left panel reports results for the VAR specification with constant variance and right panel reports results for the VAR specification with stochastic volatility. The numbers reported in the row labeled Raw are the mean squared error (MSE), the three rows immediately below report relative MSE: Baseline relative to Raw, Hybrid relative to Raw, and Hybrid relative to Baseline. The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *10 percent, **5 percent, and ***1 percent significance levels, respectively. The test statistics use two-sided standard normal critical values.

Appendix B

Chapter 3 Appendix

Chapter 3: Real-Time Density Nowcasts of US Inflation: A Model-Combination Approach.

B1. Description of Mixed-Frequency Models and Simulation Procedures

B.1.1. MIDAS Model

Following Knotek and Zaman (2017, KZ), a general representation of an ADL-MIDAS model with leads takes the following form,

$$\pi_{t+h} = \alpha_h + \sum_{j=0}^{P(M)-1} \chi_{j+1,(h)} \pi_{t-j} + \sum_{j=0}^{P(M)-1} \gamma_{j+1,(h)} Z_{t-j} + \beta_h \sum_{j=0}^{P(HF)-1} \omega_{P(HF)-j}(\theta_{(h)}^{HF}) X_{P(HF)-j,t+1}^{HF} + e_{t+h} \quad (\text{B.1})$$

where Z refers to other monthly variables; $P(M)$ refers to the number of lags of the monthly regressors (we set to 1); and $P(HF)$ refers to the number of high-frequency observations, $X_{1,t+1}^{HF}, \dots, X_{P(HF),t+1}^{HF}$ in month $t+1$ (i.e., the target nowcast month). The notation (h) indicates that the coefficients are independently estimated for each forecast horizon (h) . In nowcasting monthly inflation, h will range from 1 to 2, whereas in nowcasting quarterly inflation, h will range from 1 to 4. An assumption of $\sum_{j=0}^{P(HF)-1} \omega_{P(HF)-j}(\theta_{(h)}^{HF}) = 1$ helps identify β_h .

Density construction: Drawing errors from the normal distribution

Let T be the total number of observations (i.e., the length of the estimation window).

1. For $h = 1, \dots, 4$
2. Estimate the model specified in equation (1) using nonlinear least squares to obtain the parameter estimates $\hat{\alpha}_{(h)}, \hat{\chi}_{(h)}, \hat{\gamma}_{(h)}, \hat{\beta}_{(h)}(\hat{\theta}_{(h)})$

3. Based on the estimates in the previous step, compute the sequence of residuals \hat{e}_{t+h}

4. For $d = 1, \dots, D$

a. Sample e_{t+h}^* from the empirical distribution of $e_{t+h} \sim N(0, \text{var}(\check{e}_{t+h}))$, where $\check{e}_{t+h} = (\frac{T}{T-k})^{0.5} \hat{e}_{t+h}$ and k is the number of regressors in eq.(1).

b. Generate a simulated series π_{t+h}^* using

$$\pi_{t+h}^*{}^{(d)} = \hat{\alpha}_h + \sum_{j=0}^{P(M)-1} \hat{\chi}_{j+1,(h)} \pi_{t-j} + \sum_{j=0}^{P(M)-1} \hat{\gamma}_{j+1,(h)} Z_{t-j} + \hat{\beta}_h \sum_{j=0}^{P(HF)-1} \omega_{P(HF)-j}(\hat{\theta}_{(h)}^{HF}) X_{P(HF)-j,t+1}^{HF} + e_{t+h}^*$$

c. REPEAT

5. The empirical distribution $\{\pi_{t+h}^*\}_{d=1}^D$ constitutes the estimate of the density nowcast corresponding to the forecast horizon, h

Note that, in step 4a above, the draws are obtained from a distribution of modified residuals because the variance of the modified residuals is a better estimate of the true variance of the least squares estimate of the error term e_{t+h} in equation (1). To further explain why this is the case, recall that the variance of the residuals \hat{e}_{t+h} is the sum of the squared residuals divided by T , whereas the variance of the least squares estimate should be divided by $T - k$, where k is the number of regressors in the regression. Therefore, the original series of residuals are rescaled to correct the variance (see Davidson and MacKinnon, 2006).

This simple procedure accounts for shock uncertainty only; i.e., it does not account for the parameter uncertainty. However, in preliminary exercises, the difference in the density accuracy between this procedure and a bootstrapping procedure that also takes into account parameter uncertainty was very small.

B.1.2. DFM Model

Our implementation of the mixed-frequency DFM follows Modugno (2013) and KZ.

The dynamic factor model takes the general form:

$$y_\tau = C f_\tau + \varepsilon_\tau, \quad \varepsilon_\tau \sim N(0, \Sigma) \tag{B.2}$$

where τ refers to the trading-day frequency, y_τ is a vector of observations, C is a block diagonal matrix of factor loadings, ε_τ is a vector of idiosyncratic components, and f_τ is a vector of latent

common factors following VAR dynamics:

$$Bf_\tau = A(L)f_{\tau-1} + u_\tau, \quad u_\tau \sim N(0, Q) \quad (\text{B.3})$$

where B and $A(L)$ are coefficient matrices governing factor dynamics, some of which may be time-varying, and u_τ is a vector of residuals.

With monthly, weekly, and daily data, $y_\tau = [y_\tau^M, y_\tau^W, y_\tau^D]'$, we have three corresponding factors, $f_\tau = [f_\tau^M, f_\tau^W, f_\tau^D]'$, each of dimension $r \times 1$. The monthly factor(s) f_τ^M and the weekly factor(s) f_τ^W are a function of the daily factors(s) f_τ^D . Thus equations (B.2) and (B.3) can be written as:

$$\begin{bmatrix} y_\tau^M \\ y_\tau^W \\ y_\tau^D \end{bmatrix} = \begin{bmatrix} C_M & 0 & 0 \\ 0 & C_W & 0 \\ 0 & 0 & C_D \end{bmatrix} \begin{bmatrix} f_\tau^M \\ f_\tau^W \\ f_\tau^D \end{bmatrix} + \begin{bmatrix} \varepsilon_\tau^M \\ \varepsilon_\tau^W \\ \varepsilon_\tau^D \end{bmatrix} \quad (\text{B.4})$$

and

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_\tau^M \\ f_\tau^W \\ f_\tau^D \end{bmatrix} = \begin{bmatrix} \Theta_\tau^M & 0 & 0 \\ 0 & \Theta_\tau^W & 0 \\ 0 & 0 & A_D \end{bmatrix} \begin{bmatrix} f_{\tau-1}^M \\ f_{\tau-1}^W \\ f_{\tau-1}^D \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ u_\tau^D \end{bmatrix} \quad (\text{B.5})$$

The matrices C_M, C_W , and C_D are the loadings for the monthly, weekly, and daily variables. Θ_τ^M and Θ_τ^W are time-varying coefficients: Θ_τ^M is equal to zero the day after the release of the monthly data and is equal to one elsewhere; similarly, Θ_τ^W is equal to zero the day after the release of the weekly data and is equal to one elsewhere.

Assuming that the monthly variables and weekly variables in our system at any time τ represent a stock (i.e., a snapshot), accordingly the monthly first difference (or growth rate) and weekly first difference (or growth rate) of those variables can be formed by summing up their respective daily first differences (or growth rates).

To produce forecasts far into the future, the daily factors are forecast via the transition equation (B.5) and are translated to daily nowcasts and aggregated to weekly and monthly nowcasts via equation (B.4). Following Modugno (2013), we estimate the model with the expectation-maximization (EM) algorithm as detailed in Bańbura and Modugno (2014).

Density construction: Standard bootstrapping procedure

Our procedure closely follows the factor model bootstrapping procedure detailed in Aastveit et al. (2014).

Let T be the number of observations (i.e., the length of the estimation window).

1. Estimate the model specified in equations (B.2) and (B.3) to obtain parameter estimates

$\hat{A}^{(0)}, \hat{B}^{(0)}, \hat{C}^{(0)}, \hat{Q}^{(0)}, \hat{\Sigma}^{(0)}, \hat{f}^{(0)}$. Let $\hat{A} = \hat{A}^{(0)}, \hat{B} = \hat{B}^{(0)}, \hat{C} = \hat{C}^{(0)}, \hat{Q} = \hat{Q}^{(0)}, \hat{\Sigma} = \hat{\Sigma}^{(0)}$.

2. For $d=1, \dots, D$, do the following

a. Simulate draws u_τ^* from the empirical distribution of $u_\tau \sim N(0, \hat{Q})$

b. Generate bootstrap series f_τ^* using $\hat{B}f_\tau^* = \hat{A}(L)f_{\tau-1}^* + u_\tau^*$ where u_τ^* is obtained in the previous step

c. Simulate draws ε_τ^* from the empirical distribution of $\varepsilon_\tau \sim N(0, \hat{\Sigma})$

d. Generate bootstrap series y_τ^* using $y_\tau^* = \hat{C}f_\tau^* + \varepsilon_\tau^*$ where ε_τ^* and f_τ^* are obtained in the previous two steps.

e. Using y_τ^* re-estimate the model in equations (B.2) and (B.3) to obtain an updated set of parameter and factor estimates, $\hat{A}^{(d)}, \hat{B}^{(d)}, \hat{C}^{(d)}, \hat{Q}^{(d)}, \hat{\Sigma}^{(d)}, \hat{f}^{(d)}$. Set $\hat{A} = \hat{A}^{(d)}, \hat{B} = \hat{B}^{(d)}, \hat{C} = \hat{C}^{(d)}, \hat{Q} = \hat{Q}^{(d)}, \hat{\Sigma} = \hat{\Sigma}^{(d)}$.

f. Based on the parameter and factor estimates obtained in the previous step construct forecasts of factors via equation (B.3), which are then aggregated up to produce nowcasts (and forecasts) for monthly inflation, $\pi_{t+h}^*{}^{(d)}$ via equation (B.2).

g. REPEAT

3. The empirical distribution $\{\pi_{t+h}^*\}_{d=1}^D$ constitutes the estimate of the density nowcast corresponding to the forecast horizon, h

B.1.3. DMS Model

As discussed in the body of chapter 3, the DMS model is essentially a collection of univariate and multivariate regressions applied to disaggregate components and aggregate inflation. To appropriately account for uncertainty, we devise two separate bootstrapping algorithms for univariate and multivariate formulations. The difference between these two algorithms is only slight but it helps improve the density accuracy of monthly inflation.

We first describe the general-purpose bootstrap algorithm for the multivariate regression followed by the description for the univariate regression.

A general representation for a multivariate regression can be written as follows,

$$y_t = \beta_0 + \alpha X_t + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma^2) \quad (\text{B.6})$$

Assume that $\hat{\beta}_0$, $\hat{\alpha}$, $\hat{\sigma}^2$ are the OLS estimates obtained through the estimation of equation (B.6) over the sample $1, \dots, T$. $\hat{\varepsilon}_t$ are the least squares residuals with mean 0 and variance $\hat{\sigma}^2$.

Density construction, algorithm 1: Wild block bootstrap for density forecasts

For $d=1, \dots, D$ do the following.

1. Construct a transformed series of residuals $\{\tilde{\varepsilon}_t\}_{t=1}^T$ from the OLS residuals $\{\hat{\varepsilon}_t\}_{t=1}^T$, where $\tilde{\varepsilon}_t = h(\hat{\varepsilon}_t)u_t$ and $u_t \sim N(0, 1)$. h is a transformation function that modifies the original least squares residuals to correct them for possible heteroscedasticity. Various choices for h have been suggested in the literature. Following Chernick and LaBudde (2011, Ch. 6, Section 6.6), we set

$$h(\hat{\varepsilon}_t) = \frac{\hat{\varepsilon}_t}{1-H} \quad \text{where } H = X(X'X)^{-1}X'$$

We also tried $h(\hat{\varepsilon}_t) = \frac{\hat{\varepsilon}_t}{(1-H)^{1/2}}$, another widely used transformation.

2. Sampling from $\tilde{\varepsilon}$

a. To correct for possible serial correlation (following Aastveit et al., 2014), we draw blocks of consecutive errors from $\tilde{\varepsilon}$. We define the block size, $b_{size} = 4$; it is common to set it greater than or equal to the forecast horizon; T is the number of observations; and $b_{number} = \text{ceil}(\frac{T}{b_{size}})$, is an integer that denotes the number of non-overlapping blocks of consecutive errors.

b. For $l = 1, \dots, b_{size}$ and $j = 1, \dots, b_{number}$ construct the bootstrap sample for y^*

$$y_{(j-1)b_{size}+l}^* = \hat{\beta}_0 + \hat{\alpha}X_{(j-1)b_{size}+l} + \varepsilon_{(j-1)b_{size}+l}^*$$

where $\varepsilon_{(j-1)b_{size}+l}^* = \tilde{\varepsilon}_{(j-1)b_{size}+l} \cdot \delta_j$, and δ_j is set as a Rademacher variable, following Davidson and Flachaire (2008) and Aastveit et al (2014):

$$\delta_j = \begin{cases} +1, & \text{with probability 0.5} \\ -1, & \text{with probability 0.5} \end{cases}$$

We also experimented with $\delta_j \sim N(0, 1)$, but doing so slightly worsened the accuracy of the density forecasts.

3. Based on the bootstrap sample y^* (constructed in the previous step), re-estimate the model in equation (B.6) to obtain updated estimates $\hat{\beta}_0^{(d)}, \hat{\alpha}^{(d)}, \hat{\sigma}^{2(d)}$.

4. Use $\hat{\beta}_0^{(d)}$ and $\hat{\alpha}^{(d)}$ in equation (B.6) to generate iterative forecasts, $\hat{y}_{t+h}^{(d)}$ up to h periods ahead. (We also experimented with a modified step 4: when generating iterative forecasts $\hat{y}_{t+h}^{(d)}$ we drew from $\varepsilon^* \sim N(0, var(\varepsilon^*))$ for each h . This alternative made no difference to the overall results.)

5. REPEAT

6. The empirical distribution of $\{\hat{y}_{t+h}\}_{d=1}^D$ constitutes our estimate of the h -step-ahead density.

Next, we describe the algorithm that we apply to the univariate AR regressions. A general representation for a univariate AR regression is:

$$y_t = \beta_0 + \sum_{j=1}^P \alpha_j y_{t-j} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (\text{B.7})$$

Assume that $\hat{\beta}_0, [\hat{\alpha}_j]_{j=1}^P, \hat{\sigma}^2$ are the OLS estimates obtained through the estimation of equation (B.7) over the sample consisting of $1, \dots, T$ observations. $\hat{\varepsilon}_t$ are the least squares residuals with mean 0 and variance $\hat{\sigma}^2$.

Density construction, algorithm 2: Parametric bootstrap for density forecasts

For $d=1, \dots, D$ do the following.

1. Construct a transformed series of residuals $\{\tilde{\varepsilon}_t\}_{t=1}^T$ from the residuals $\{\hat{\varepsilon}_t\}_{t=1}^T$, where $\tilde{\varepsilon}_t = (\frac{T}{T-k})^{0.5} \hat{\varepsilon}_t$ and k is the number of regressors, in this case $k = P + 1$; P is the number of lags of the dependent variable. We also experimented with $\tilde{\varepsilon}_t = h(\hat{\varepsilon}_t)u_t$ and $u_t \sim N(0, 1)$ but this produced inferior nowcasts.

2. Sample a sequence of $\{\varepsilon_t^*\}_{t=1}^T$ from $\varepsilon \sim N(0, var(\tilde{\varepsilon}))$ and then construct a bootstrap sample of $\{y_t^*\}_{t=1}^T$ using

$$y_t^* = \hat{\beta}_0 + \sum_{j=1}^P \hat{\alpha}_j y_{t-j}^* + \varepsilon_t^*$$

3. Based on the bootstrap sample y^* re-estimate the model in equation (B.7) to obtain updated estimates $\hat{\beta}_0^{(d)}, [\hat{\alpha}_j^{(d)}]_{j=1}^P$

4. Use $\hat{\beta}_0^{(d)}$, $[\hat{\alpha}_j^{(d)}]_{j=1}^P$ in equation (B.7) to iteratively generate forecasts, $\hat{y}_{t+h}^{(d)}$ up to h periods ahead. (We also experimented with a modified step 4: when generating iterative forecasts $\hat{y}_{t+h}^{(d)}$ we draw from $\varepsilon \sim N(0, \text{var}(\varepsilon))$ for each h . This alternative made no difference to the overall results.)

5. REPEAT

6. The empirical distribution of $\{\hat{y}_{t+h}\}_{d=1}^D$ constitutes our estimate of the h -step-ahead density.

Using the same notation as in KZ, the general representation of the DMS model for monthly headline (or core) inflation is

$$A_{s(\tau)}Z_t = B_{s(\tau)} + C_{s(\tau)}X_t + \sum_{j=1}^J D_{j,s(\tau)}Z_{t-j} + \varepsilon_{s(\tau)} \quad (\text{B.8})$$

where Z_t is an $n \times 1$ vector of aggregates, X_t is an $m \times 1$ vector of disaggregates that are informative over Z_t , and $\varepsilon_{s(\tau)} \sim N(0, \Sigma)$. The coefficient matrices A , B , C , and D_j are $n \times n$, $n \times 1$, $n \times m$, and $n \times n$, respectively, and are allowed to vary over time depending on the available information set, denoted $s(\tau)$; in particular, C and D_j measure the weights put on the disaggregates and lagged aggregates, respectively.

Nowcasting core inflation

Let $Z_t = [\pi_t^{\text{CoreCPI}}, \pi_t^{\text{CorePCE}}]'$ and $X_t = 0$ in equation (B.8). We specify two possible regression specifications for core inflation. The first one is a univariate AR, and the second is a bridge equation (i.e., multivariate regression), which regresses core CPI on core PCE and a constant. Conditional on the available information, equation (B.8) reduces to either a univariate AR or a combination of a univariate AR and bridge equation.

Univariate AR: $\pi_t^{\text{Core}} = \beta_0 + \sum_{j=1}^P \alpha_j \pi_{t-j}^{\text{Core}} + \varepsilon_t$

Bridge equation: $\pi_t^{\text{CorePCE}} = \gamma_0 + \theta \pi_t^{\text{CoreCPI}} + u_t$

In cases where we have an additional monthly release of core CPI compared with core PCE, and only core PCE remains to be nowcasted: (1) The forecasts of core CPI are produced using a univariate AR, and algorithm 2 is used to produce density forecasts. (2) The nowcast of core

PCE is produced using a bridge regression. The forecasts up to h steps ahead are produced using a univariate regression that treats the nowcast from a bridge regression as an initial value. To produce density estimates (nowcasts and forecasts), algorithm 2 is used. In all other cases, both core CPI and core PCE are nowcasted (and forecasted) using a univariate AR model. The density estimates are computed based on algorithm 2.

Nowcasting food inflation

Nowcasts for food inflation are produced and used to nowcast headline inflation in all cases except: (1) when we are unable to produce a nowcast for gasoline inflation, and (2) when we have an additional reading for PCE inflation (π^{PCE}) compared to CPI inflation (π^{CPI}). Similar to core PCE, we adopt a parsimonious approach to produce nowcasts of food inflation by simply estimating a univariate AR,

$$\pi_t^{food} = \beta_0 + \sum_{j=1}^P \alpha_j \pi_{t-j}^{food} + \varepsilon_t$$

Density nowcasts (and forecasts) are produced using algorithm 2.

Nowcasting gasoline inflation

Following KZ, we generate nowcasts (and forecasts) for gasoline inflation based on the availability of weekly gasoline prices and daily oil prices. If weekly gasoline prices are available in the current month, these form the basis for that month's gasoline inflation nowcast. We use a daily random walk in oil prices to extend (i.e., forecast) the oil price series by one additional month. If oil price data or a forecast for oil prices is available for a month but gasoline prices are not available from within that month, then we produce nowcasts or forecasts for gasoline inflation ($\hat{\pi}^{Gasoline}$) via a two-stage regression procedure (see KZ for details). In the first stage, a longer-run relationship between monthly gasoline prices and monthly oil prices is assumed via the following regression:

$$P_{t-1}^{Gasoline(NSA)} = \alpha + \beta P_{t-1}^{Oil} + e_{1,t-1} \quad (\text{B.9})$$

Denote $\tilde{P}_{t-1}^{Gasoline(NSA)}$ as the fitted monthly gasoline prices obtained by estimating equation (B.9). In the second stage, we estimate an error correction model that uses the lagged gap between gasoline prices and their predicted (longer-run) values obtained in the first stage via the following regression:

$$\Delta P_{t-1}^{Gasoline(NSA)} = b \Delta P_{t-1}^{Oil} + c (P_{t-2}^{Gasoline(NSA)} - \tilde{P}_{t-2}^{Gasoline(NSA)}) + e_{2,t-1} \quad (\text{B.10})$$

Using the estimated coefficients in equations (B.9) and (B.10) and iterating forward equations (B.9) and (B.10) we generate $\hat{P}_{t-1+h}^{Gasoline(NSA)}$ and $\hat{\Delta P}_{t-1+h}^{Gasoline(NSA)}$ and in turn estimates of $\hat{\pi}_{t-1+h}^{Gasoline(NSA)}$. The estimates are seasonally adjusted to produce $\hat{\pi}_{t-1+h}^{Gasoline}$. The density forecasts are produced by applying algorithm 1 sequentially to equations (B.9) and (B.10). For each simulation d , $\hat{\pi}_{t-1+h}^{Gasoline(NSA),d}$ is seasonally adjusted to obtain the corresponding $\hat{\pi}_{t-1+h}^{Gasoline,d}$.

Nowcasting headline inflation

Let $Z_t = [\pi_t^{CPI}, \pi_t^{PCE}]'$ and $X_t = [\pi_t^{CoreCPI}, \pi_t^{CorePCE}, \pi_t^{Food}, \pi_t^{Gasoline}]'$. In cases where we have an additional release of π_t^{CPI} , equation (B.8) reduces to a bridge equation for π_t^{PCE} and a univariate AR for π_t^{CPI} .

$$\text{Univariate AR: } \pi_t^{CPI} = \beta_0 + \sum_{j=1}^P \alpha_j \pi_{t-j}^{CPI} + \varepsilon_t$$

$$\text{Bridge equation: } \pi_t^{PCE} = \gamma_0 + \theta \pi_t^{CPI} + u_t$$

Density estimates are constructed using algorithm 2. In cases where we have nowcasts of $\pi_t^{Gasoline}$, equation (B.8) reduces to a multivariate regression,

$$\pi_t^{CPI} = b_1 + c_{11}\pi_t^{CoreCPI} + c_{13}\pi_t^{Food} + c_{14}\pi_t^{Gasoline} + e_t^{CPI} \quad (\text{B.11})$$

$$\pi_t^{PCE} = b_2 + c_{22}\pi_t^{CorePCE} + c_{23}\pi_t^{Food} + c_{24}\pi_t^{Gasoline} + e_t^{PCE} \quad (\text{B.12})$$

The density nowcasts (and forecasts) for CPI and PCE inflation are produced by separately applying algorithm 1 to equations (B.11) and (B.12). In very few cases, where we lack estimates of $\tilde{\pi}_t^{Gasoline}$ and do not have an additional reading for π_t^{CPI} , equation (B.8) reduces to univariate AR,

$$\pi_t^{CPI} = \beta_1 + \sum_{j=1}^P \alpha_j^{CPI} \pi_{t-j}^{CPI} + \varepsilon_t^{CPI} \quad (\text{B.13})$$

$$\pi_t^{PCE} = \beta_2 + \sum_{j=1}^P \alpha_j^{PCE} \pi_{t-j}^{PCE} + \varepsilon_t^{PCE} \quad (\text{B.14})$$

The density nowcasts (and forecasts) are generated by separately applying algorithm 2 on equations (B.13) and (B.14).

In all of our simulation procedures, $D=500$. Early experimentation suggested that we would normally obtain similar results if we instead set $D=1000$.

B2. Mechanics of Density Combination and Graphical Illustration

Assume at time t , we have $i = 1, \dots, M$ (potentially different) empirical distributions $f_{i,t}(y_t)$ for a variable y_t . We wish to combine them using a given set of M weights, $w_{i,t}$.

Step 1: Looking across all the M empirical distributions, $f_{i,t}(y_t)$, determine the (global) minimum value and (global) maximum value of y_t . Denote x_t^{min} as the minimum value and x_t^{max} as the maximum value.

Step 2: Define a grid $x_t \in \{x_t^{min}, \dots, x_t^{max}\}$ of S equally spaced intervals such that $x_{k-1} < x_k$.

Step 3: Transform each of the $i = 1, \dots, M$ empirical distributions $f_{i,t}(y_t)$ to a probability density function (pdf), $p_{i,t}(y_t)$ using the grid x_t as the domain. The Gaussian kernel function (Matlab: `ksdensity` function) is applied to construct a smoothed $p_{i,t}(y_t)$. Using the same grid x_t to construct each of the M pdfs will guarantee that all the pdfs that are to be combined together at time t have the same domain; that is, they are all positioned over the same grid.

Step 4: With all pdfs positioned over the same domain (grid), the combination can be achieved by simply adding up the M different densities using the corresponding weights $w_{i,t}$ (for linear combination) or raised to a power of $w_{i,t}$ for a log pool combination. The combined density $g_t(y_t)$ will also be positioned over the same grid (domain) x_t as the M individual densities.

We set $S=500$. Early experimentation suggested that the results were very similar if we set $S=1000$.

Note that our procedure dynamically adjusts the grid x_t at each time t . Alternatively, we could just set it to a predefined interval but then the interval has to be wide enough to encompass all the individual empirical distributions for all $t = 1, \dots, T$ (i.e., over the evaluation sample). Given the breadth of our analysis, including the number of variables considered and both monthly and quarterly rates, having a grid that adjusts dynamically was more efficient for our application.

In implementing our algorithm, we have benefitted from and are grateful for the PROFOR Matlab toolbox (developed by researchers at the Norges Bank, Bank of England, and Warwick Business School). We have modified some of the functions of the toolbox to fit our needs.

B3. Comparing Properties of Grand Combinations across Weighting Schemes

Figure 3.8 in the body of chapter 3 shows the weights and higher-order moments from using the log score weighting scheme to generate the stage 1 and stage 2 combinations. Figures B15, B16, and B17 (in this appendix) show the weights and higher-order moments from the CMG, Ganics, and CRPS weighting schemes, respectively. We summarize six key results from this comparison.

First, for CPI inflation and PCE inflation, the DMS combination gets the highest weight in all weighting schemes with the exception of Ganics. Furthermore, the DMS maintains its ranking with incoming information over the course of the month.

Second, the CMG and Ganics grand combinations for CPI inflation and PCE inflation provide stronger evidence of both kurtosis and skewness than the combination based on log score weights. This finding is associated with the grand combination being composed of more diverse components in these cases; that is, the DMS combination, the DFM combination, and the MIDAS combination are all assigned nonzero weights in the grand combination. Different weighting schemes can lead to combinations with very different compositions, as is evident by very different profiles of the weights assigned to the three model classes over time. In general, the greater the diversity in the composition of the grand combination, the greater is the evidence of skewness and kurtosis.¹ But greater flexibility in terms of accommodating skewness and kurtosis does not necessarily translate into improved accuracy. We say this because for CPI inflation the grand combination based on the log-score weighting scheme is more accurate than grand combinations based on other schemes, yet it displays less evidence of skewness and kurtosis on average compared with other grand combinations. This improved accuracy is mainly coming from the significantly more accurate mean of the density nowcast constructed from the log-score scheme, which puts high weight on the stage 1 DMS combination, compared with grand combinations based on other weighting schemes.

Third, in the case of core inflation, the patterns observed in the properties of the grand combination are generally comparable across the various weighting schemes, even though the weights assigned to the densities of the three modeling classes differ. This result stems from the fact that the estimates of density nowcasts for core inflation are generally similar across the different modeling classes; so irrespective of the approach used to combine the component density nowcasts, the resulting estimates of the combined density nowcasts are similar. This

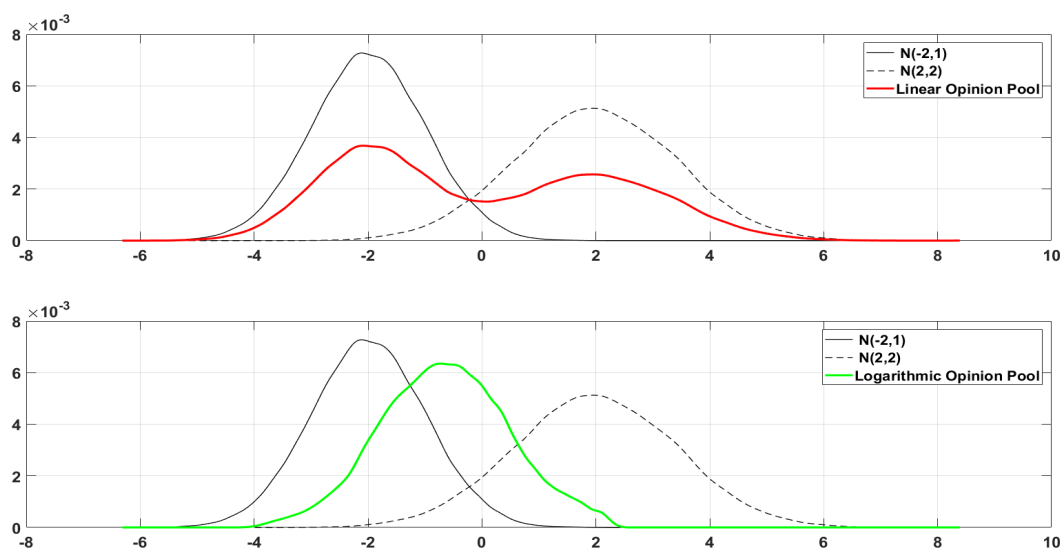
¹We highlight a result in regard to grand combinations for CPI inflation (case 4) produced using the log score weighting scheme (see Figure 3.8) and the CMG weighting scheme (Figure B15). Both schemes assign a weight of 100% to DMS at least in the last few years of the evaluation sample, yet the profiles of the kurtosis property of the grand combinations across the two schemes are very different for this period. This finding arises in part because the underlying composition of the two respective stage 1 DMS combinations is quite different; see Figure B18.

latter pattern also explains the comparable accuracy results for core inflation shown in Figures 3.4 and 3.5 of chapter 3 (especially in the case of core CPI). Relatedly, the weight profiles across different weighting schemes (for core inflation) indicate a high incidence of fast switching across the three combinations. The evidence of time-varying switching across density combinations highlights the importance of combining density estimates from a range of models to circumvent the instability issues of using a single model.

Fourth, the CRPS weighting scheme assigns positive weights to the three combinations across all inflation measures and at all representative dates (shown for cases 1 and 4), reflecting the generous assessment of the CRPS metric. In the case of core inflation, the weights are pretty evenly distributed across the DMS, DFM, and MIDAS combinations.

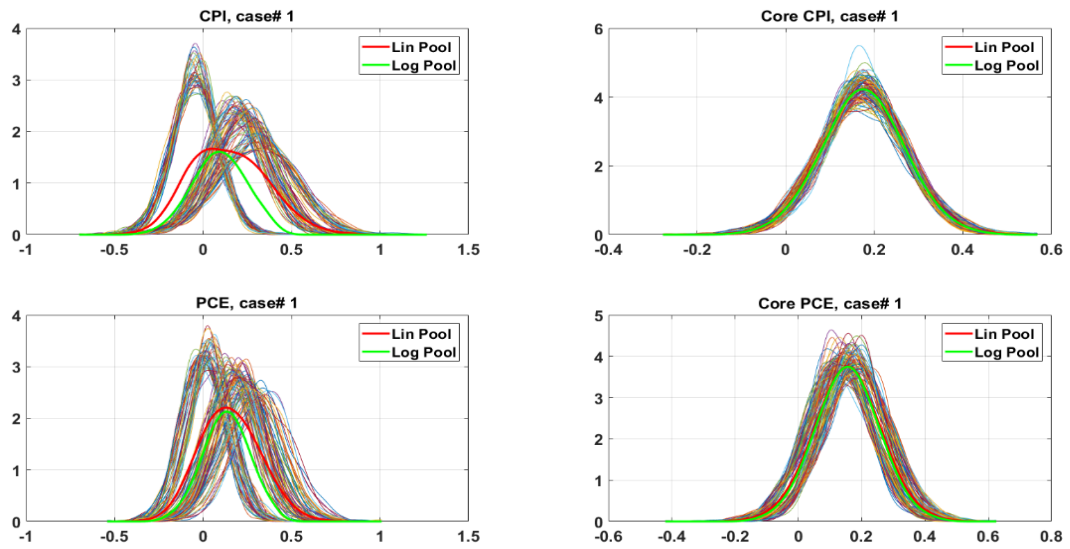
Fifth, in our application, the two optimal combination weighting schemes (CMG and Ganics) yield weight profiles that are remarkably different, especially in the case of CPI inflation and PCE inflation. However, the different profiles are not unexpected, given the earlier results that showed MIDAS and DFM combinations producing well-calibrated densities compared with DMS, which tends to do quite well in relative accuracy scoring. The weights produced from the Ganics approach display quite a bit of variability early in the sample. This variability is also present to a degree in the results reported in Ganics (2017) using industrial production data.

Figure B1: Illustration of Combining Densities with Linear and Log Opinion Pools



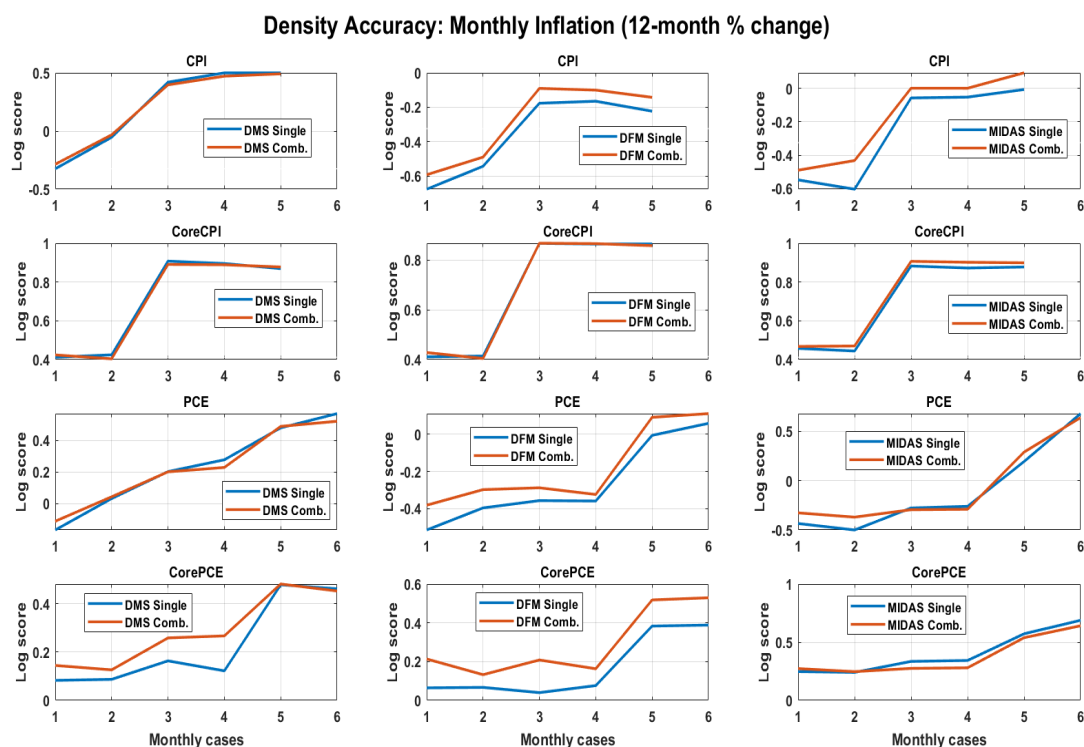
Notes: A simple example (motivated by Kascha and Ravazzolo, 2010) on combining two densities with very different mean and variances via two different functional forms.

Figure B2: Example Stage 1 DMS Combination



Notes: Single specification density nowcasts (thin lines) underlying the stage 1 DMS combination, linear pool nowcasts (thick red lines), and log pool nowcasts (thick green lines) for case 1 for the month of September 2000.

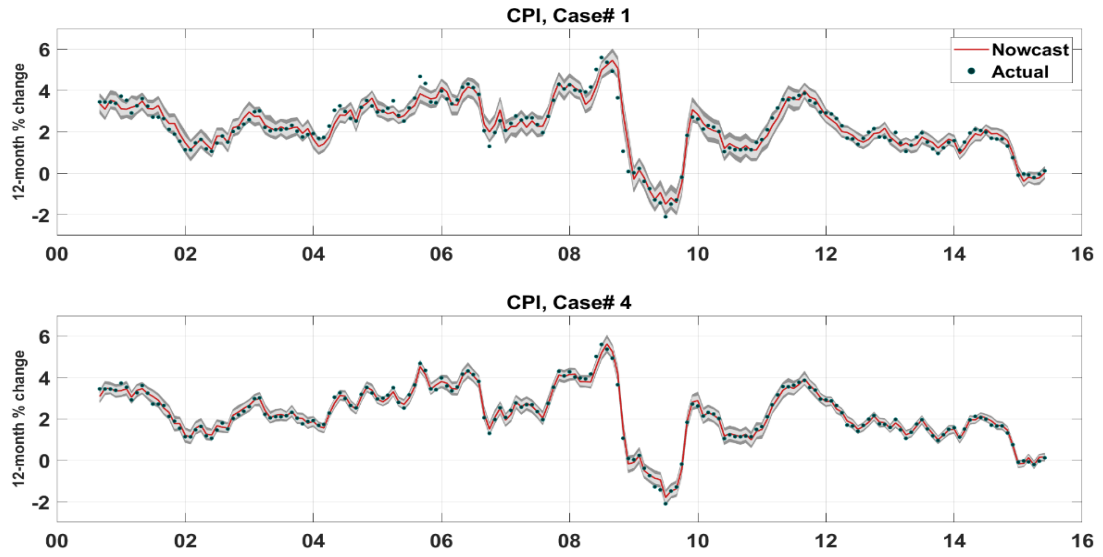
Figure B3: Comparisons between Single Specifications vs. Stage 1 Combinations



Notes: Average log scores at different nowcast origins for single specifications and stage 1 combinations within model classes. The evaluation sample is September 2000 through June 2015. We exclude September 2001 and October 2001 from the average log score calculations for PCE inflation and core PCE inflation.

Figure B4: Real-Time Density Nowcasts

(a) CPI inflation



(b) Core CPI inflation

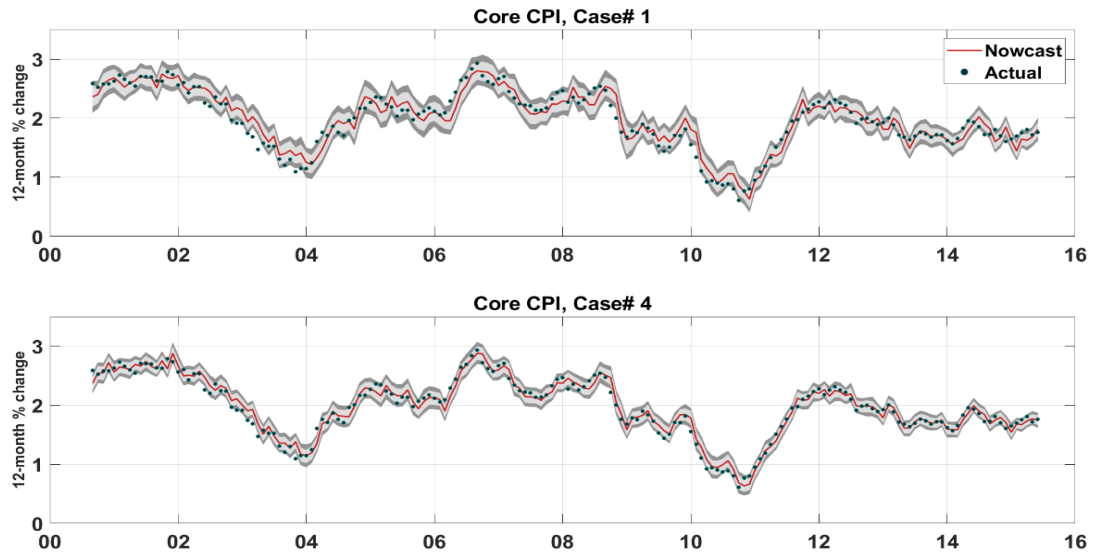
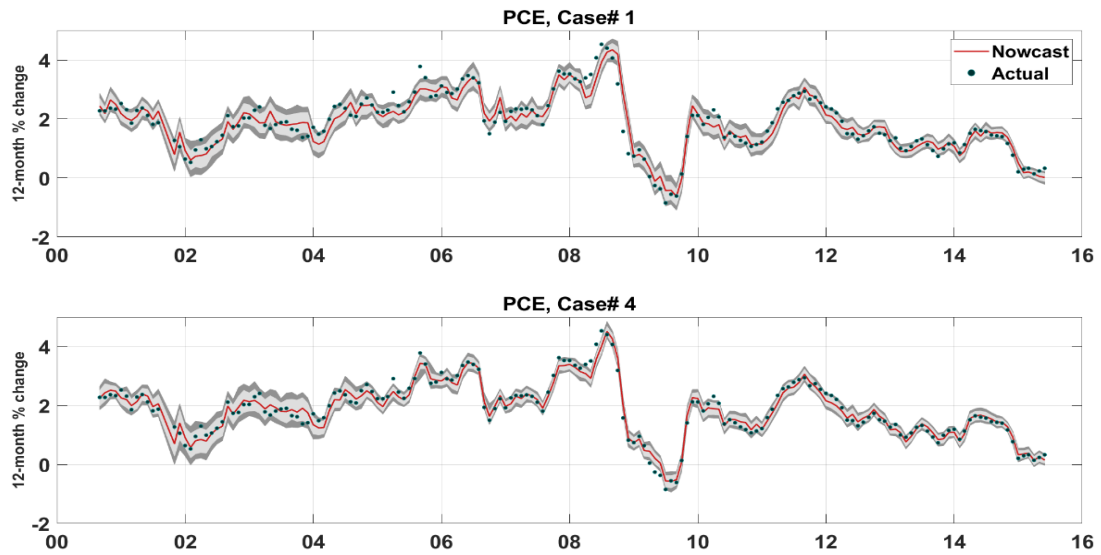
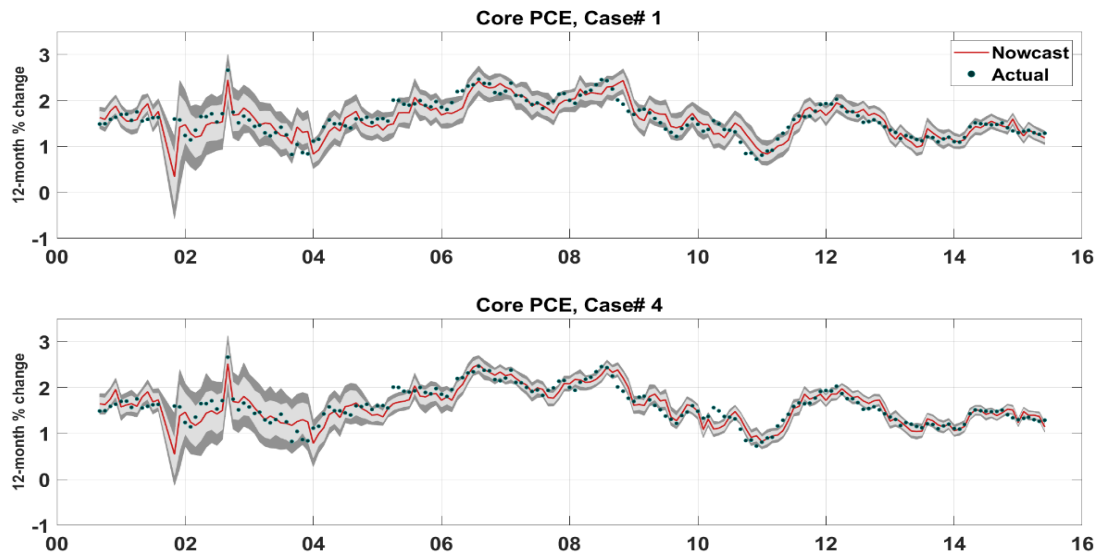


Figure B4: Real-Time Density Nowcasts (continued)

(c) PCE inflation

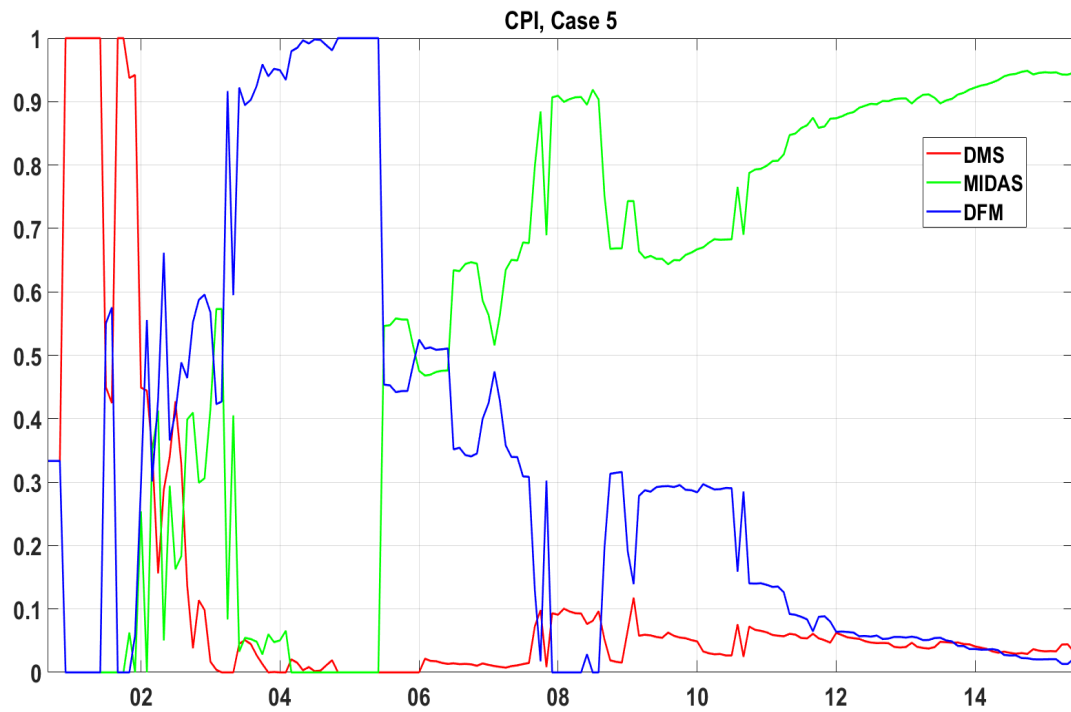


(d) Core PCE inflation



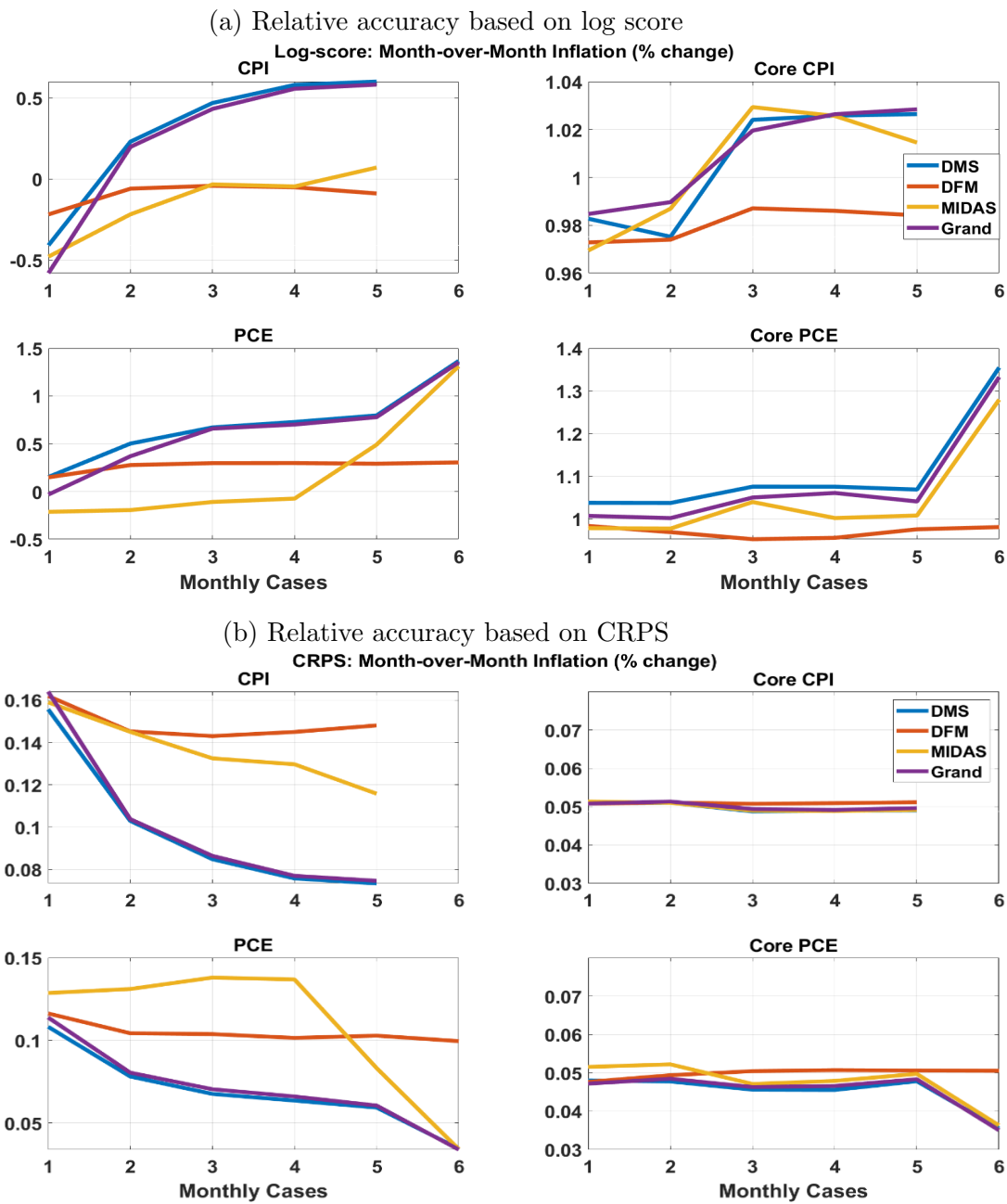
Notes: The figure shows the out-of-sample nowcasts generated using real-time data from the grand combination with the log score weighting scheme and the flexible aggregation strategy at two different points in each month (case 1 and case 4) for the 12-month trailing inflation rate. The shaded areas represent 70% and 90% prediction intervals. The sample period spans September 2000 through June 2015.

Figure B5: Weights Underlying Grand Combination based on Ganics Weighting Scheme



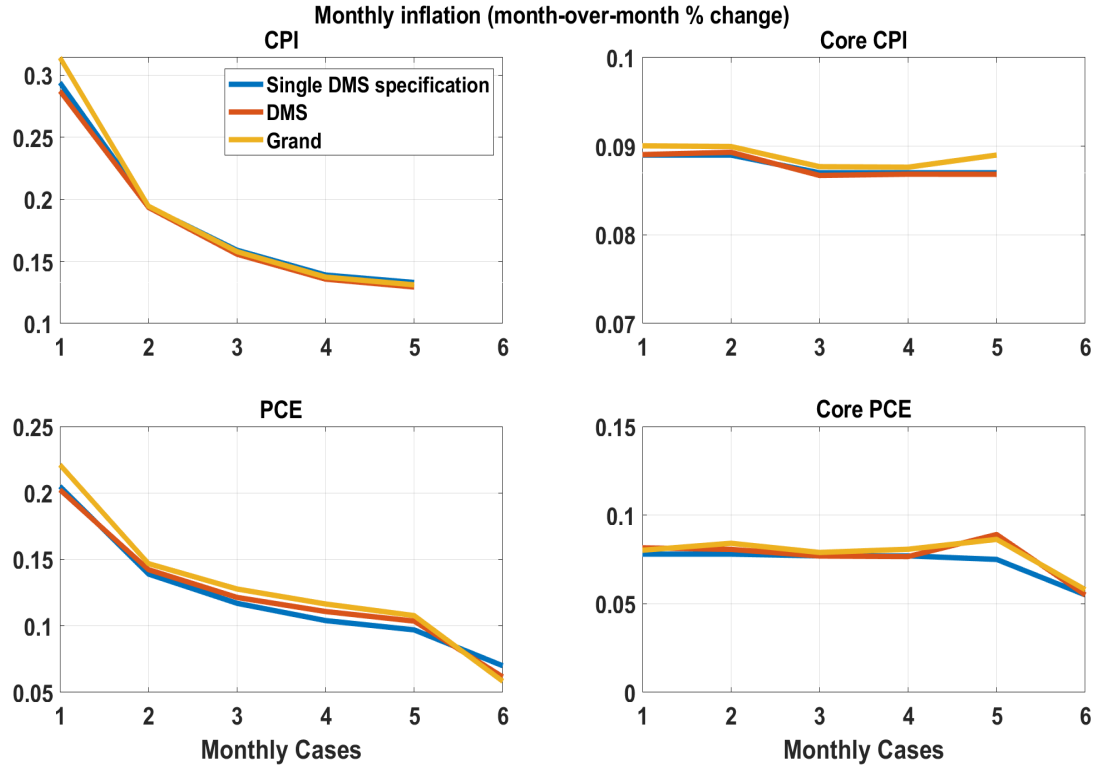
Notes: The figure plots the evolution of the weights applied to each of the stage 1 density combinations from the DMS, MIDAS, and DFM model classes to form the stage 2 combination, based on nowcasts generated for monthly (year-over-year) inflation at case 5 for nowcasting CPI inflation. The sample period spans September 2000 through June 2015.

Figure B6: Density Performance of Grand Combination vs. Its Components: Month-Over-Month Inflation



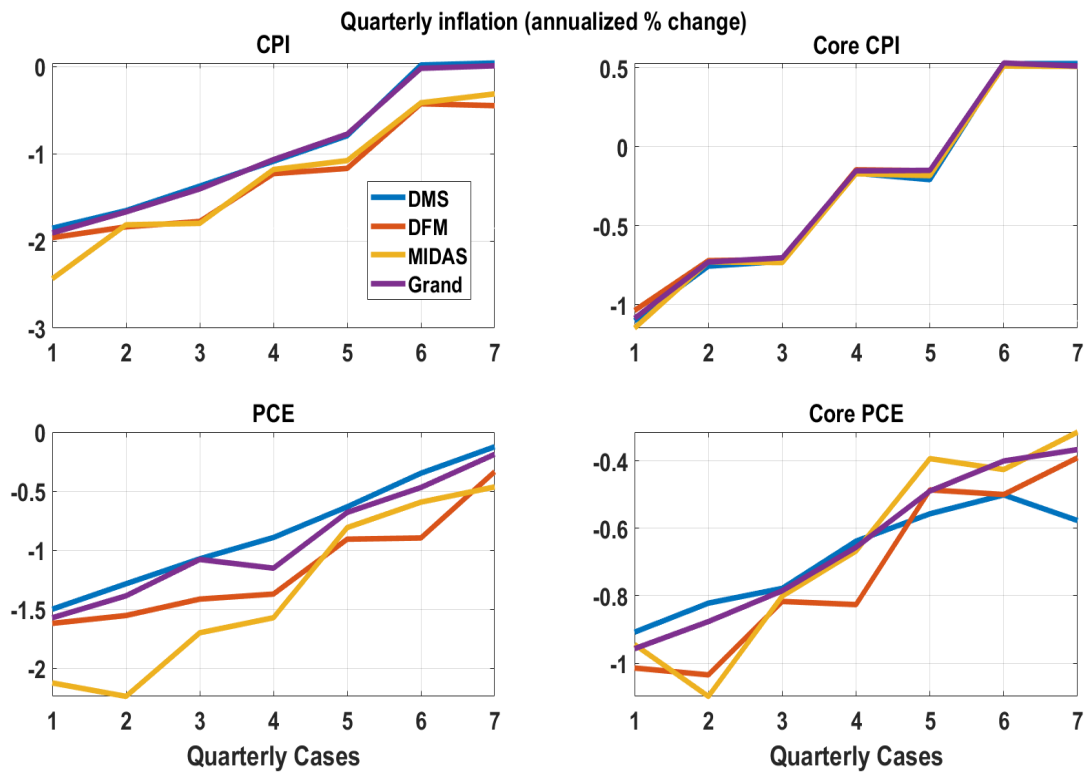
Notes: The top panel plots the average log score and the bottom panel plots the average CRPS for the grand combination based on the log score weighting scheme and combinations based on the DMS model class, MIDAS model class, and DFM model class, where each individual model class uses the log score weighting scheme. The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure B7: Point Nowcasting Performance, Grand Combination vs. DMS: Month-Over-Month Inflation



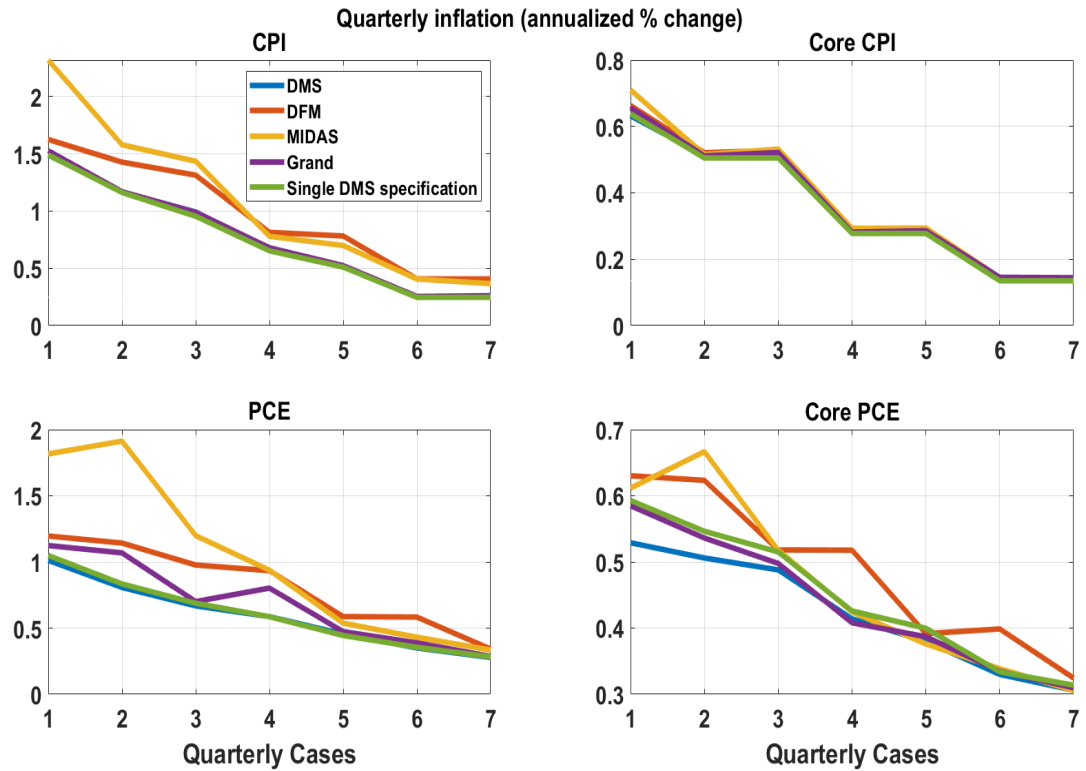
Notes: The figure plots the RMSE for the grand combination based on log score and using the flexible aggregation strategy; the stage 1 combination from the DMS model class; and a single specification from the DMS model class based on Knotek and Zaman (2017). The cases reflect the point in time when each nowcast was made relative to the target nowcast month; see Table 3.2 (in chapter 3). The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure B8: Density Performance of Grand Combination vs. Its Components: Quarterly Inflation



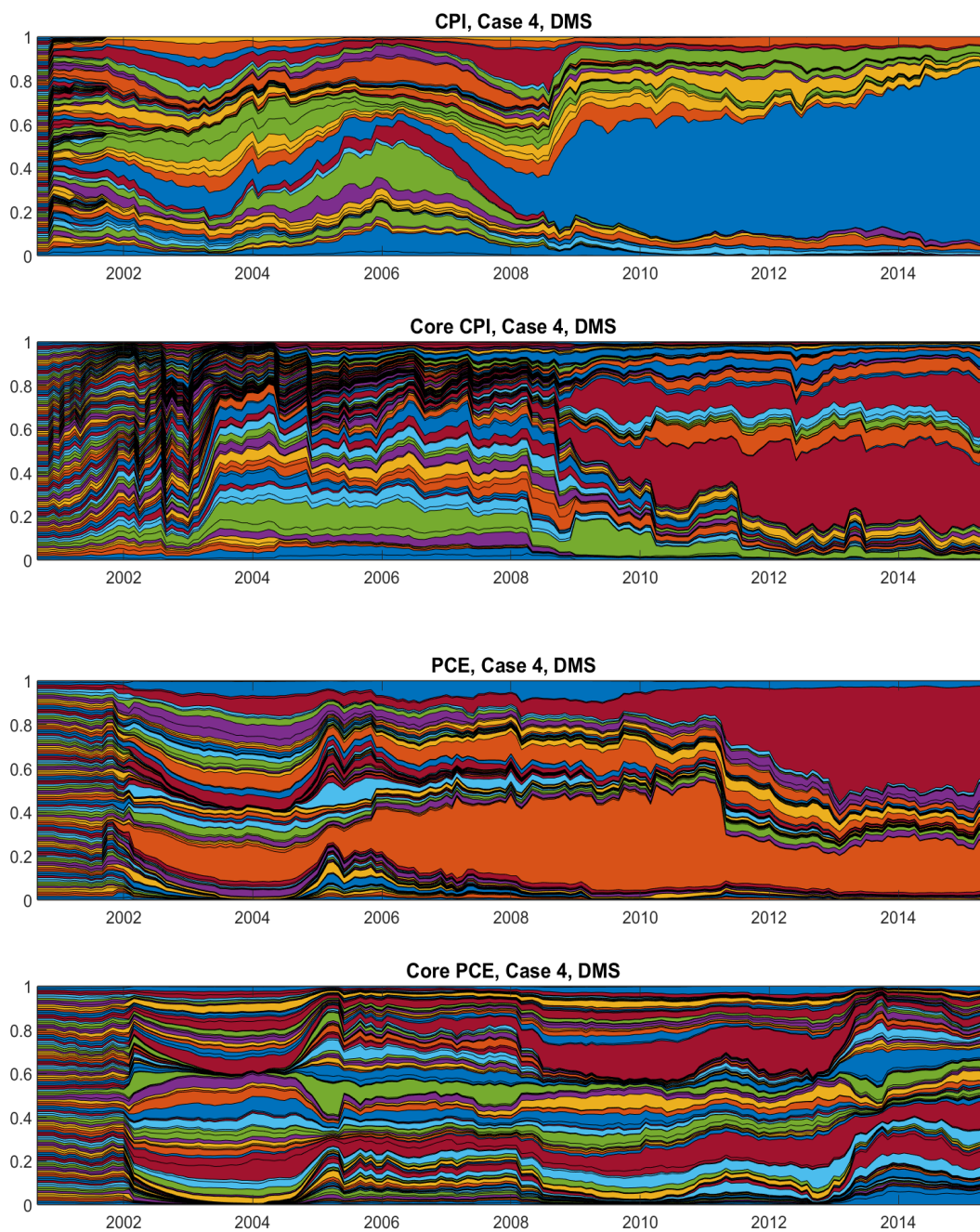
Notes: Average log score for the grand combination based on the log score weighting scheme and combinations based on the DMS model class, MIDAS model class, and DFM model class, where each individual model class uses the log score weighting scheme. The evaluation sample runs from 2000Q4 through 2015Q2; we omit 2001Q3 and 2001Q4 for PCE inflation and core PCE inflation calculations.

Figure B9: Point Nowcasting Performance, Grand Combination vs. Other Combinations and Single DMS Specification: Quarterly Inflation



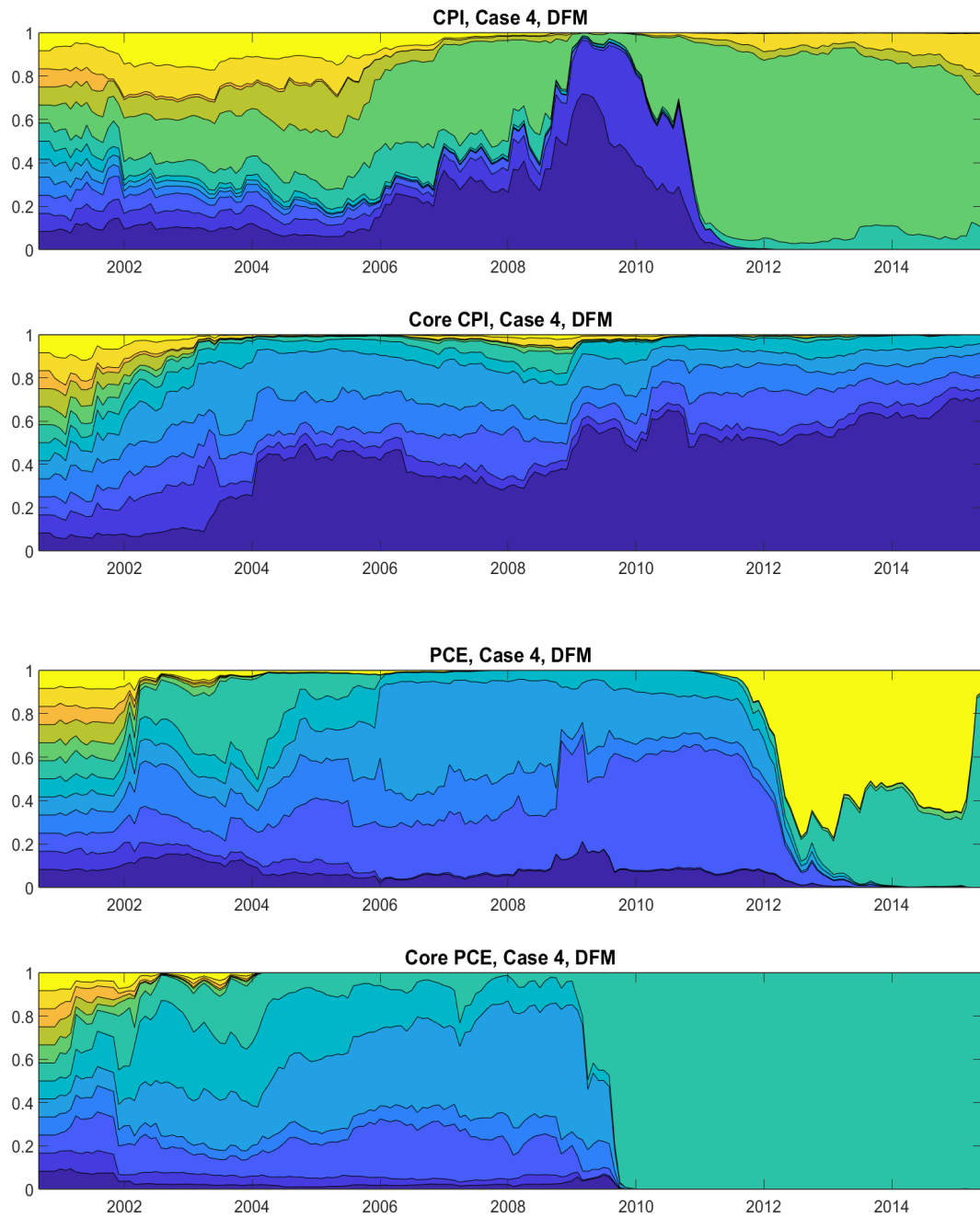
Notes: The figure plots the RMSE for the grand combination based on log score and using the flexible aggregation strategy; the stage 1 combinations from the DMS model class, DFM model class, and MIDAS model class; and a single specification from the DMS model class based on Knotek and Zaman (2017). The cases reflect the point in time when each nowcast was made relative to the target nowcast quarter; see Table B1. The evaluation sample runs from 2000Q4 through 2015Q2; we omit 2001Q3 and 2001Q4 for PCE inflation and core PCE inflation calculations.

Figure B10: Weights for Stage 1 DMS Combinations, Log Score Weighting Scheme



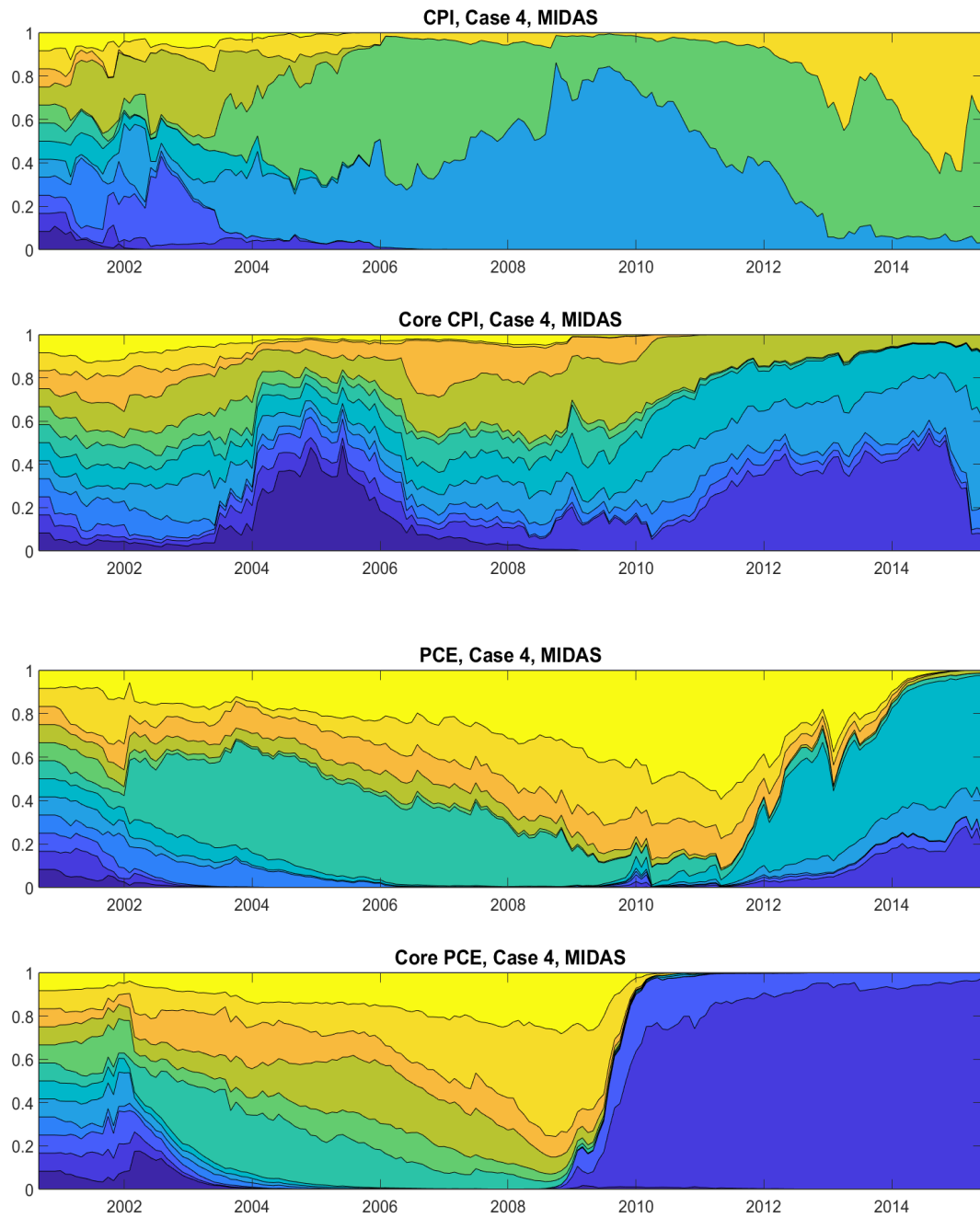
Notes: The figure plots the evolution of the weights for underlying individual candidate densities for the stage 1 DMS combination at case 4. Each color shade represents a particular individual candidate density. There are 108 candidate densities. The sample period spans September 2000 through June 2015.

Figure B11: Weights for Stage 1 DFM Combinations, Log Score Weighting Scheme



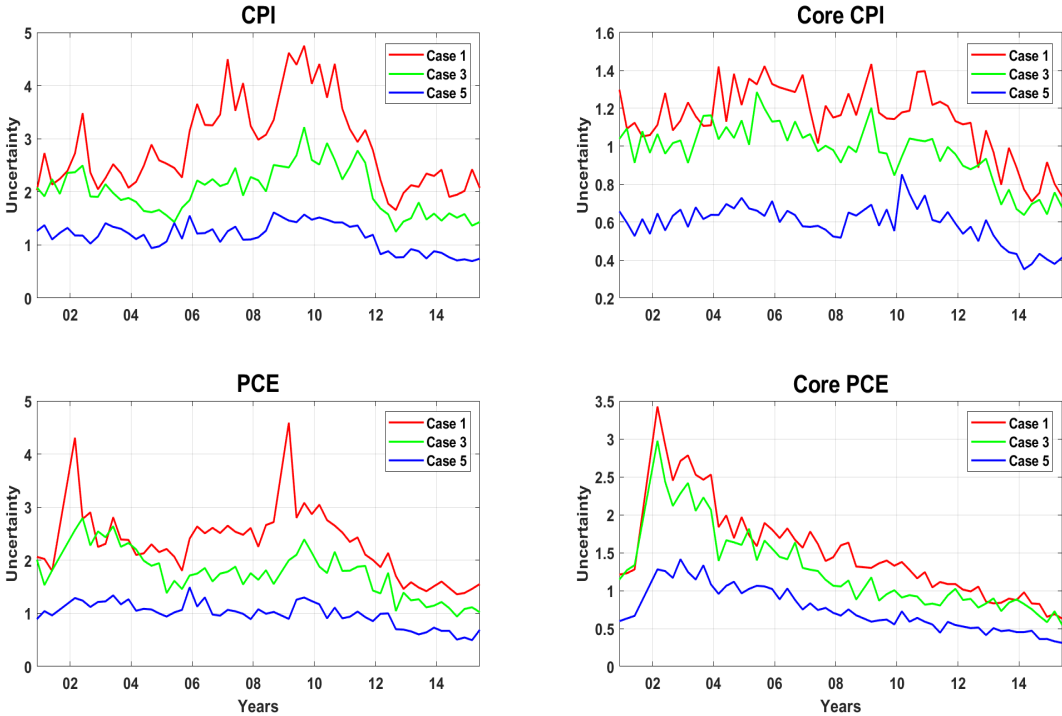
Notes: The figure plots the evolution of the weights for of underlying individual candidate densities for the stage 1 DFM combination at case 4. Each color shade represents a particular individual candidate density. There are 12 candidate densities. The sample period spans September 2000 through June 2015.

Figure B12: Weights for Stage 1 MIDAS Combinations, Log Score Weighting Scheme



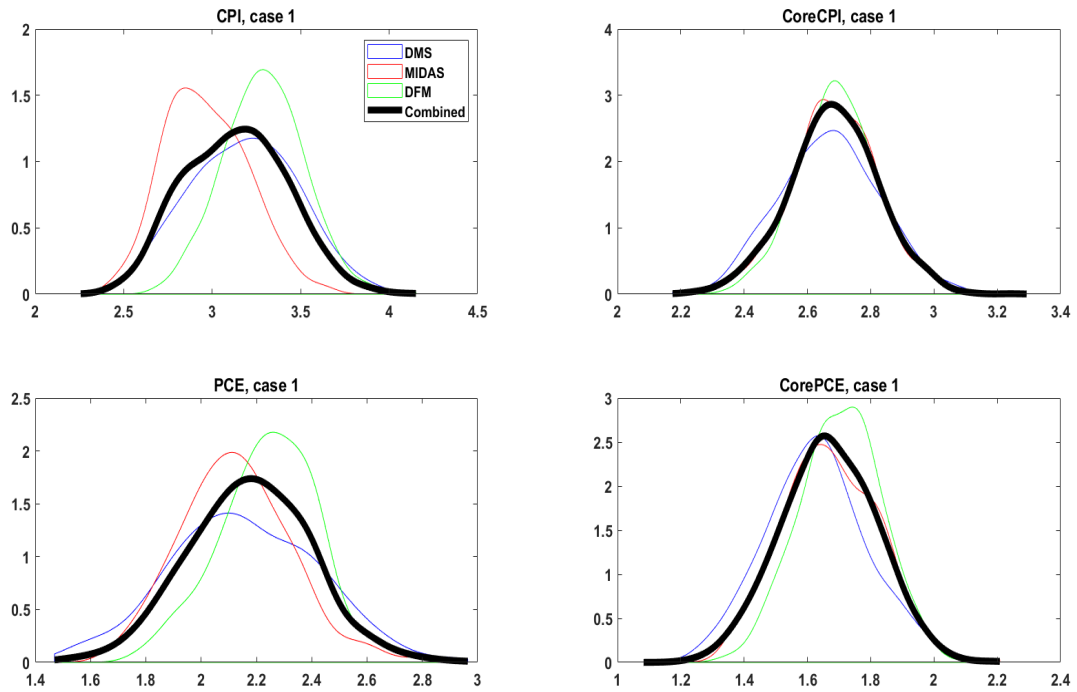
Notes: The figure plots the evolution of the weights for underlying individual candidate densities for the stage 1 MIDAS combination at case 4. Each color shade represents a particular individual candidate density. There are 12 candidate densities. The sample period spans September 2000 through June 2015.

Figure B13: Time-Varying Uncertainty Estimates for Density Nowcasts of Quarterly Inflation



Notes: Uncertainty is measured as the width of the 70% prediction intervals. Estimates are for the grand combination based on the flexible aggregation strategy and log score weighting scheme for case 1 (last day of the preceding quarter), case 3 (last day of the first month of the quarter), and case 5 (last day of the second month of the quarter); see Table B1. The sample period spans 2000Q4 through 2015Q2.

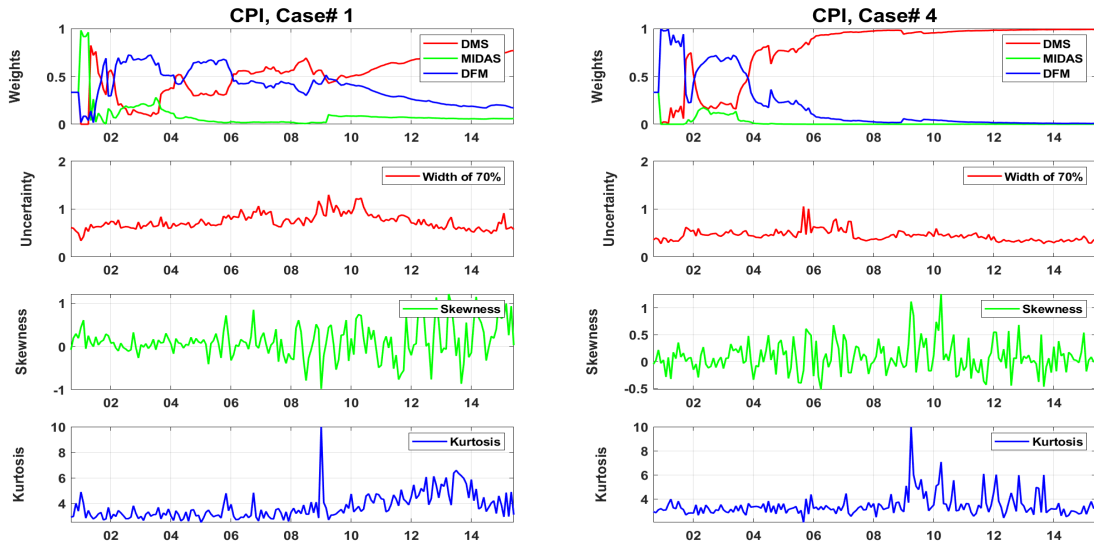
Figure B14: Stage 2 Grand Combination of DMS, DFM, and MIDAS Combinations



Notes: The figure illustrates a grand combination for 12-month inflation rates as of case 1 (the last day of the previous month) for nowcasting the target month of January 2001 and the three stage 1 combinations from the DMS, MIDAS, and DFM model classes that are used to construct the grand combination.

Figure B15: Weights and Higher-Order Moments, CMG Weighting Scheme

(a) CPI inflation



(b) Core CPI inflation

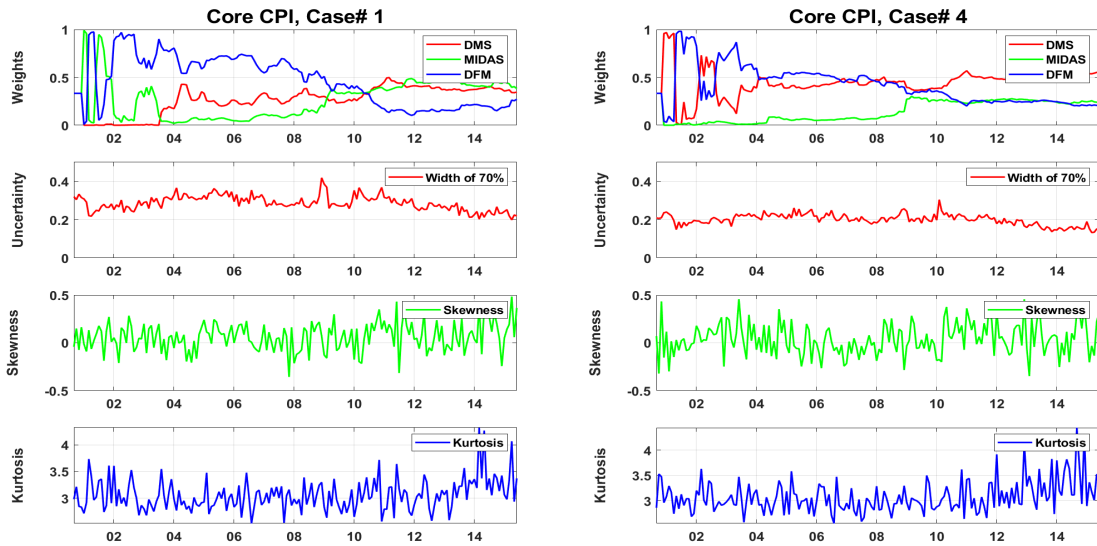
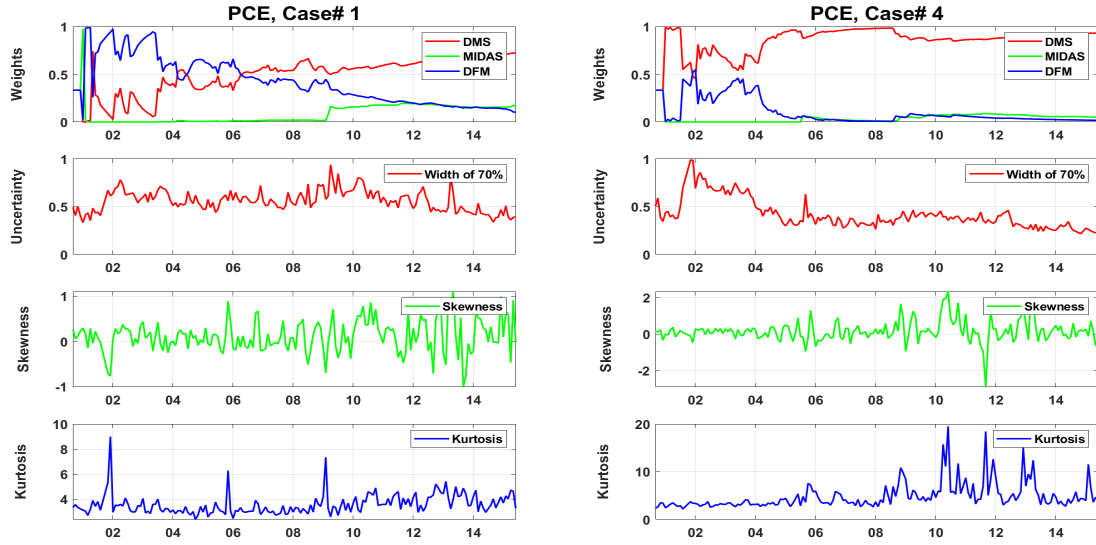
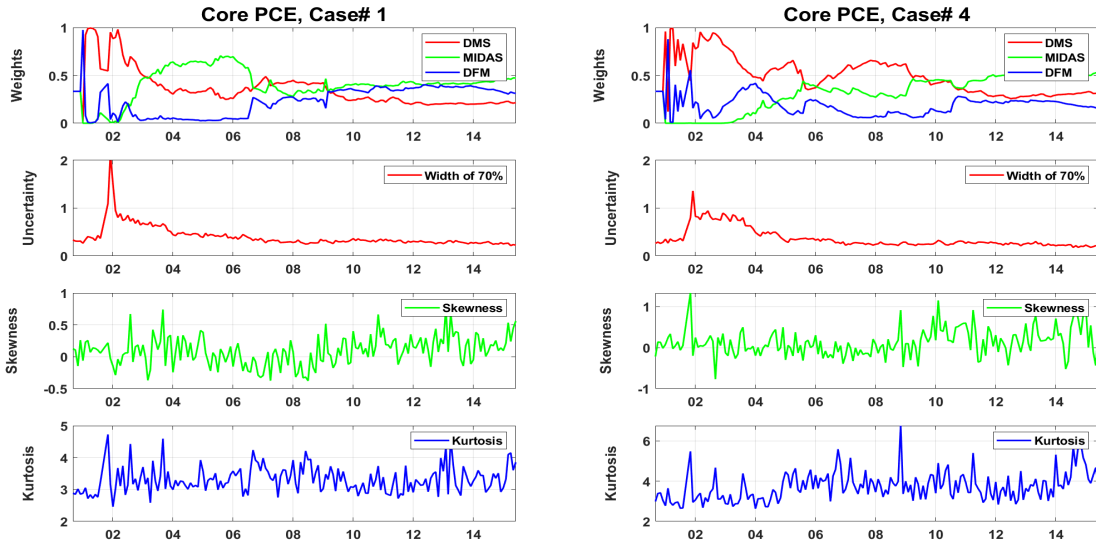


Figure B15: Weights and Higher-Order Moments, CMG Weighting Scheme (continued)

(c) PCE inflation



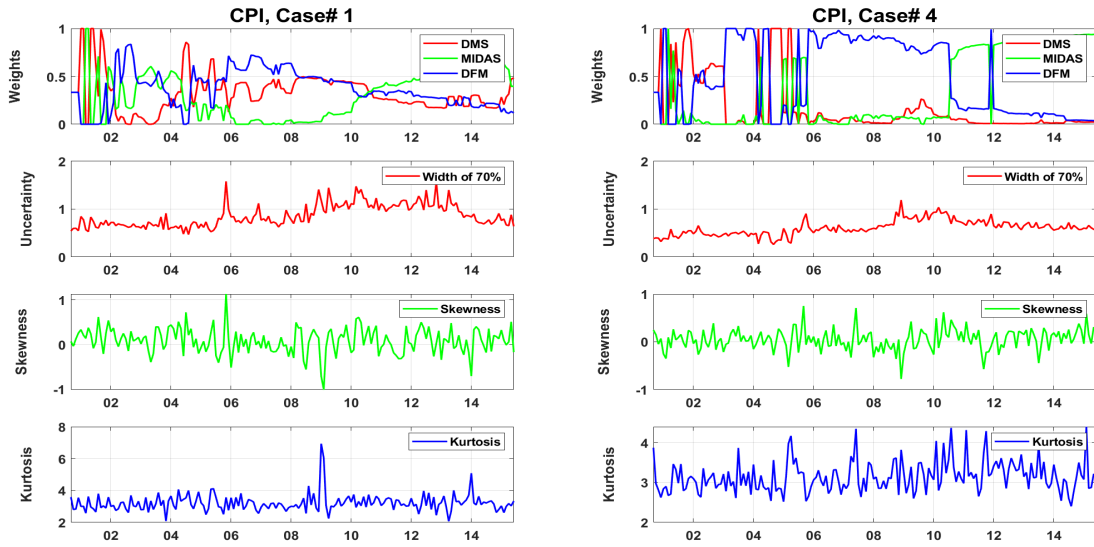
(d) Core PCE inflation



Notes: The first row of each panel plots the evolution of the weights for the three model classes underlying the grand combination, based on the flexible aggregation strategy and CMG weighting scheme. (Each model class is a combination of multiple model specifications.) The second row plots estimates of dynamic uncertainty, defined as the width of the 70% prediction intervals. The last two rows plot time-varying estimates of skewness and kurtosis. The sample period spans September 2000 through June 2015.

Figure B16: Weights and Higher-Order Moments, Ganics Weighting Scheme

(a) CPI inflation



(b) Core CPI inflation

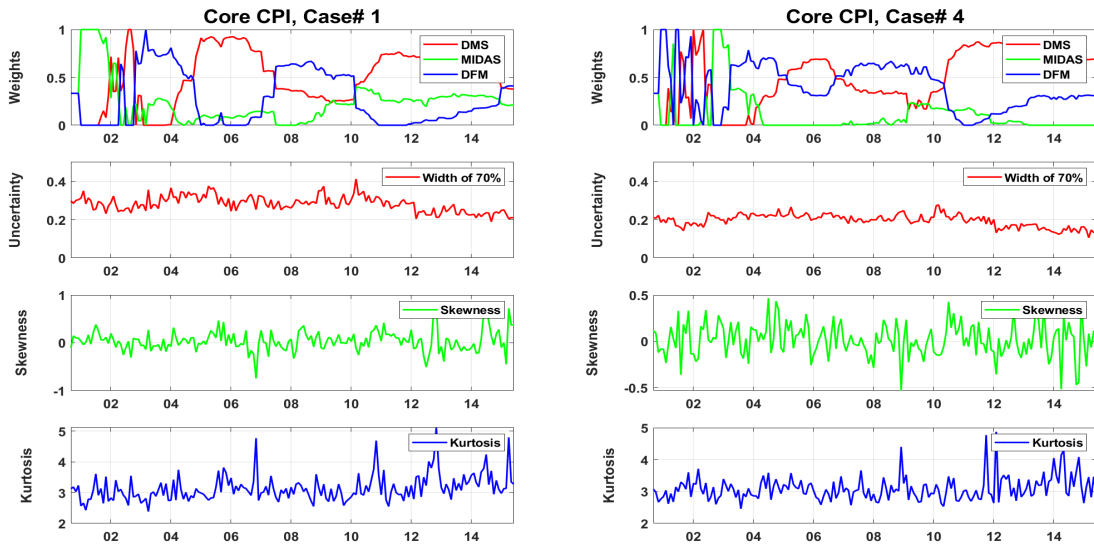
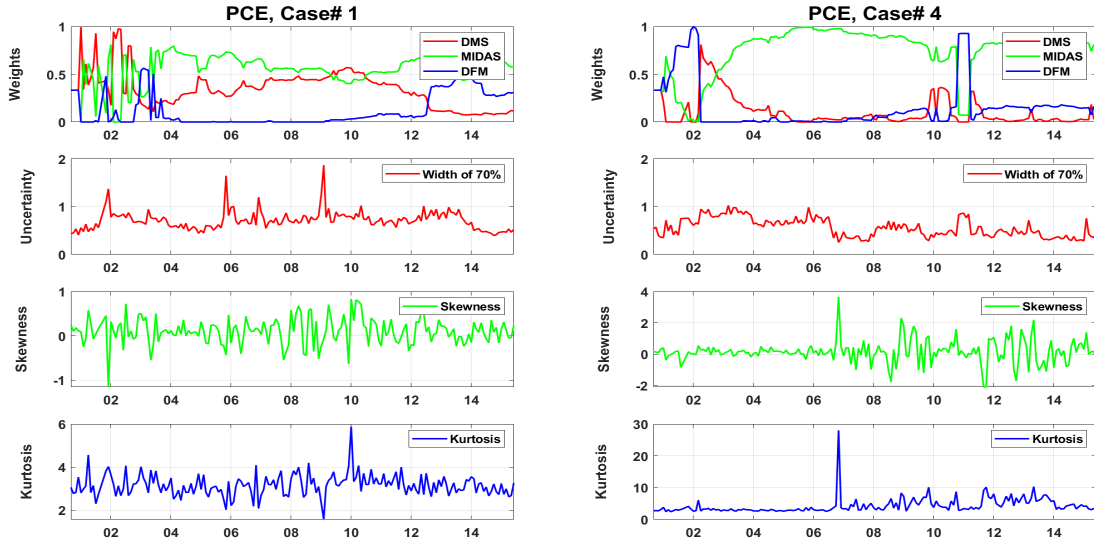
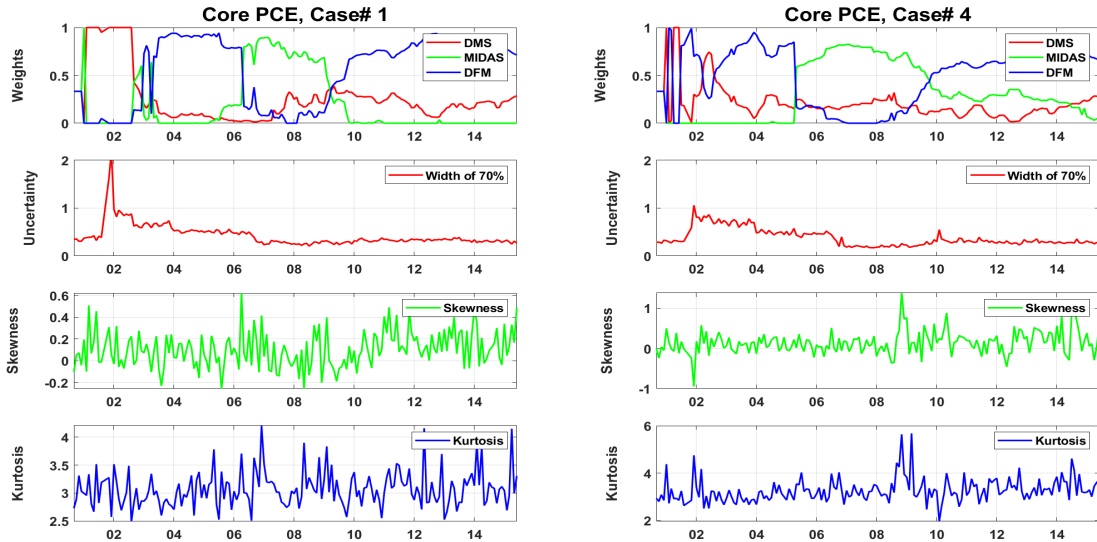


Figure B16: Weights and Higher-Order Moments, Ganics Weighting Scheme (continued)

(c) PCE inflation



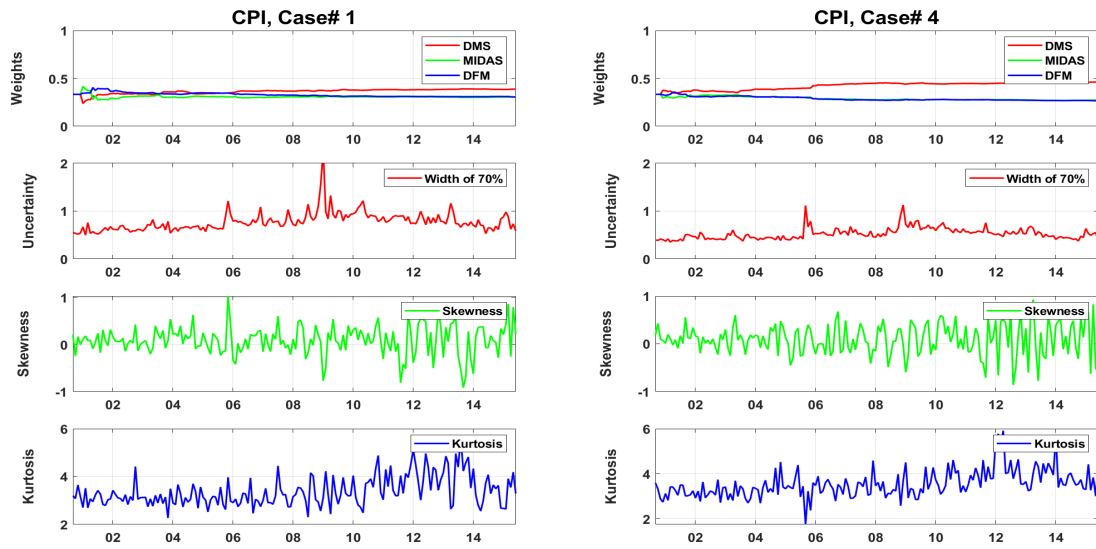
(d) Core PCE inflation



Notes: The first row of each panel plots the evolution of the weights for the three model classes underlying the grand combination, based on the flexible aggregation strategy and the Ganics weighting scheme. (Each model class is a combination of multiple model specifications.) The second row plots estimates of dynamic uncertainty, defined as the width of the 70% prediction intervals. The last two rows plot time-varying estimates of skewness and kurtosis. The sample period spans September 2000 through June 2015.

Figure B17: Weights and Higher-Order Moments, CRPS Weighting Scheme

(a) CPI inflation



(b) Core CPI inflation

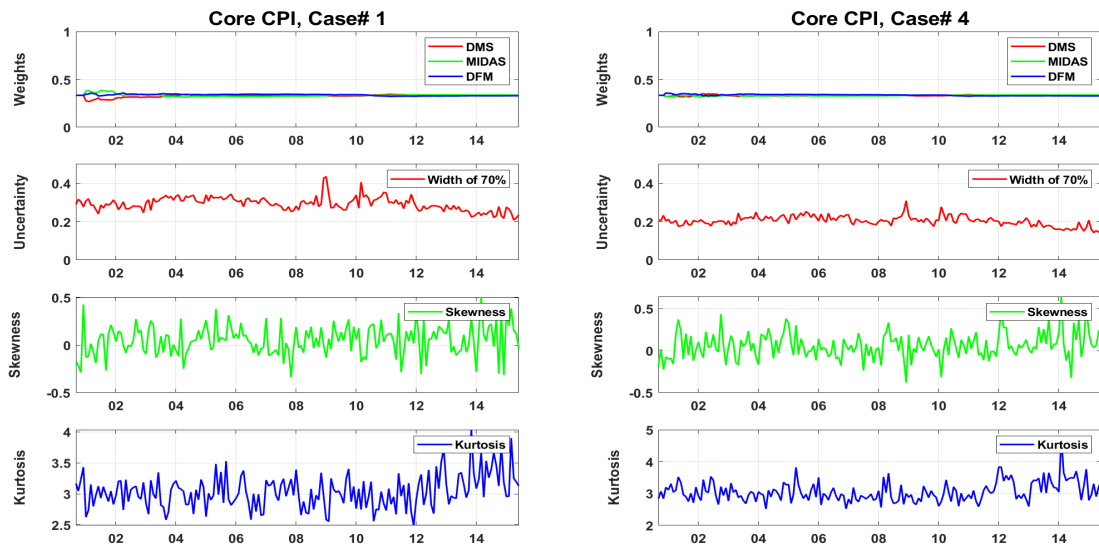
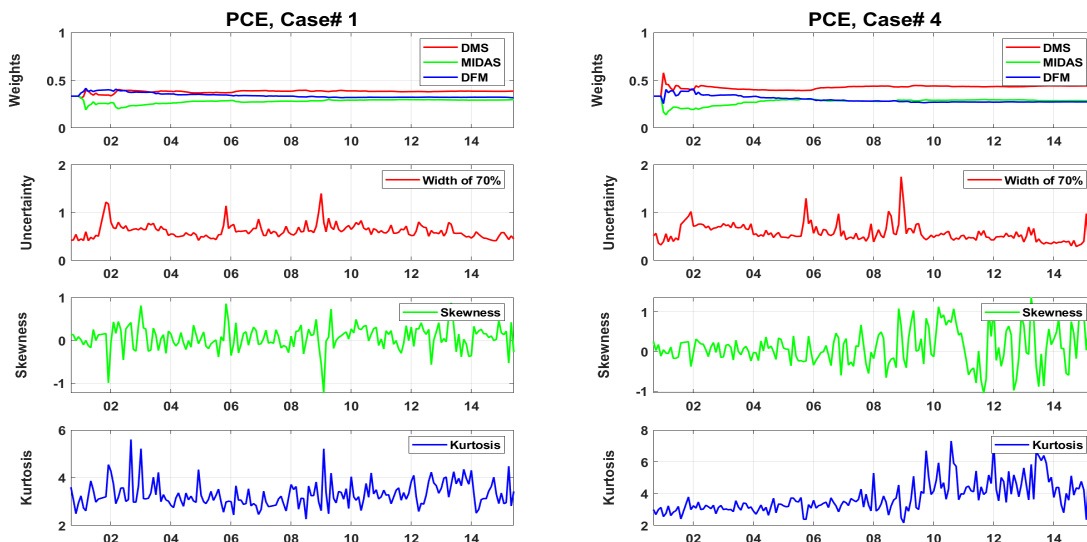
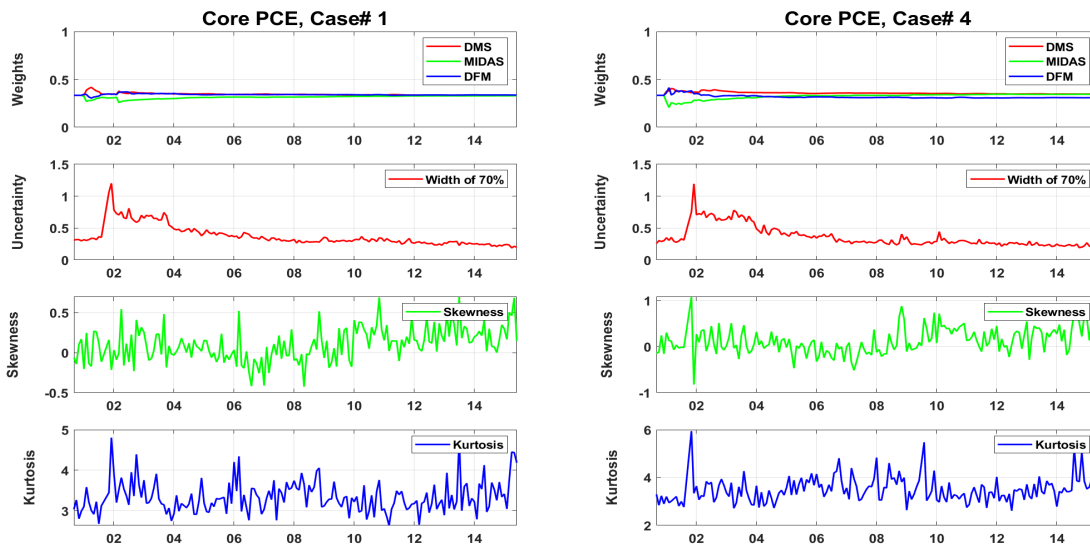


Figure B17: Weights and Higher-Order Moments, CRPS Weighting Scheme (continued)

(c) PCE inflation

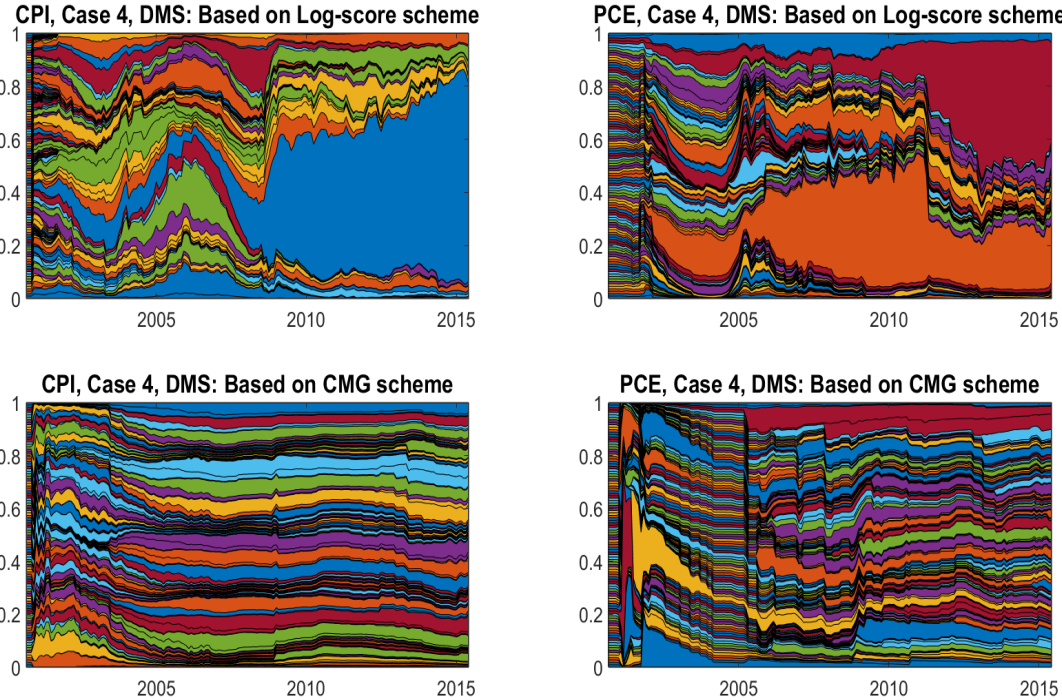


(d) Core PCE inflation



Notes: The first row of each panel plots the evolution of the weights for the three model classes underlying the grand combination, based on the flexible aggregation strategy and the CRPS weighting scheme. (Each model class is a combination of multiple model specifications.) The second row plots estimates of dynamic uncertainty, defined as the width of the 70% prediction intervals. The last two rows plot time-varying estimates of skewness and kurtosis. The sample period spans September 2000 through June 2015.

Figure B18: Comparison of Weights within the DMS Model Class, Log Score Weighting Scheme vs. CMG Weighting Scheme



Notes: The figure plots the evolution of weights of the underlying individual candidate densities. Each color shade represents a particular individual candidate density. There are 108 candidate densities. The richness in the color variation indicates that no single candidate density dominates others. The left panel displays the weights for the stage 1 DMS combination constructed using the log score weighting scheme, and the right panel displays weights for the stage 1 DMS combination constructed using the CMG weighting scheme. The flexible aggregation method is used in both cases. The sample period spans September 2000 through June 2015.

Table B1: Representative Dates for Quarterly Nowcasting Performance

Case	Date	Information Set (Example: Nowcasting target quarter is Q1)	Months to Forecast
1	Last day of the previous month	December 31: Have CPI and PCE through November; high-frequency information through December 31	CPI: h=4 (Dec., Jan., Feb., Mar.) PCE: h=4 (Dec., Jan., Feb., Mar.)
2	Day 15 of month 1 of the target quarter	January 15: Receive CPI for December and have PCE through November; high-frequency information through end of second week of January, which includes two weekly retail gasoline readings from January	CPI: h=3 (Jan., Feb., Mar.) PCE: h=4 (Dec., Jan., Feb., Mar.)
3	Last day of month 1 of the target quarter	January 31: Have CPI for December and receive PCE for December; high-frequency information for all of January, which includes all four weekly retail gasoline readings from January	CPI: h=3 (Jan., Feb., Mar.) PCE: h=3 (Jan., Feb., Mar.)
4	Day 15 of month 2 of the target quarter	February 15: Receive CPI for January and have PCE through December; high-frequency information through end of second week of February, which includes two weekly retail gasoline readings from February	CPI: h=2 (Feb., Mar.) PCE: h=3 (Jan., Feb., Mar.)
5	Last day of month 2 of the target quarter	February 28: Have CPI for January and receive PCE for January; high-frequency information for all of February, which includes all four weekly retail gasoline readings from February	CPI: h=2 (Feb., Mar.) PCE: h=2 (Feb., Mar.)
6	Day 15 of month 3 of the target quarter	March 15: Receive CPI for February and have PCE through January; high-frequency information through end of second week of March, which includes two weekly retail gasoline readings from March	CPI: h=1 (Mar.) PCE: h=2 (Feb., Mar.)
7	Last day of month 3 of the target quarter	March 31: Have CPI for February and receive PCE for February; high-frequency information for all of March, which includes all four weekly retail gasoline readings from March	CPI: h=1 (Mar.) PCE: h=1 (Mar.)

Appendix C

Chapter 4 Appendix

Chapter 4: A Unified Framework to Estimate Macroeconomic Stars.

C1. Bayesian Estimation Details

C1.a. Base Model equations

For convenience, we list all model equations keeping the numbering as in the main text.

$$U_t = U_t^* + U_t^c \quad (\text{C.6})$$

$$U_t - U_t^* = \rho_1^u(U_{t-1} - U_{t-1}^*) + \rho_2^u(U_{t-2} - U_{t-2}^*) + \phi^u \text{ogap}_t + \varepsilon_t^u, \quad \varepsilon_t^u \sim N(0, \sigma_u^2) \quad (\text{C.7})$$

where, $\rho_1^u + \rho_2^u < 1$, $\rho_2^u - \rho_1^u < 1$, and $|\rho_2^u| < 1$; $\phi^u < 0$

$$U_t^* = U_{t-1}^* + \varepsilon_t^{u*}, \quad \varepsilon_t^{u*} \sim TN(a_u - U_{t-1}^*, b_u - U_{t-1}^*; 0, \sigma_{u*}^2) \quad (\text{C.8})$$

$$Z_t^u = C_t^u + \beta^u U_t^* + \varepsilon_t^{zu}, \quad \varepsilon_t^{zu} \sim N(0, \sigma_{zu}^2) \quad (\text{C.9})$$

$$C_t^u = C_{t-1}^u + \varepsilon_t^{cu}, \quad \varepsilon_t^{cu} \sim N(0, \sigma_{cu}^2) \quad (\text{C.10})$$

$$\text{gdp}_t = \text{gdp}_t^* + \text{ogap}_t \quad (\text{C.11})$$

$$\text{gdp}_t^* = 2\text{gdp}_{t-1}^* - \text{gdp}_{t-2}^* + \varepsilon_t^{\text{gdp}*}, \quad \varepsilon_t^{\text{gdp}*} \sim N(0, \sigma_{\text{gdp}*}^2) \quad (\text{C.12})$$

$$g_t^* \equiv \Delta \text{gdp}_t^*$$

$$g_t^* = g_{t-1}^* + \varepsilon_t^{\text{gdp}*} \quad (\text{C.13})$$

$$\text{ogap}_t = \rho_1^g(\text{ogap}_{t-1}) + \rho_2^g(\text{ogap}_{t-2}) + a^r(r_t - r_t^*) + \lambda^g(U_t - U_t^*) + \varepsilon_t^{\text{ogap}} \quad (\text{C.14})$$

where, $\varepsilon_t^{ogap} \sim N(0, \sigma_{ogap}^2)$, $\rho_1^g + \rho_2^g < 1$, $\rho_2^g - \rho_1^g < 1$, and $|\rho_2^g| < 1$; $\lambda^g < 0$

$$Z_t^g = C_t^g + \beta^g * 4 * g_t^* + \varepsilon_t^{zg}, \quad \varepsilon_t^{zg} \sim N(0, \sigma_{zg}^2) \quad (C.15)$$

$$C_t^g = C_{t-1}^g + \varepsilon_t^{cg}, \quad \varepsilon_t^{cg} \sim N(0, \sigma_{cg}^2) \quad (C.16)$$

$$P_t - P_t^* = \rho^p(P_{t-1} - P_{t-1}^*) + \lambda_t^p(U_t - U_t^*) + \varepsilon_t^p, \quad \varepsilon_t^p \sim N(0, e^{h_t^p}) \quad (C.17)$$

where, $|\rho^p| < 1$

$$\lambda_t^p = \lambda_{t-1}^p + \varepsilon_t^{\lambda p}, \quad \varepsilon_t^{\lambda p} \sim N(0, \sigma_{\lambda p}^2) \quad (C.18)$$

$$h_t^p = h_{t-1}^p + \varepsilon_t^{hp}, \quad \varepsilon_t^{hp} \sim N(0, \sigma_{hp}^2) \quad (C.19)$$

$$P_t^* = P_{t-1}^* + \varepsilon_t^{p*}, \quad \varepsilon_t^{p*} \sim N(0, \sigma_{p*}^2) \quad (C.20)$$

$$\pi_t - \pi_t^* = \rho_t^\pi(\pi_{t-1} - \pi_{t-1}^*) + \lambda_t^\pi(U_t - U_t^*) + \varepsilon_t^\pi, \quad \varepsilon_t^\pi \sim N(0, e^{h_t^\pi}) \quad (C.21)$$

$$\rho_t^\pi = \rho_{t-1}^\pi + \varepsilon_t^{\rho\pi}, \quad \varepsilon_t^{\rho\pi} \sim TN(0 - \rho_{t-1}^\pi, 1 - \rho_{t-1}^\pi; 0, \sigma_{\rho\pi}^2) \quad (C.22)$$

where, ρ^π are truncated so that $0 < \rho_t^\pi < 1$.

$$\lambda_t^\pi = \lambda_{t-1}^\pi + \varepsilon_t^{\lambda\pi}, \quad \varepsilon_t^{\lambda\pi} \sim TN(-1 - \lambda_{t-1}^\pi, 0 - \lambda_{t-1}^\pi; 0, \sigma_{\lambda\pi}^2) \quad (C.23)$$

λ^π is the slope of price Phillips curve and is constrained in the interval $(-1, 0)$.

$$h_t^\pi = h_{t-1}^\pi + \varepsilon_t^{h\pi}, \quad \varepsilon_t^{h\pi} \sim N(0, \sigma_{h\pi}^2) \quad (C.24)$$

$$\pi_t^* = \pi_{t-1}^* + \varepsilon_t^{\pi*}, \quad \varepsilon_t^{\pi*} \sim N(0, \sigma_{\pi*}^2) \quad (C.25)$$

$$Z_t^\pi = C_t^\pi + \beta^\pi \pi_t^* + \varepsilon_t^{z\pi}, \quad \varepsilon_t^{z\pi} \sim N(0, \sigma_{z\pi}^2) \quad (C.26)$$

$$C_t^\pi = C_{t-1}^\pi + \varepsilon_t^{c\pi}, \quad \varepsilon_t^{c\pi} \sim N(0, \sigma_{c\pi}^2) \quad (C.27)$$

$$W_t^* = \pi_t^* + P_t^* + \varepsilon_t^{w*}, \quad \varepsilon_t^{w*} \sim N(0, \sigma_{w*}^2) \quad (C.28)$$

$$W_t - W_t^* = \rho_t^w(W_{t-1} - W_{t-1}^*) + \lambda_t^w(U_t - U_t^*) + \kappa_t^w(\pi_t - \pi_t^*) + \varepsilon_t^w, \quad \varepsilon_t^w \sim N(0, e^{h_t^w}) \quad (C.29)$$

$$h_t^w = h_{t-1}^w + \varepsilon_t^{hw}, \quad \varepsilon_t^{hw} \sim N(0, \sigma_{hw}^2) \quad (C.30)$$

$$\rho_t^w = \rho_{t-1}^w + \varepsilon_t^{\rho w}, \quad \varepsilon_t^{\rho w} \sim TN(0 - \rho_{t-1}^w, 1 - \rho_{t-1}^w; 0, \sigma_{\rho w}^2) \quad (C.31)$$

$$\lambda_t^w = \lambda_{t-1}^w + \varepsilon_t^{\lambda w}, \quad \varepsilon_t^{\lambda w} \sim TN(-1 - \lambda_{t-1}^w, 0 - \lambda_{t-1}^w; 0, \sigma_{\lambda w}^2) \quad (C.32)$$

λ^w is the slope of wage Phillips curve and is constrained in the interval $(-1, 0)$.

$$\kappa_t^w = \kappa_{t-1}^w + \varepsilon_t^{\kappa w}, \quad \varepsilon_t^{\kappa w} \sim N(0, \sigma_{\kappa w}^2) \quad (C.33)$$

$$i_t - \pi_t^* - r_t^* = \rho^i(i_{t-1} - \pi_{t-1}^* - r_{t-1}^*) + \lambda^i(U_t - U_t^*) + \kappa^i(\pi_t - \pi_t^*) + \varepsilon_t^i, \quad \varepsilon_t^i \sim N(0, e^{h_t^i}) \quad (C.34)$$

where, ρ^i is truncated so that $0 < \rho^i < 1$.

$$h_t^i = h_{t-1}^i + \varepsilon_t^{hi}, \quad \varepsilon_t^{hi} \sim N(0, \sigma_{hi}^2) \quad (\text{C.35})$$

$$r_t^* = \zeta g_t^* + D_t. \quad (\text{C.36})$$

$$D_t = D_{t-1} + \varepsilon_t^d, \quad \varepsilon_t^d \sim N(0, \sigma_d^2) \quad (\text{C.37})$$

$$Z_t^r = C_t^r + \beta^r r_t^* + \varepsilon_t^{zr}, \quad \varepsilon_t^{zr} \sim N(0, \sigma_{zr}^2) \quad (\text{C.38})$$

$$C_t^r = C_{t-1}^r + \varepsilon_t^{cr}, \quad \varepsilon_t^{cr} \sim N(0, \sigma_{cr}^2) \quad (\text{C.39})$$

C1.b. Prior Elicitation

Our prior settings are similar to those used in Chan, Koop, and Potter (2016) [CKP], Chan, Clark, and Koop (2018) [CCK], and Gonzalez-Astudillo and Laforte (2020). As discussed in CCK, UC models with several unobserved variables, such as the one developed in this chapter, require informative priors. That said, our priors settings for most variables are only slightly informative. The use of inequality restrictions on some parameters such as the Phillips curve, persistence, bounds on u^* could be viewed as additional sources of information that eliminates the need for tight priors, something also noted by CKP. The parameters for which there is a strong agreement in the empirical literature on their values, such as the Taylor-rule equation parameters, we use relatively tight priors, such that prior distributions are centered on prior means with small variance.

In the table below, the notation $N(a, b)$ denotes Normal distribution with mean a , and variance b ; and $IG(\nu, S)$ denotes Inverse Gamma distribution with degrees of freedom parameter ν , and scale parameter S .

Table C.1: Prior Settings

Parameter	Parameter Description	Prior
a^r	Coefficient on interest-rate gap in output gap equation	$N(0, 1)$
ρ_1^g	Persistence in output gap: lag 1	$N(1.3, 0.1^2)$
ρ_2^g	Persistence in output gap: lag 2	$N(-0.5, 0.1^2)$
ρ_1^u	Persistence in UR gap: lag 1	$N(1.3, 0.1^2)$
ρ_2^u	Persistence in UR gap: lag 2	$N(-0.5, 0.1^2)$
ρ^p	Persistence in productivity gap	$N(0.1, 1)$
ζ	Relationship between r^* and g^*	$N(1, 0.1)$
ρ^i	Persistence in interest-rate gap	$N(0.85, 0.1^2)$
λ^i	Interest rate sensitivity to UR gap: $(-2 * (1 - \rho^i))$	$N(-0.3, 0.1^2)$
κ^i	Interest rate sensitivity to inflation: $(0.5 * (1 - \rho^i))$	$N(0.075, 0.1^2)$
λ^g	Output gap response to UR gap	$N(-0.02, 1)$
ϕ^u	UR gap response to Output gap	$N(-0.02, 1)$
β^g	Link between g^* and survey	$N(1, 0.1^2)$
β^u	Link between u^* and survey	$N(1, 0.05^2)$
β^r	Link between r^* and survey	$N(1, 0.1^2)$
β^π	Link between π^* and survey	$N(1, 0.05^2)$
$\sigma_{\pi^*}^2$	Var. of the shocks to π^*	$IG(10, 0.1^2 \times 9)$
$\sigma_{p^*}^2$	Var. of the shocks to p^*	$IG(10, 0.142^2 \times 9)$
$\sigma_{u^*}^2$	Var. of the shocks to u^*	$IG(10, 0.1^2 \times 9)$

Continued on next page

Table C1 – continued from previous page

Parameter	Parameter Description	Prior
$\sigma_{gdp^*}^2$	Var. of the shocks to gdp^*	$IG(10, 0.01^2 \times 9)$
σ_d^2	Var. of the shocks to d	$IG(10, 0.1^2 \times 9)$
$\sigma_{w^*}^2$	Var. of the shocks to w^*	$IG(10, 0.142^2 \times 9)$
σ_{ogap}^2	Var. of the shocks to Ogap	$IG(10, 1 \times 9)$
σ_u^2	Var. of the shocks to UR gap	$IG(10, 0.707^2 \times 9)$
σ_{hp}^2	Var. of the volatility – Productivity eq.	$IG(10, 0.316^2 \times 9)$
σ_h^2	Var. of the volatility – Price Inf. eq.	$IG(10, 0.316^2 \times 9)$
σ_{hw}^2	Var. of the volatility – Wage Inf. eq.	$IG(10, 0.316^2 \times 9)$
σ_{hi}^2	Var. of the volatility – Interest rate eq.	$IG(10, 0.316^2 \times 9)$
$\sigma_{\lambda^\pi}^2$	Var. of the shocks to TVP λ^π , Price Phillips curve	$IG(10, 0.04^2 \times 9)$
$\sigma_{\lambda^w}^2$	Var. of the shocks to TVP λ^w , Wage Phillips curve	$IG(10, 0.04^2 \times 9)$
$\sigma_{\lambda^p}^2$	Var. of the shocks to TVP λ^p , Cyc. Productivity	$IG(10, 0.04^2 \times 9)$
$\sigma_{\kappa^w}^2$	Var. of the shocks to TVP κ^w , PT: π to Wages	$IG(10, 0.04^2 \times 9)$
$\sigma_{\rho^w}^2$	Var. of the shocks to TVP ρ^w , Persist. Wage-gap	$IG(10, 0.04^2 \times 9)$
$\sigma_{\rho^\pi}^2$	Var. of the shocks to TVP ρ^π , Persist. Inflation-gap	$IG(10, 0.04^2 \times 9)$
C_0^π	Time-varying Intercept in eq. linking survey to pi-star	$N(0, 0.1)$
C_0^u	Time-varying Intercept in eq. linking survey to u-star	$N(0, 0.1)$
C_0^g	Time-varying Intercept in eq. linking survey to g-star	$N(0, 0.1)$
C_0^r	Time-varying Intercept in eq. linking survey to r-star	$N(0, 0.1)$
$\sigma_{c^\pi}^2$	Var. of the shocks to TVP C^π	$IG(10, 0.1^2 \times 9)$
$\sigma_{c^u}^2$	Var. of the shocks to TVP C^u	$IG(10, 0.1^2 \times 9)$
$\sigma_{c^g}^2$	Var. of the shocks to TVP C^g	$IG(10, 0.1^2 \times 9)$
$\sigma_{c^r}^2$	Var. of the shocks to TVP C^r	$IG(10, 0.1^2 \times 9)$
$\sigma_{z^\pi}^2$	Var. of the shocks in measurement eq. Z^π ,	$IG(10, 0.2 \times 9)$
$\sigma_{z^u}^2$	Var. of the shocks in measurement eq. Z^u ,	$IG(10, 0.3 \times 9)$
$\sigma_{z^g}^2$	Var. of the shocks in measurement eq. Z^g ,	$IG(10, 0.1 \times 9)$
$\sigma_{z^r}^2$	Var. of the shocks in measurement eq. Z^r ,	$IG(10, 0.2 \times 9)$
π_0^*	Initial value of pi-star	$N(3, 5^2)$
u_0^*	Initial value of u-star, $t = 0$	$N(5, 5^2)$
u_{-1}^*	Initial value of u-star, $t = -1$	$N(5, 5^2)$
p_0^*	Initial value of p-star	$N(3, 5^2)$
w_0^*	Initial value of w-star, $E(p_0^*) + E(\pi_0^*) = 6$	$N(6, 5^2)$
D_0	Initial value of D, "catch-all" component of r-star	$N(0, 0.3162^2)$
gdp_0^*	Initial value of gdp-star, $t = 0$	$N(750, 10^2)$
gdp_{-1}^*	Initial value of gdp-star, $t = -1$	$N(750, 10^2)$

C1.c. MCMC Algorithm

The estimation of our complex UC model and sampling from its joint posterior distribution reduces to sequentially drawing from a set of conditional posterior densities, some of which are standard and some that are non-standard.

Collect all the time-invariant model parameters into θ :

$$\theta = (\rho_1^u, \rho_2^u, \sigma_u^2, \phi_u, \sigma_{u*}^2, \beta^u, \sigma_{zu}^2, \sigma_{cu}^2, \sigma_{gdp*}^2, \rho_1^g, \rho_2^g, a^r, \lambda^g, \sigma_{ogap}^2, \sigma_{zg}^2, \sigma_{cg}^2, \beta^g, \rho^p, \sigma_{hp}^2, \sigma_{p*}^2, \sigma_{\lambda\pi}^2, \dots, \sigma_{\rho\pi}^2, \sigma_{h\pi}^2, \sigma_{\pi*}^2, \sigma_{z\pi}^2, \sigma_{c\pi}^2, \beta^\pi, \sigma_{w*}^2, \sigma_{hw}^2, \sigma_{\rho w}^2, \sigma_{\lambda w}^2, \sigma_{\kappa w}^2, \rho^i, \lambda^i, \kappa^i, \sigma_{hi}^2, \sigma_{zr}^2, \sigma_{cr}^2, \beta^r, \sigma_d^2)$$

We denote \bullet as representing all other model parameters.

1. $p(U^*|Y, \bullet)$
2. $p(gdp^*|Y, \bullet)$
3. $p(P^*|Y, \bullet)$
4. $p(\pi^*|Y, \bullet)$
5. $p(w^*|Y, \bullet)$
6. $p(r^*|Y, \bullet)$
7. $p(\lambda^p|Y, \bullet)$
8. $p(\rho^\pi|Y, \bullet)$
9. $p(\lambda^\pi|Y, \bullet)$
10. $p(\rho^w|Y, \bullet)$
11. $p(\lambda^w|Y, \bullet)$
12. $p(\kappa^w|Y, \bullet)$
13. $p(h^p, h^\pi, h^w, h^i|Y, \bullet)$
14. $p(C^u, C^g, C^\pi, C^r|Y, \bullet)$
15. $p(D|Y, \bullet)$
16. $p(\theta|Y, \bullet)$

Step 1. Derive the conditional distribution $p(U^*|Y, \bullet)$

The derivation of this distribution is most complex because the information about U^* comes from eight sources (i.e., model equations). Below, we derive an expression for each of the eight sources.

The first source is the state equation of U^* . We rewrite it in a matrix notation as follows,

$$HU^* = \alpha_u + \varepsilon^{u*} \quad \varepsilon^{u*} \sim N(0, \Omega_{u*}), \quad \text{where } \Omega_{u*} = \text{diag}(\omega_{u*}^2, \sigma_{u*}^2, \dots, \sigma_{u*}^2) \quad (\text{C.40})$$

where,

$$\alpha_u = \begin{pmatrix} U_0^* \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{pmatrix}$$

That is, the prior density for U^* is given by

$$p(U^*|\sigma_{U^*}^2) \propto -\frac{1}{2}(U^* - H^{-1}\alpha_u)' H' \Omega_{u*}^{-1} H (U^* - H^{-1}\alpha_u) + g_{u*}(U^*, \sigma_{u*}^2)$$

where,

$a_u < U^* < b_u$ for $t = 1, \dots, T$, and

$$g_{u^*}(U^*, \sigma_{u^*}^2) = -\log \left(\Phi \left(\frac{b_u}{\omega_{u^*}} \right) - \Phi \left(\frac{a_u}{\omega_{u^*}} \right) \right) - \sum_{t=2}^T \log \left(\Phi \left(\frac{b_u - U_{t-1}^*}{\sigma_{u^*}} \right) - \Phi \left(\frac{a_u - U_{t-1}^*}{\sigma_{u^*}} \right) \right)$$

The second source of information comes from the unemployment measurement equation. Rewrite the equation in a matrix notation,

$$K_u U = \mu^u + K_u U^* + \varepsilon^u \quad \varepsilon^u \sim N(0, \Omega_u), \quad \text{where } \Omega_u = I_T \otimes \sigma_u^2 \quad (\text{C.41})$$

and,

$$\mu_u = \begin{pmatrix} \rho_1^u(U_0 - U_0^*) + \rho_2^u(U_{-1} - U_{-1}^*) \\ \rho_2^u(U_0 - U_0^*) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad K_u = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho_1^u & 1 & 0 & \cdots & 0 \\ -\rho_2^u & -\rho_1^u & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & -\rho_2^u & -\rho_1^u & 1 \end{pmatrix}$$

Ignoring any terms not involving U^* , we have

$$\log p(U|U^*, \bullet) \propto -\frac{1}{2}(U - K_u^{-1}\mu_u - U^*)' K_u' \Omega_u^{-1} K_u (U - K_u^{-1}\mu_u - U^*)$$

The third source of information comes from the inflation measurement equation. Rewrite the equation in a matrix notation,

$$Z = \Lambda^\pi U^* + \varepsilon^\pi \quad \varepsilon^\pi \sim N(0, \Omega_\pi), \quad \text{where } \Omega_\pi = \text{diag}(e^{h_1^\pi}, e^{h_2^\pi}, \dots, e^{h_T^\pi}) \quad (\text{C.42})$$

where,

$$z_t = (\pi_t - \pi_t^*) - \rho_t^\pi(\pi_{t-1} - \pi_{t-1}^*) - \lambda_t^\pi U_t,$$

$$Z = (z_1, \dots, z_T)' \text{ and } \Lambda^\pi = \text{diag}(-\lambda_1^\pi, \dots, -\lambda_T^\pi)$$

Ignoring any terms not involving U^* , we have

$$\log p(\pi|U^*, U, \pi^*, h^\pi, \rho^\pi, \bullet) \propto -\frac{1}{2}(Z - \Lambda^\pi U^*)' \Omega_\pi^{-1} (Z - \Lambda^\pi U^*)$$

The fourth source of information comes from the productivity measurement equation. Rewrite

the equation in a matrix notation,

$$M^P = \Lambda^P U^* + \varepsilon^P \quad \varepsilon^P \sim N(0, \Omega_P), \quad \text{where } \Omega_P = \text{diag}(e^{h_1^P}, e^{h_2^P}, \dots, e^{h_T^P}) \quad (\text{C.43})$$

where,

$$m_t = (P_t - P_t^*) - \rho^P (P_{t-1} - P_{t-1}^*) - \lambda_t^P U_t,$$

$$M^P = (m_1, \dots, m_T)' \text{ and } \Lambda^P = \text{diag}(-\lambda_1^P, \dots, -\lambda_T^P)$$

Ignoring any terms not involving U^* , we have

$$\log p(P|U^*, U, P^*, h^P, \rho^P, \bullet) \propto -\frac{1}{2}(M^P - \Lambda^P U^*)' \Omega_P^{-1} (M^P - \Lambda^P U^*)$$

The fifth source of information comes from the wage measurement equation. Rewrite the equation in a matrix notation,

$$M^w = \Lambda^w U^* + \varepsilon^w \quad \varepsilon^w \sim N(0, \Omega_w), \quad \text{where } \Omega_w = \text{diag}(e^{h_1^w}, e^{h_2^w}, \dots, e^{h_T^w}) \quad (\text{C.44})$$

where,

$$m_t^w = (W_t - W_t^*) - \rho_t^W (W_{t-1} - W_{t-1}^*) - \lambda_t^W U_t - \kappa_t^W (\pi_t - \pi_t^*),$$

$$M^w = (m_1^w, \dots, m_T^w)' \text{ and } \Lambda^w = \text{diag}(-\lambda_1^W, \dots, -\lambda_T^W)$$

Ignoring any terms not involving U^* , we have

$$\log p(W|U^*, W, W^*, h^w, \rho^W, \bullet) \propto -\frac{1}{2}(M^w - \Lambda^w U^*)' \Omega_w^{-1} (M^w - \Lambda^w U^*)$$

The sixth source of information comes from the output gap measurement equation. Rewrite the equation in a matrix notation,

$$M^g = \Lambda^g U^* + \varepsilon^g \quad \varepsilon^g \sim N(0, \Omega_{ogap}), \quad \text{where } \Omega_{ogap} = \text{diag}(\sigma_{ogap}^2, \dots, \sigma_{ogap}^2) \quad (\text{C.45})$$

where,

$$m_t^g = ogap_t - \rho_1^g (ogap_{t-1}) - \rho_2^g (ogap_{t-2}) - \lambda^g U_t - a^r (r_t - r_t^*),$$

$$M^g = (m_1^g, \dots, m_T^g)' \text{ and } \Lambda^g = \text{diag}(-\lambda^g, \dots, -\lambda^g)$$

Ignoring any terms not involving U^* , we have

$$\log p(\text{ogap}|U^*, U, \bullet) \propto -\frac{1}{2}(M^g - \Lambda^g U^*)' \Omega_{\text{ogap}}^{-1} (M^g - \Lambda^g U^*)$$

The seventh source of information comes from the Taylor-type rule measurement equation. Rewrite the equation in a matrix notation,

$$M^{ui} = \Gamma^{ui} U^* + \varepsilon^i \quad \varepsilon^i \sim N(0, \Omega_i), \quad \text{where } \Omega_i = \text{diag}(e^{h_1^i}, e^{h_2^i}, \dots, e^{h_T^i}) \quad (\text{C.46})$$

where,

$$m_t^{ui} = i_t - \pi_t^* - r_t^* - \rho^i(i_{t-1} - \pi_{t-1}^* - r_{t-1}^*) - \kappa^i(\pi_t - \pi_t^*) - \lambda^i U_t,$$

$$M^{ui} = (m_1^{ui}, \dots, m_T^{ui})' \text{ and } \Gamma^{ui} = \text{diag}(-\lambda^i, \dots, -\lambda^i)$$

Ignoring any terms not involving U^* , we have

$$\log p(i|U^*, U, \pi, \bullet) \propto -\frac{1}{2}(M^{ui} - \Gamma^{ui} U^*)' \Omega_i^{-1} (M^{ui} - \Gamma^{ui} U^*)$$

The eighth source of information comes from the measurement equation that links survey to U^* . Rewrite the equation in a matrix notation,

$$F^u = \beta^u U^* + \varepsilon^{zu} \quad \varepsilon^{zu} \sim N(0, \Omega_{zu}), \quad \text{where } \Omega_{zu} = \text{diag}(\sigma_{zu}^2, \dots, \sigma_{zu}^2) \quad (\text{C.47})$$

where,

$$f_t^u = Z_t^u - C_t^u,$$

$$F^u = (f_1^u, \dots, f_T^u)'$$

Ignoring any terms not involving U^* , we have

$$\log p(Z^u|U^*, U, \pi, \bullet) \propto -\frac{1}{2}(F^u - \beta^u U^*)' \Omega_{zu}^{-1} (F^u - \beta^u U^*)$$

Combining the above eight conditional densities we obtain,

$$\log p(U^*|Y, \bullet) \propto -\frac{1}{2}(U^* - \hat{U}^*)' D_{U^*}^{-1} (U^* - \hat{U}^*) + g_{u^*}(U^*, \sigma_{u^*}^2)$$

where,

$$D_{U^*} = (H' \Omega_{U^*}^{-1} H + K_u' \Omega_u^{-1} K_u + \Lambda \pi' \Omega_\pi^{-1} \Lambda \pi + \Lambda^w' \Omega_w^{-1} \Lambda^w + \Lambda^g' \Omega_{ogap}^{-1} \Lambda^g + \Gamma^{ui'} \Omega_i^{-1} \Gamma^{ui} + \Lambda^P' \Omega_P^{-1} \Lambda^P + (\beta^u)^2 \Omega_{zu}^{-1})^{-1}$$

$$\hat{U}^* = D_{U^*} (H' \Omega_{U^*}^{-1} \alpha_u + K_u' \Omega_u^{-1} K_u (U - K_u^{-1} \mu_u) + \Lambda \pi' \Omega_\pi^{-1} Z + \Lambda^w' M^w + \Lambda^w + \Lambda^g' \Omega_{ogap}^{-1} M^g + \Gamma^{ui'} \Omega_i^{-1} M^{ui} + \Lambda^P' \Omega_P^{-1} M^P + \beta^u \Omega_{zu}^{-1} F^u)$$

The addition of the term $g_{u^*}(U^*, \sigma_{u^*}^2)$ leads to a non-standard density. Accordingly, we sample U^* using an independence-chain Metropolis-Hastings (MH) procedure. This involves first generating candidate draws from $N(\hat{U}^*, D_{U^*})$ using the precision based algorithm that are then accepted or rejected based on accept-reject Metropolis-Hastings (ARMH) algorithm (discussed in Chan and Strachan, 2012).

Step 2. Derive the conditional distribution $p(gdp^*|Y, \bullet)$

The information about gdp^* comes from five sources. Below, we derive an expression for each of these sources.

The first source is the state equation of gdp^* . We rewrite it in a matrix notation as follows,

$$H_2 gdp^* = \alpha_{gdp^*} + \varepsilon^{gdp^*} \quad \varepsilon^{gdp^*} \sim N(0, \Omega_{gdp^*}), \quad \text{where } \Omega_{gdp^*} = \text{diag}(\omega_{gdp^*}^2, \sigma_{gdp^*}^2, \dots, \sigma_{gdp^*}^2) \quad (\text{C.48})$$

where,

$$\alpha_{gdp^*} = \begin{pmatrix} gdp_0^* + \Delta gdp_0^* \\ -gdp_0^* \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}$$

H_2 is a band matrix with unit determinant and hence is invertible.

The prior density for gdp^* is given by

$$p(gdp^* | \sigma_{gdp^*}^2) \propto -\frac{1}{2}(gdp^* - H_2^{-1}\alpha_{gdp^*})' H_2' \Omega_{gdp^*}^{-1} H_2 (gdp^* - H_2^{-1}\alpha_{gdp^*})$$

The second source of information about gdp^* is from the output gap measurement equation. Rewrite in matrix form,

$$H_{rhog}gdp = H_{rhog}gdp^* + a^r \tilde{r} + \lambda^g \tilde{u} + \alpha_{gmore} + \varepsilon^{ogap} \quad \varepsilon^{ogap} \sim N(0, \Omega_{ogap}), \quad \text{where } \Omega_{ogap} = \text{diag}(\sigma_{ogap}^2, \dots, \sigma_{ogap}^2) \quad (\text{C.49})$$

where,

$$\alpha_{gmore} = \begin{pmatrix} \rho_1^g(gdp_0 - gdp_0^*) + \rho_2^g(gdp_{-1} - gdp_{-1}^*) \\ \rho_2^g(gdp_0 - gdp_0^*) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad H_{rhog} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -\rho_1^g & 1 & 0 & 0 & \cdots & 0 \\ -\rho_2^g & -\rho_1^g & 1 & 0 & \cdots & 0 \\ 0 & -\rho_2^g & -\rho_1^g & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\rho_2^g & -\rho_1^g & 1 \end{pmatrix},$$

$$\tilde{r} = \begin{pmatrix} r_1 - r_1^* \\ r_2 - r_2^* \\ r_3 - r_3^* \\ \vdots \\ r_T - r_T^* \end{pmatrix} \quad \tilde{u} = \begin{pmatrix} U_1 - U_1^* \\ U_2 - U_2^* \\ U_3 - U_3^* \\ \vdots \\ U_T - U_T^* \end{pmatrix}$$

$$\log p(gdp | gdp^*, \bullet) \propto -\frac{1}{2}(gdp - H_{rhog}^{-1}(H_{rhog}gdp^* + a^r \tilde{r} + \lambda^g \tilde{u} + \alpha_{gmore}))' H_{rhog}' \Omega_{ogap}^{-1} H_{rhog} (gdp - H_{rhog}^{-1}(H_{rhog}gdp^* + a^r \tilde{r} + \lambda^g \tilde{u} + \alpha_{gmore}))$$

The third source of information comes from the unemployment gap measurement equation. Rewrite that equation in matrix notation,

$$Y^{ugdp} = \Gamma^u gdp^* + \varepsilon^u \quad \varepsilon^u \sim N(0, \Omega_u), \quad \text{where } \Omega_u = \text{diag}(\sigma_u^2, \dots, \sigma_u^2) \quad (\text{C.50})$$

where,

$$y_t^{ugdp} = \tilde{u}_t - \rho_1^u u_{t-1} - \rho_2^u u_{t-2} - \phi^u gdp, \quad \text{where } \tilde{u}_t = (U_t - U_t^*)$$

$$Y^{ugdp} = (y_1^{ugdp}, \dots, y_T^{ugdp})'$$

Ignoring any terms not involving gdp^* , we have

$$\log p(U|gdp^*, \bullet) \propto -\frac{1}{2}(Y^{ugd}p - \Gamma^u gdp^*)' \Omega_u^{-1} (Y^{ugd}p - \Gamma^u gdp^*)$$

The fourth source of information comes from the equation linking r-star to g-star, i.e.,

$$r_t^* = \zeta(gdp_t^* - gdp_{t-1}^*) + D_t \quad (\text{C.51})$$

Rewrite this equation in matrix notation,

$$r^* = \zeta H gdp^* + \alpha_{gr} + D \quad (\text{C.52})$$

where,

$$\alpha_{gr} = (-\zeta gdp_0^*, 0, 0, \dots, 0)'$$

Ignoring any terms not involving gdp^* , we have

$$\log p(r^*|gdp^*, D, \bullet) \propto -\frac{1}{2}(r^* - (\zeta H gdp^* + \alpha_{gr} + D))'(r^* - (\zeta H gdp^* + \alpha_{gr} + D))$$

The fifth source of information comes from the measurement equation that links survey to g^* . Rewrite the equation in a matrix notation,

$$F^g = \beta^g (H gdp^* - \alpha_g) + \varepsilon^{zg} \quad \varepsilon^{zg} \sim N(0, \Omega_{zg}), \quad \text{where } \Omega_{zg} = \text{diag}(\sigma_{zg}^2, \dots, \sigma_{zg}^2) \quad (\text{C.53})$$

where,

$$f_t^g = Z_t^g - C_t^g, \quad F^g = (f_1^g, \dots, f_T^g)'$$

$$\alpha_g = (gdp_0^*, 0, 0, \dots, 0)' \text{ is a } T \times 1 \text{ vector.}$$

Ignoring any terms not involving gdp^* , we have

$$\log p(Z^g|gdp^*, \bullet) \propto -\frac{1}{2}(F^g - \beta^g (H gdp^* - \alpha_g))' \Omega_{zg}^{-1} (F^g - \beta^g (H gdp^* - \alpha_g))$$

Combining the above five conditional densities we obtain,

$$\log p(gdp^*|Y, \bullet) \propto -\frac{1}{2}(gdp^* - \hat{gdp}^*)' D_{gdp^*}^{-1} (gdp^* - \hat{gdp}^*)$$

where,

$$D_{gdp^*} = (H_2' \Omega_{gdp^*}^{-1} H_2 + H_{rhog}' \Omega_{ogap}^{-1} H_{rhog} + \Gamma^{u'} \Omega_u^{-1} \Gamma^u + (\zeta H)' (\zeta H) + \beta^g H' \Omega_{zg}^{-1} \beta^g H)^{-1}$$

$$\hat{gdp}^* = D_{gdp^*} (H_2' \Omega_{gdp^*}^{-1} H_2 \alpha_{gdp^*} + H_{rhog}' \Omega_{ogap}^{-1} (H_{rhog} gdp - a^r \tilde{r} - \lambda^g \tilde{u} - \alpha_{gmore}) + \Gamma^{u'} \Omega_u^{-1} Y^{ugdp} + (\zeta H)' (r^* - \alpha_{gr} + D) + \beta^g H' \Omega_{zg}^{-1} F^g)$$

Step 3. Derive the conditional distribution $p(P^*|Y, \bullet)$

First, rewrite the productivity measurement eq. as

$$K_P P = \mu_P + K_P P^* + \varepsilon^P \quad \varepsilon^P \sim N(0, \Omega_P), \quad \text{where } \Omega_P = \text{diag}(e^{h_1^P}, e^{h_2^P}, \dots, e^{h_T^P}) \quad (\text{C.54})$$

$$\mu_P = \begin{pmatrix} \rho_1^P (P_0 - P_0^*) + \lambda_1^P (U_1 - U_1^*) \\ \lambda_2^P (U_2 - U_2^*) \\ \lambda_3^P (U_3 - U_3^*) \\ \vdots \\ \lambda_T^P (U_T - U_T^*) \end{pmatrix}, \quad K_P = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho_2^P & 1 & 0 & \cdots & 0 \\ 0 & -\rho_3^P & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho_T^P & 1 \end{pmatrix}, \quad P^* = \begin{pmatrix} P_1^* \\ P_2^* \\ P_3^* \\ \vdots \\ P_T^* \end{pmatrix}$$

Since $|K_P| = 1$ for any ρ_P , K_P is invertible. Therefore, we have likelihood

$$p(P|P^*, U, \bullet) \sim N(K_P^{-1} \mu_P + P^*, (K_P' \Omega_P^{-1} K_P)^{-1})$$

i.e.,

$$\log p(P|U, \bullet) \propto -\frac{1}{2} \iota_T h^P - \frac{1}{2} (P - K_P^{-1} \mu_P - P^*)' K_P' \Omega_P^{-1} K_P (P - K_P^{-1} \mu_P - P^*),$$

where ι_T is a $T \times 1$ columns of ones.

Similarly, rewrite the state equation for P^* as

$$H P^* = \alpha_P + \varepsilon^{P^*} \quad \varepsilon^{P^*} \sim N(0, \Omega_{P^*}), \quad \text{where } \Omega_{P^*} = \text{diag}(\omega_{P^*}^2, \sigma_{P^*}^2, \dots, \sigma_{P^*}^2) \quad (\text{C.55})$$

where,

$$\alpha_p = \begin{pmatrix} P_0^* \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad K_P = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

That is, the prior density for P^* is given by

$$p(P^* | \sigma_{P^*}^2) \propto -\frac{1}{2}(P^* - H^{-1}\alpha_p)' H' \Omega_{P^*}^{-1} H (P^* - H^{-1}\alpha_p)$$

Now account for the third source of information about P^* in the equation $W^* = P^* + \pi^* + \varepsilon^{w^*}$,

$$p(P^* | W^*, \pi^*, \sigma_{W^*}^2) \propto -\frac{1}{2}(P^* - (W^* - \pi^*))' \Omega_{W^*}^{-1} (P^* - (W^* - \pi^*))$$

where,

$$\Omega_{W^*} = \text{diag}(\sigma_{W^*}^2, \sigma_{W^*}^2, \dots, \sigma_{W^*}^2), \quad W^* = (W_1^*, \dots, W_T^*)', \quad \pi^* = (\pi_1^*, \dots, \pi_T^*)'$$

Combining the above three conditional densities we obtain,

$$\log p(P^* | Y, \bullet) \propto -\frac{1}{2}(P^* - \hat{P}^*)' D_{P^*}^{-1} (P^* - \hat{P}^*)$$

where,

$$D_{P^*} = (H' \Omega_{P^*}^{-1} H + K_P' \Omega_P^{-1} K_P + \Omega_{W^*}^{-1})^{-1}$$

$$\hat{P}^* = D_{P^*} (H^{-1} \Omega_{P^*}^{-1} \alpha_p + K_P' \Omega_P^{-1} K_P (P - K_P^{-1} \mu_P) + \Omega_{W^*}^{-1} (W^* - \pi^*))$$

The candidate draws are sampled from $N(\hat{P}^*, D_{P^*})$ using the precision based algorithm.

Step 4. Derive the conditional distribution $p(\pi^* | Y, \bullet)$

The information about π^* comes from six sources. Below, we derive an expression for each of these sources.

The first source is the inflation measurement equation. Rewrite it in a matrix notation as,

$$K_\pi \pi = \mu_\pi + K_\pi \pi^* + \varepsilon^\pi \quad \varepsilon^\pi \sim N(0, \Omega_\pi), \quad \text{where } \Omega_\pi = \text{diag}(e^{h_1^\pi}, e^{h_2^\pi}, \dots, e^{h_T^\pi}) \quad (\text{C.56})$$

where,

$$\mu_\pi = \begin{pmatrix} \rho_1^\pi(\pi_0 - \pi_0^*) + \lambda_1^\pi(U_1 - U_1^*) \\ \lambda_2^\pi(U_2 - U_2^*) \\ \lambda_3^\pi(U_3 - U_3^*) \\ \vdots \\ \lambda_T^\pi(U_T - U_T^*) \end{pmatrix}, \quad K_\pi = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho_2^\pi & 1 & 0 & \cdots & 0 \\ 0 & -\rho_3^\pi & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho_T^\pi & 1 \end{pmatrix}$$

Since $|K_\pi| = 1$ for any ρ_π , K_π is invertible. Therefore, we have likelihood

$$\log p(\pi|U, U^*, \bullet) \propto -\frac{1}{2} \iota_T h^\pi - \frac{1}{2} (\pi - (K_\pi^{-1} \mu_\pi + \pi^*))' K_\pi' \Omega_\pi^{-1} K_\pi (\pi - (K_\pi^{-1} \mu_\pi + \pi^*))$$

The second source of information is from the state equation of π^* . Rewrite it in a matrix notation,

$$H\pi^* = \alpha_\pi + \varepsilon^{\pi^*} \quad \varepsilon^{\pi^*} \sim N(0, \Omega_{\pi^*}), \quad \text{where } \Omega_{\pi^*} = \text{diag}(\omega_{\pi^*}^2, \sigma_{\pi^*}^2, \dots, \sigma_{\pi^*}^2) \quad (\text{C.57})$$

where,

$$\alpha_\pi = \begin{pmatrix} \pi_0^* \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

That is, the prior density for π^* is given by

$$p(\pi^* | \sigma_{\pi^*}^2) \propto -\frac{1}{2} (\pi^* - H^{-1} \alpha_\pi)' H' \Omega_{\pi^*}^{-1} H (\pi^* - H^{-1} \alpha_\pi)$$

Now account for the third source of information about π^* in the equation $W^* = P^* + \pi^* + \varepsilon^{w^*}$,

$$p(\pi^* | W^*, P^*, \sigma_{W^*}^2) \propto -\frac{1}{2} (\pi^* - (W^* - P^*))' \Omega_{W^*}^{-1} (\pi^* - (W^* - P^*))$$

where,

$$\Omega_{W^*} = \text{diag}(\sigma_{W^*}^2, \sigma_{W^*}^2, \dots, \sigma_{W^*}^2), \quad W^* = (W_1^*, \dots, W_T^*)', \quad P^* = (P_1^*, \dots, P_T^*)'$$

The fourth source of information is from the wage measurement equation. Rewrite in matrix notation,

$$M^{w\pi} = X_{w\pi}\pi^* + \varepsilon^w \quad \varepsilon^w \sim N(0, \Omega_w), \quad \text{where } \Omega_w = \text{diag}(e^{h_1^w}, e^{h_2^w}, \dots, e^{h_T^w}) \quad (\text{C.58})$$

where,

$$m_t^{w\pi} = w_t - w_t^* - \rho_t^w(w_{t-1} - w_{t-1}^*) - \lambda_t^w(U_t - U_t^*) - \kappa_t^w \pi_t$$

$$M^{w\pi} = (m_1^{w\pi}, m_2^{w\pi}, \dots, m_T^{w\pi})$$

$$X_{w\pi} = \begin{pmatrix} -\kappa_1^w & 0 & 0 & \dots & 0 \\ 0 & -\kappa_2^w & 0 & \dots & 0 \\ 0 & 0 & -\kappa_3^w & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & -\kappa_T^w \end{pmatrix}$$

$$\log p(W|\pi^*, \bullet) \propto -\frac{1}{2}(M^{w\pi} - X_{w\pi}\pi^*)'\Omega_w^{-1}(M^{w\pi} - X_{w\pi}\pi^*)$$

The fifth source is the Taylor-rule equation. Rewrite the equation in the matrix notation,

$$M^{\pi i} = \alpha_{\pi i} + (K_{\pi i} + \Gamma_{\pi})\pi^* + \varepsilon^i \quad \varepsilon^i \sim N(0, \Omega_i), \quad \text{where } \Omega_i = \text{diag}(e^{h_1^i}, e^{h_2^i}, \dots, e^{h_T^i}) \quad (\text{C.59})$$

where,

$$m_t^{\pi i} = i_t - \rho^i i_{t-1} - r_t^* + \rho^i r_{t-1}^* - \lambda^i(U_t - U_t^*) - \kappa^i \pi_t$$

$$M^{\pi i} = (m_1^{\pi i}, m_2^{\pi i}, \dots, m_T^{\pi i})'$$

$$K_{\pi i} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho^i & 1 & 0 & \cdots & 0 \\ 0 & -\rho^i & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho^i & 1 \end{pmatrix}, \quad \Gamma_{\pi} = \begin{pmatrix} -\kappa^i & 0 & 0 & \cdots & 0 \\ 0 & -\kappa^i & 0 & \cdots & 0 \\ 0 & 0 & -\kappa^i & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & -\kappa^i \end{pmatrix}, \quad \alpha_{\pi i} = \begin{pmatrix} -\rho^i \pi_0^* \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\log p(i|\pi^*, \pi, \bullet) \propto -\frac{1}{2}(M^{\pi i} - (\alpha_{\pi i} + (K_{\pi i} + \Gamma_{\pi})\pi^*))' \Omega_i^{-1} (M^{\pi i} - (\alpha_{\pi i} + (K_{\pi i} + \Gamma_{\pi})\pi^*))$$

The sixth source of information comes from the measurement equation that links survey to π^* . Rewrite the equation in a matrix notation,

$$F^{\pi} = \beta^{\pi} \pi^* + \varepsilon^{z^{\pi}} \quad \varepsilon^{z^{\pi}} \sim N(0, \Omega_{z^{\pi}}), \quad \text{where } \Omega_{z^{\pi}} = \text{diag}(\sigma_{z^{\pi}}^2, \dots, \sigma_{z^{\pi}}^2) \quad (\text{C.60})$$

where,

$$f_t^{\pi} = Z_t^{\pi} - C_t^{\pi},$$

$$F^{\pi} = (f_1^{\pi}, \dots, f_T^{\pi})'$$

Ignoring any terms not involving π^* , we have

$$\log p(Z^{\pi}|\pi^*, \pi, \bullet) \propto -\frac{1}{2}(F^{\pi} - \beta^{\pi} \pi^*)' \Omega_{z^{\pi}}^{-1} (F^{\pi} - \beta^{\pi} \pi^*)$$

Combining the above six conditional densities we obtain,

$$\log p(\pi^*|Y, \bullet) \propto -\frac{1}{2}(\pi^* - \hat{\pi}^*)' D_{\pi^*}^{-1} (\pi^* - \hat{\pi}^*)$$

where,

$$D_{\pi^*} = (H' \Omega_{\pi^*}^{-1} H + K_{\pi}^{\prime} \Omega_{\pi}^{-1} K_{\pi} + \Omega_w^{-1} + X'_{w\pi} \Omega_w^{-1} X_{w\pi} + (K'_{\pi i} + \Gamma_{\pi})' \Omega_i^{-1} (K'_{\pi i} + \Gamma_{\pi})' + (\beta^{\pi})^2 \Omega_{z^{\pi}}^{-1})^{-1}$$

$$\hat{\pi}^* = D_{\pi^*} (H' \Omega_{\pi^*}^{-1} \alpha_{\pi} + K_{\pi}^{\prime} \Omega_{\pi}^{-1} K_{\pi} (\pi - K_{\pi}^{-1} \mu_{\pi}) + \Omega_w^{-1} (W^* - P^*) + X'_{w\pi} \Omega_w^{-1} M^{w\pi} + (K'_{\pi i} + \Gamma_{\pi})' \Omega_i^{-1} (M^{\pi i} - \alpha_{\pi i}) + \beta^{\pi} \Omega_{z^{\pi}}^{-1} F^{\pi})$$

The candidate draws are sampled from $N(\hat{\pi}^*, D_{\pi^*})$ using the precision based algorithm.

Step 5. Derive the conditional distribution $p(w^*|Y, \bullet)$

The information about w^* comes from two sources. Below, we derive an expression for each of these sources.

The first source is the nominal wage measurement equation. Rewrite it in a matrix notation as,

$$K_w W = \mu_w + K_w W^* + \varepsilon^w \quad \varepsilon^w \sim N(0, \Omega_w), \quad \text{where } \Omega_w = \text{diag}(e^{h_1^w}, e^{h_2^w}, \dots, e^{h_T^w}) \quad (\text{C.61})$$

where,

$$\mu_w = \begin{pmatrix} \rho_1^w(W_0 - W_0^*) + \lambda_1^w(U_1 - U_1^*) + \kappa_1^w(\pi_1 - \pi_1^*) \\ \lambda_2^w(U_2 - U_2^*) + \kappa_2^w(\pi_2 - \pi_2^*) \\ \lambda_3^w(U_3 - U_3^*) + \kappa_3^w(\pi_3 - \pi_3^*) \\ \vdots \\ \lambda_T^w(U_T - U_T^*) + \kappa_T^w(\pi_T - \pi_T^*) \end{pmatrix}, \quad K_w = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho_2^w & 1 & 0 & \cdots & 0 \\ 0 & -\rho_3^w & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho_T^w & 1 \end{pmatrix}$$

Since $|K_w| = 1$ for any ρ_w , K_w is invertible. Therefore, we have likelihood

Ignoring any terms not involving w^* , we have

$$\log p(W|W^*, \bullet) \propto -\frac{1}{2} \iota_T h^w - \frac{1}{2} (W - (K_w^{-1} \mu_w + W^*))' K_w' \Omega_w^{-1} K_w (W - (K_w^{-1} \mu_w + W^*))$$

The second source is the state equation of W^* , which describes W^* as the sum of P^* and π^* . This equation can be thought of describing prior density for W^* . Rewrite it in a matrix form.

$$W^* = P^* + \pi^* + \varepsilon^{w*} \quad \varepsilon^{w*} \sim N(0, \Omega_{w*}) \quad (\text{C.62})$$

$$p(W^*|P^*, \pi^*, \sigma_{w*}^2) \propto -\frac{1}{2} (W^* - (P^* + \pi^*))' \Omega_{w*}^{-1} (W^* - (P^* + \pi^*))$$

Combining the above two conditional densities we obtain,

$$\log p(W^*|Y, \bullet) \propto -\frac{1}{2} (W^* - \hat{W}^*)' D_{W^*}^{-1} (W^* - \hat{W}^*)$$

where,

$$D_{W^*} = (K_w' \Omega_w^{-1} K_w + \Omega_{W^*}^{-1})^{-1}$$

$$\hat{W}^* = D_{W^*} (K_w' \Omega_w^{-1} (K_w W - \mu_w) + \Omega_{w^*}^{-1} (P^* + \pi^*))$$

The candidate draws are sampled from $N(\hat{W}^*, D_{W^*})$ using the precision based algorithm.

Step 6. Derive the conditional distribution $p(r^*|Y, \bullet)$

The information about r^* comes from four sources. Below, we derive an expression for each of these sources.

The first source is the output gap measurement equation. We rewrite it in a matrix notation as follows,

$$H_{rhog} ogap = \alpha_{ogap} - a^r r^* + \varepsilon^{ogap} \quad \varepsilon^{ogap} \sim N(0, \Omega_{ogap}) \quad (\text{C.63})$$

where,

$$\alpha_{ogap} = \begin{pmatrix} \rho_1^g(ogap_0) + \rho_2^g(ogap_{-1}) + a^r r_1 + \lambda^g(U_1 - U_1^*) \\ \rho_2^g(ogap_0) + a^r r_2 + \lambda^g(U_2 - U_2^*) \\ a^r r_3 + \lambda^g(U_3 - U_3^*) \\ \vdots \\ a^r r_T + \lambda^g(U_T - U_T^*) \end{pmatrix}$$

Ignoring any terms not involving r^* , we have

$$\log p(ogap|r^*, \bullet) \propto -\frac{1}{2}(ogap - H_{rhog}^{-1}(\alpha_{ogap} - a^r r^*))' H_{rhog}' \Omega_{ogap}^{-1} H_{rhog} (ogap - H_{rhog}^{-1}(\alpha_{ogap} - a^r r^*))$$

The second source is the state equation linking r^* to g^* . We rewrite it in a matrix notation as follows,

$$r^* = \zeta \Delta gdp^* + H^{-1} \varepsilon^d \quad \varepsilon^d \sim N(0, \Omega_d), \quad \text{where } \Omega_d = \text{diag}(\omega_d^2, \sigma_d^2, \dots, \sigma_d^2) \quad (\text{C.64})$$

Ignoring any terms not involving r^* , the prior density for r^* is given by

$$\log p(r^*|gdp^*, \sigma_d^2, \bullet) \propto -\frac{1}{2}(r^* - \zeta \Delta gdp^*)' H' \Omega_d^{-1} H (r^* - \zeta \Delta gdp^*)$$

The third source is the Taylor-type rule equation. We rewrite it in a matrix notation as follows,

$$M^{ri} = \alpha_{ri} + K_{\pi i} r^* + \varepsilon^i \quad \varepsilon^i \sim N(0, \Omega_i), \quad \text{where } \Omega_i = \text{diag}(e^{h_1^i}, e^{h_2^i}, \dots, e^{h_T^i}) \quad (\text{C.65})$$

where,

$$m_t^{ri} = i_t - \rho^i i_{t-1} - \pi_t^* + \rho^i \pi_{t-1}^* - \lambda^i (U_t - U_t^*) - \kappa^i (\pi_t - \pi_t^*),$$

$$M^{ri} = (m_1^{ri}, m_2^{ri}, \dots, m_T^{ri})'$$

$$\alpha_{ri} = \begin{pmatrix} -\rho^i r_0^* \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad K_{\pi i} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\rho^i & 1 & 0 & \dots & 0 \\ 0 & -\rho^i & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & -\rho^i & 1 \end{pmatrix}$$

Ignoring any terms not involving r^* , we have

$$\log p(i|r^*, \bullet) \propto -\frac{1}{2} \iota_T h^i - \frac{1}{2} (M^{ri} - (\alpha_{ri} + K_{\pi i} r^*))' \Omega_i^{-1} (M^{ri} - (\alpha_{ri} + K_{\pi i} r^*))$$

The fourth source of information comes from the measurement equation that links survey to r^* . Rewrite the equation in a matrix notation,

$$F^r = \beta^r r^* + \varepsilon^{zr} \quad \varepsilon^{zr} \sim N(0, \Omega_{zr}), \quad \text{where } \Omega_{zr} = \text{diag}(\sigma_{zr}^2, \dots, \sigma_{zr}^2) \quad (\text{C.66})$$

where,

$$f_t^r = Z_t^r - C_t^r,$$

$$F^r = (f_1^r, \dots, f_T^r)'$$

Ignoring any terms not involving r^* , we have

$$\log p(Z^r|r^*, \bullet) \propto -\frac{1}{2} (F^r - \beta^r r^*)' \Omega_{zr}^{-1} (F^r - \beta^r r^*)$$

Combining the above four conditional densities we obtain,

$$\log p(r^*|Y, \bullet) \propto -\frac{1}{2} (r^* - \hat{r}^*)' D_{r^*}^{-1} (r^* - \hat{r}^*)$$

where,

$$D_{r*} = ((-a^r)^2 \Omega_{ogap}^{-1} + H' \Omega_d^{-1} H + K'_{\pi i} \Omega_i^{-1} K_{\pi i} + (\beta^r)(2) \Omega_{zr}^{-1})^{-1}$$

$$\hat{r}^* = D_{r*} (-a^r \Omega_{ogap}^{-1} (H_{rhogogap} - \alpha_{ogap}) + H' \Omega_d^{-1} H \zeta \Delta g d p^* + K'_{\pi i} \Omega_i^{-1} (M^{ri} - \alpha_{ri}) + \beta^r \Omega_{zr}^{-1} F^r)$$

The candidate draws are sampled from $N(\hat{r}^*, D_{r*})$ using the precision based algorithm.

Step 7. Derive the conditional distribution $p(\lambda^p|Y, \bullet)$

The information about λ^p comes from two sources. Below, we derive an expression for each of these two sources.

The first source is the productivity measurement equation. Rewrite it in a matrix notation,

$$B = X_u \lambda^p + \varepsilon^p \quad \varepsilon^p \sim N(0, \Omega_p) \quad (\text{C.67})$$

where,

$$\begin{aligned} B &= (\tilde{p}_1 - \rho^p \tilde{p}_0, \dots, \tilde{p}_T - \rho^p \tilde{p}_{T-1}) \\ \tilde{p}_t &= p_t - p_t^* \\ \tilde{u}_t &= U_t - U_t^* \\ X_u &= \text{diag}(\tilde{u}_1, \dots, \tilde{u}_T) \end{aligned}$$

Ignoring any terms not involving λ^p , we have the likelihood

$$\log p(p|\lambda^p, \bullet) \propto -\frac{1}{2} (B - X_u \lambda^p)' \Omega_p^{-1} (B - X_u \lambda^p)$$

The second source of information comes from state equation for λ^p . We rewrite it in a matrix notation as follows,

$$H \lambda^p = \varepsilon^{\lambda^p} \quad \varepsilon^{\lambda^p} \sim N(0, \Omega_{\lambda^p}), \quad \text{where } \Omega_{\lambda^p} = \text{diag}(\omega_{\lambda^p}^2, \sigma_{\lambda^p}^2, \dots, \sigma_{\lambda^p}^2) \quad (\text{C.68})$$

Ignoring any terms not involving λ^p , the prior density for λ^p is given by

$$\log p(\lambda^p | \sigma_{\lambda^p}^2, \Omega_{\lambda^p}) \propto -\frac{1}{2} (\lambda^p)' H' \Omega_{\lambda^p}^{-1} H (\lambda^p)$$

Combining the above two conditional densities we obtain,

$$\log p(\lambda^p|Y, \bullet) \propto -\frac{1}{2}(\lambda^p - \hat{\lambda}^p)' D_{\lambda^p}^{-1}(\lambda^p - \hat{\lambda}^p)$$

where,

$$D_{\lambda^p} = (H' \Omega_{\lambda^p}^{-1} H + X_u' \Omega_p^{-1} X_u)^{-1}$$

$$\hat{\lambda}^p = D_{\lambda^p} (X_u' \Omega_p^{-1} B)$$

The candidate draws are sampled from $N(\hat{\lambda}^p, D_{\lambda^p})$ using the precision based algorithm.

Step 8. Derive the conditional distribution $p(\rho^\pi|Y, \bullet)$

The information about ρ^π comes from two sources. Below, we derive an expression for each of these two sources.

First, we define some notation,

$$\begin{aligned} \tilde{\pi}_t &= \pi_t - \pi_t^* \\ \tilde{u}_t &= U_t - U_t^* \\ \tilde{\Pi} &= (\tilde{\pi}_1, \dots, \tilde{\pi}_T)' \\ \tilde{u} &= (\tilde{u}_1, \dots, \tilde{u}_T)' \end{aligned}$$

The first source is the price inflation measurement equation. Rewrite it in a matrix notation,

$$\tilde{\Pi} + \Lambda \tilde{u} = X_\pi \rho^\pi + \varepsilon^\pi \quad \varepsilon^\pi \sim N(0, \Omega_\pi) \quad (\text{C.69})$$

where,

$$\begin{aligned} X_\pi &= \text{diag}(\tilde{\pi}_0, \dots, \tilde{\pi}_{T-1}) \\ \Lambda &= \text{diag}(-\lambda_1^\pi, \dots, -\lambda_T^\pi) \end{aligned}$$

Ignoring any terms not involving ρ^π , we have the likelihood

$$\log p(\pi|\rho^\pi, \bullet) \propto -\frac{1}{2}(\tilde{\Pi} - (X_\pi \rho^\pi - \Lambda \tilde{u}))' \Omega_\pi^{-1} (\tilde{\Pi} - (X_\pi \rho^\pi - \Lambda \tilde{u}))$$

The second source comes from state equation for ρ^π . We rewrite it in a matrix notation as

follows,

$$H\rho^\pi = \varepsilon^{\rho^\pi} \quad \varepsilon^{\rho^\pi} \sim N(0, \Omega_{\rho^\pi}), \quad \text{where } \Omega_{\rho^\pi} = \text{diag}(\omega_{\rho^\pi}^2, \sigma_{\rho^\pi}^2, \dots, \sigma_{\rho^\pi}^2) \quad (\text{C.70})$$

$$0 < \rho_t^\pi < 1 \text{ for } t=1, \dots, T$$

Ignoring any terms not involving ρ^π , the prior density for ρ^π is given by

$$\log p(\rho^\pi | \sigma_{\rho^\pi}^2, \Omega_{\rho^\pi}) \propto -\frac{1}{2}(\rho^\pi)' H' \Omega_{\rho^\pi}^{-1} H (\rho^\pi) + g_{\rho^\pi}(\rho^\pi, \sigma_{\rho^\pi}^2)$$

where,

$$g_{\rho^\pi}(\rho^\pi, \sigma_{\rho^\pi}^2) = -\sum_{t=2}^T \log \left(\Phi \left(\frac{1 - \rho_{t-1}^\pi}{\sigma_{\rho^\pi}} \right) - \Phi \left(\frac{0 - \rho_{t-1}^\pi}{\sigma_{\rho^\pi}} \right) \right)$$

Combining the above two conditional densities we obtain,

$$\log p(\rho^\pi | Y, \bullet) \propto -\frac{1}{2}(\rho^\pi - \hat{\rho}^\pi)' D_{\rho^\pi}^{-1} (\rho^\pi - \hat{\rho}^\pi) + g_{\rho^\pi}(\rho^\pi, \sigma_{\rho^\pi}^2)$$

where,

$$D_{\rho^\pi} = (H' \Omega_{\rho^\pi}^{-1} H + X_\pi' \Omega_\pi^{-1} X_\pi)^{-1}$$

$$\hat{\rho}^\pi = D_{\rho^\pi} (X_\pi' \Omega_\pi^{-1} (\tilde{\Pi} + \Lambda \tilde{u}))$$

The addition of the term $g_{\rho^\pi}(\rho^\pi, \sigma_{\rho^\pi}^2)$ leads to a non-standard density. Accordingly, we sample ρ^π using an independence-chain Metropolis-Hastings (MH) procedure. This involves first generating candidate draws from $N(\hat{\rho}^\pi, D_{\rho^\pi})$ using the precision based algorithm that are then accepted or rejected based on accept-reject Metropolis-Hastings (ARMH) algorithm (discussed in Chan and Strachan, 2012).

Step 9. Derive the conditional distribution $p(\lambda^\pi | Y, \bullet)$

The information about λ^π comes from two sources. Below, we derive an expression for each of these two sources.

First, we define some notation,

$$\tilde{\pi}_t = \pi_t - \pi_t^*$$

$$\begin{aligned}\tilde{u}_t &= U_t - U_t^* \\ NW &= (\tilde{\pi}_1 - \rho_1^\pi \tilde{\pi}_0, \dots, \tilde{\pi}_T - \rho_T^\pi \tilde{\pi}_{T-1})'\end{aligned}$$

The first source is the price inflation measurement equation. Rewrite it in a matrix notation,

$$NW = X_u \lambda^\pi + \varepsilon^\pi \quad \varepsilon^\pi \sim N(0, \Omega_\pi) \quad (\text{C.71})$$

where,

$$X_u = \text{diag}(\tilde{u}_1, \dots, \tilde{u}_T)$$

Ignoring any terms not involving λ^π , we have the likelihood

$$\log p(\pi | \lambda^\pi, \bullet) \propto -\frac{1}{2} (NW - X_u \lambda^\pi)' \Omega_\pi^{-1} (NW - X_u \lambda^\pi)$$

The second source comes from state equation for λ^π . We rewrite it in a matrix notation as follows,

$$H \lambda^\pi = \varepsilon^{\lambda^\pi} \quad \varepsilon^{\lambda^\pi} \sim N(0, \Omega_{\lambda^\pi}), \quad \text{where } \Omega_{\lambda^\pi} = \text{diag}(\omega_{\lambda^\pi}^2, \sigma_{\lambda^\pi}^2, \dots, \sigma_{\lambda^\pi}^2) \quad (\text{C.72})$$

$$-1 < \lambda_t^\pi < 0 \text{ for } t=1, \dots, T$$

Ignoring any terms not involving λ^π , the prior density for λ^π is given by

$$\log p(\lambda^\pi | \sigma_{\lambda^\pi}^2, \Omega_{\lambda^\pi}) \propto -\frac{1}{2} (\lambda^\pi)' H' \Omega_{\lambda^\pi}^{-1} H (\lambda^\pi) + g_{\lambda^\pi}(\lambda^\pi, \sigma_{\lambda^\pi}^2)$$

where,

$$g_{\lambda^\pi}(\lambda^\pi, \sigma_{\lambda^\pi}^2) = -\sum_{t=2}^T \log \left(\Phi \left(\frac{0 - \lambda_{t-1}^\pi}{\sigma_{\lambda^\pi}} \right) - \Phi \left(\frac{-1 - \lambda_{t-1}^\pi}{\sigma_{\lambda^\pi}} \right) \right)$$

Combining the above two conditional densities we obtain,

$$\log p(\lambda^\pi | Y, \bullet) \propto -\frac{1}{2} (\lambda^\pi - \hat{\lambda}^\pi)' D_{\lambda^\pi}^{-1} (\lambda^\pi - \hat{\lambda}^\pi) + g_{\lambda^\pi}(\lambda^\pi, \sigma_{\lambda^\pi}^2)$$

where,

$$D_{\lambda^\pi} = (H' \Omega_{\lambda^\pi}^{-1} H + X_u' \Omega_\pi^{-1} X_u)^{-1}$$

$$\hat{\lambda}^\pi = D_{\lambda^\pi} (X_u' \Omega_\pi^{-1} N W)$$

The addition of the term $g_{\lambda^\pi}(\lambda^\pi, \sigma_{\lambda^\pi}^2)$ leads to a non-standard density. Accordingly, we sample λ^π using an independence-chain Metropolis-Hastings (MH) procedure. This involves first generating candidate draws from $N(\hat{\lambda}^\pi, D_{\lambda^\pi})$ using the precision based algorithm that are then accepted or rejected based on accept-reject Metropolis-Hastings (ARMH) algorithm (discussed in Chan and Strachan, 2012).

Step 10. Derive the conditional distribution $p(\rho^w|Y, \bullet)$

The information about ρ^w comes from two sources. Below, we derive an expression for each of these two sources.

First, we define some notation,

$$\tilde{w}_t = w_t - w_t^*$$

$$\tilde{u}_t = U_t - U_t^*$$

$$\tilde{w} = (\tilde{w}_1, \dots, \tilde{w}_T)'$$

$$\tilde{u} = (\tilde{u}_1, \dots, \tilde{u}_T)'$$

$$\tilde{\pi}_t = \pi_t - \pi_t^*$$

$$\tilde{\pi} = (\tilde{\pi}_1, \dots, \tilde{\pi}_T)'$$

The first source is the wage inflation measurement equation. Rewrite it in a matrix notation,

$$\tilde{w} + \Lambda^w \tilde{u} + \Lambda^{w\pi} \tilde{\pi} = X_w \rho^w + \varepsilon^{\rho^w} \quad \varepsilon^{\rho^w} \sim N(0, \Omega_w) \quad (\text{C.73})$$

where,

$$X_w = \text{diag}(\tilde{w}_0, \dots, \tilde{w}_{T-1})$$

$$\Lambda^w = \text{diag}(-\lambda_1^w, \dots, -\lambda_T^w)$$

$$\Lambda^{w\pi} = \text{diag}(-\kappa_1^w, \dots, -\kappa_T^w)$$

Ignoring any terms not involving ρ^w , we have the likelihood

$$\log p(w|\rho^w, \bullet) \propto -\frac{1}{2}(\tilde{w} - (X_w \rho^w - \Lambda^w \tilde{u} - \Lambda^{w\pi} \tilde{\pi}))' \Omega_w^{-1} (\tilde{w} - (X_w \rho^w - \Lambda^w \tilde{u} - \Lambda^{w\pi} \tilde{\pi}))$$

The second source comes from state equation for ρ^w . We rewrite it in a matrix notation as follows,

$$H\rho^w = \varepsilon^{\rho^w} \quad \varepsilon^{\rho^w} \sim N(0, \Omega_{\rho^w}), \quad \text{where } \Omega_{\rho^w} = \text{diag}(\omega_{\rho^w}^2, \sigma_{\rho^w}^2, \dots, \sigma_{\rho^w}^2) \quad (\text{C.74})$$

$$0 < \rho_t^w < 1 \text{ for } t=1, \dots, T$$

Ignoring any terms not involving ρ^w , the prior density for ρ^w is given by

$$\log p(\rho^w | \sigma_{\rho^w}^2, \Omega_{\rho^w}) \propto -\frac{1}{2}(\rho^w)' H' \Omega_{\rho^w}^{-1} H(\rho^w) + g_{\rho^w}(\rho^w, \sigma_{\rho^w}^2)$$

where,

$$g_{\rho^w}(\rho^w, \sigma_{\rho^w}^2) = -\sum_{t=2}^T \log \left(\Phi \left(\frac{1 - \rho_{t-1}^w}{\sigma_{\rho^w}} \right) - \Phi \left(\frac{0 - \rho_{t-1}^w}{\sigma_{\rho^w}} \right) \right)$$

Combining the above two conditional densities we obtain,

$$\log p(\rho^w | Y, \bullet) \propto -\frac{1}{2}(\rho^w - \hat{\rho}^w)' D_{\rho^w}^{-1} (\rho^w - \hat{\rho}^w) + g_{\rho^w}(\rho^w, \sigma_{\rho^w}^2)$$

where,

$$D_{\rho^w} = (H' \Omega_{\rho^w}^{-1} H + X_w' \Omega_w^{-1} X_w)^{-1}$$

$$\hat{\rho}^w = D_{\rho^w} (X_w' \Omega_w^{-1} (\tilde{w} + \Lambda^w \tilde{u} + \Lambda^{w\pi} \tilde{\pi}))$$

The addition of the term $g_{\rho^\pi}(\rho^\pi, \sigma_{\rho^\pi}^2)$ leads to a non-standard density. Accordingly, we sample ρ^π using an independence-chain Metropolis-Hastings (MH) procedure. This involves first generating candidate draws from $N(\hat{\rho}^\pi, D_{\rho^\pi})$ using the precision based algorithm that are then accepted or rejected based on accept-reject Metropolis-Hastings (ARMH) algorithm (discussed in Chan and Strachan, 2012).

Step 11. Derive the conditional distribution $p(\lambda^w | Y, \bullet)$

The information about λ^w comes from two sources. Below, we derive an expression for each of

these two sources.

First, we define some notation,

$$\begin{aligned}\tilde{w}_t &= w_t - w_t^* \\ \tilde{u}_t &= U_t - U_t^* \\ \tilde{\pi}_t &= \pi_t - \pi_t^* \\ B^w &= (\tilde{w}_1 - \rho_1^w \tilde{w}_0 - \kappa_1^w \tilde{\pi}_1, \dots, \tilde{w}_T - \rho_T^w \tilde{w}_{T-1} - \kappa_{T-1}^w \tilde{\pi}_T)'\end{aligned}$$

The first source is the wage inflation measurement equation. Rewrite it in a matrix notation,

$$B^w = X_u \lambda^w + \varepsilon^w \quad \varepsilon^w \sim N(0, \Omega_w) \quad (\text{C.75})$$

where,

$$X_u = \text{diag}(\tilde{u}_1, \dots, \tilde{u}_T)$$

Ignoring any terms not involving λ^w , we have the likelihood

$$\log p(w|\lambda^w, \bullet) \propto -\frac{1}{2}(B^w - X_u \lambda^w)' \Omega_w^{-1} (B^w - X_u \lambda^w)$$

The second source comes from state equation for λ^w . We rewrite it in a matrix notation as follows,

$$H \lambda^w = \varepsilon^{\lambda^w} \quad \varepsilon^{\lambda^w} \sim N(0, \Omega_{\lambda^w}), \quad \text{where } \Omega_{\lambda^w} = \text{diag}(\omega_{\lambda^w}^2, \sigma_{\lambda^w}^2, \dots, \sigma_{\lambda^w}^2) \quad (\text{C.76})$$

$$-1 < \lambda_t^w < 0 \text{ for } t=1, \dots, T$$

Ignoring any terms not involving λ^w , the prior density for λ^w is given by

$$\log p(\lambda^w | \sigma_{\lambda^w}^2, \Omega_{\lambda^w}) \propto -\frac{1}{2}(\lambda^w)' H' \Omega_{\lambda^w}^{-1} H (\lambda^w) + g_{\lambda^w}(\lambda^w, \sigma_{\lambda^w}^2)$$

where,

$$g_{\lambda^w}(\lambda^w, \sigma_{\lambda^w}^2) = - \sum_{t=2}^T \log \left(\Phi \left(\frac{0 - \lambda_{t-1}^w}{\sigma_{\lambda^w}} \right) - \Phi \left(\frac{-1 - \lambda_{t-1}^w}{\sigma_{\lambda^w}} \right) \right)$$

Combining the above two conditional densities we obtain,

$$\log p(\lambda^w | Y, \bullet) \propto -\frac{1}{2}(\lambda^w - \hat{\lambda}^w)' D_{\lambda^w}^{-1}(\lambda^w - \hat{\lambda}^w) + g_{\lambda^w}(\lambda^w, \sigma_{\lambda^w}^2)$$

where,

$$D_{\lambda^w} = (H' \Omega_{\lambda^w}^{-1} H + X_u' \Omega_w^{-1} X_u)^{-1}$$

$$\hat{\lambda}^w = D_{\lambda^w} (X_u' \Omega_w^{-1} B^w)$$

The addition of the term $g_{\lambda^w}(\lambda^w, \sigma_{\lambda^w}^2)$ leads to a non-standard density. Accordingly, we sample λ^w using an independence-chain Metropolis-Hastings (MH) procedure. This involves first generating candidate draws from $N(\hat{\lambda}^w, D_{\lambda^w})$ using the precision based algorithm that are then accepted or rejected based on accept-reject Metropolis-Hastings (ARMH) algorithm (discussed in Chan and Strachan, 2012).

Step 12. Derive the conditional distribution $p(\kappa^w | Y, \bullet)$

The information about κ^w comes from two sources. Below, we derive an expression for each of these two sources.

First, we define some notation,

$$\tilde{w}_t = w_t - w_t^*$$

$$\tilde{u}_t = U_t - U_t^*$$

$$\tilde{\pi}_t = \pi_t - \pi_t^*$$

$$B^{\kappa^w} = (\tilde{w}_1 - \rho_1^w \tilde{w}_0 - \lambda_1^w \tilde{u}_1, \dots, \tilde{w}_T - \rho_T^w \tilde{w}_{T-1} - \lambda_{T-1}^w \tilde{u}_T)'$$

The first source is the wage inflation measurement equation. Rewrite it in a matrix notation,

$$B^{\kappa^w} = X_{\pi} \kappa^w + \varepsilon^w \quad \varepsilon^w \sim N(0, \Omega_w) \tag{C.77}$$

where,

$$X_{\pi} = \text{diag}(\tilde{\pi}_1, \dots, \tilde{\pi}_T)$$

Ignoring any terms not involving κ^w , we have the likelihood

$$\log p(w|\kappa^w, \bullet) \propto -\frac{1}{2}(B^{\kappa w} - X_{\pi}\kappa^w)' \Omega_w^{-1} (B^{\kappa w} - X_{\pi}\kappa^w)$$

The second source comes from state equation for κ^w . We rewrite it in a matrix notation as follows,

$$H\kappa^w = \varepsilon^{\kappa w} \quad \varepsilon^{\kappa w} \sim N(0, \Omega_{\kappa w}), \quad \text{where } \Omega_{\kappa w} = \text{diag}(\omega_{\kappa w}^2, \sigma_{\kappa w}^2, \dots, \sigma_{\kappa w}^2) \quad (\text{C.78})$$

Ignoring any terms not involving κ^w , the prior density for κ^w is given by

$$\log p(\kappa^w | \sigma_{\kappa w}^2, \Omega_{\kappa w}) \propto -\frac{1}{2}(\kappa^w)' H' \Omega_{\kappa w}^{-1} H (\kappa^w)$$

Combining the above two conditional densities we obtain,

$$\log p(\kappa^w | Y, \bullet) \propto -\frac{1}{2}(\kappa^w - \hat{\kappa}^w)' D_{\kappa^w}^{-1} (\kappa^w - \hat{\kappa}^w)$$

where,

$$D_{\kappa^w} = (H' \Omega_{\kappa w}^{-1} H + X_{\pi}' \Omega_w^{-1} X_{\pi})^{-1}$$

$$\hat{\kappa}^w = D_{\kappa^w} (X_{\pi}' \Omega_w^{-1} B^{\kappa w})$$

The candidate draws are sampled from $N(\hat{\kappa}^w, D_{\kappa^w})$ using the precision based algorithm.

Step 13. Derive the conditional distribution $p(h^p, h^{\pi}, h^w, h^i | Y, \bullet)$

Given parameters and other latent states, the stochastic volatility, h^p, h^{π}, h^w, h^i are conditionally independent and so can be drawn separately. Following, Chan, Koop, and Potter (2013; 2016), we draw h^p, h^{π}, h^w, h^i using the Accept-Reject independence-chain Metropolis Hastings (ARMH) algorithm of Chan and Strachan (2012; page 32-34).

Step 14. Derive the conditional distribution $p(C^u, C^g, C^{\pi}, C^r | Y, \bullet)$

Given parameters and other latent states, C^u, C^g, C^π, C^r are conditionally independent and so can be drawn separately.

Beginning with C^u , the information about it comes from two sources. Below, we derive an expression for each of these two sources.

The first source is the measurement equation linking survey to U^* . Rewrite it in a matrix notation,

$$N^{zu} = C^u + \varepsilon^{zu} \quad \varepsilon^{zu} \sim N(0, \Omega_{zu}) \quad (\text{C.79})$$

where,

$$\begin{aligned} n_t^{zu} &= Z_t^u - \beta^u U^* \\ N^{zu} &= (n_1^{zu}, n_2^{zu}, \dots, n_T^{zu})' \\ \Omega_{zu} &= \text{diag}(\sigma_{zu}^2, \dots, \sigma_{zu}^2) \end{aligned}$$

Ignoring any terms not involving C^u , we have the likelihood

$$\log p(Z^u | C^u, \bullet) \propto -\frac{1}{2} (N^{zu} - C^u)' \Omega_{zu}^{-1} (N^{zu} - C^u)$$

The second source comes from state equation for C^u . We rewrite it in a matrix notation as follows,

$$HC^u = \alpha_{cu} + \varepsilon^{cu} \quad \varepsilon^{cu} \sim N(0, \Omega_{cu}), \quad \text{where } \Omega_{cu} = \text{diag}(\omega_{cu}^2, \sigma_{cu}^2, \dots, \sigma_{cu}^2) \quad (\text{C.80})$$

where,

$$\alpha_{cu} = \begin{pmatrix} C_0^u \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Ignoring any terms not involving C^u , the prior density for C^u is given by

$$\log p(C^u | \sigma_{cu}^2, \Omega_{cu}) \propto -\frac{1}{2} (C^u - H^{-1} \alpha_{cu})' H' \Omega_{cu}^{-1} H (C^u - H^{-1} \alpha_{cu})$$

Combining the above two conditional densities we obtain,

$$\log p(C^u|Y, \bullet) \propto -\frac{1}{2}(C^u - \hat{C}^u)' D_{C^u}^{-1} (C^u - \hat{C}^u)$$

where,

$$D_{C^u} = (H' \Omega_{cu}^{-1} H + \Omega_{zu}^{-1})^{-1}$$

$$\hat{C}^u = D_{C^u} (H' \Omega_{cu}^{-1} \alpha_{cu} + \Omega_{zu}^{-1} N^{zu})$$

The candidate draws are sampled from $N(\hat{C}^u, D_{C^u})$ using the precision based algorithm.

Following similar logic,

$$N(\hat{C}^r, D_{C^r})$$

$$D_{C^r} = (H' \Omega_{cr}^{-1} H + \Omega_{zr}^{-1})^{-1}$$

$$\hat{C}^r = D_{C^r} (H' \Omega_{cr}^{-1} \alpha_{cr} + \Omega_{zr}^{-1} N^{zr})$$

where,

$$n_t^{zr} = Z_t^r - \beta^r r^*$$

$$N^{zr} = (n_1^{zr}, n_2^{zr}, \dots, n_T^{zr})'$$

$$\Omega_{zr} = \text{diag}(\sigma_{zr}^2, \dots, \sigma_{zr}^2)$$

$$N(\hat{C}^\pi, D_{C^\pi})$$

$$D_{C^\pi} = (H' \Omega_{c\pi}^{-1} H + \Omega_{z\pi}^{-1})^{-1}$$

$$\hat{C}^\pi = D_{C^\pi} (H' \Omega_{c\pi}^{-1} \alpha_{c\pi} + \Omega_{z\pi}^{-1} N^{z\pi})$$

where,

$$n_t^{z\pi} = Z_t^\pi - \beta^\pi \pi^*$$

$$N^{z\pi} = (n_1^{z\pi}, n_2^{z\pi}, \dots, n_T^{z\pi})'$$

$$\Omega_{z\pi} = \text{diag}(\sigma_{z\pi}^2, \dots, \sigma_{z\pi}^2)$$

$$N(\hat{C}^g, D_{C^g})$$

$$D_{C^g} = (H' \Omega_{cg}^{-1} H + \Omega_{zg}^{-1})^{-1}$$

$$\hat{C}^g = D_{C^g} (H' \Omega_{cg}^{-1} \alpha_{cg} + \Omega_{zg}^{-1} N^{zg})$$

where,

$$n_t^{zg} = Z_t^g + \beta^g \alpha_g - \beta^g g d p^*$$

$$N^{zg} = (n_1^{zg}, n_2^{zg}, \dots, n_T^{zg})'$$

$$\Omega_{zg} = \text{diag}(\sigma_{zg}^2, \dots, \sigma_{zg}^2)$$

$$\alpha_g = (g d p_0^*, 0, 0, \dots, 0)'$$

Step 15. Derive the conditional distribution $p(D|Y, \bullet)$

Given the posterior draws of r^* , ζ , and g^* , the posterior draw for D is constructed as,

$$D = r^* - \zeta g^* \tag{C.81}$$

Step 16. Derive the conditional distribution $p(\theta|Y, \bullet)$

There are 40 parameters in the vector θ . These parameters are drawn in 38 separate blocks using standard regression procedures. Following similar notation to Chan, Koop, and Potter (2016), we denote θ_{-x} to refer all parameters in θ except the parameter x .

Substep 16.1 Derive the conditional distribution $p(\rho^u|Y, \bullet)$

Given the stationary constraints, $\rho_1^u + \rho_2^u < 1$, $\rho_2^u - \rho_1^u < 1$, and $|\rho_2^u| < 1$

$\rho^u = (\rho_1^u, \rho_2^u)'$ is a bivariate truncated normal. To obtain draws from this truncated normal distribution, ARMH sampling algorithm is applied to the candidate draws from the proposal density, $N(\hat{\rho}^u, D_{\rho^u})$.

$$D_{\rho u} = (V_{\rho u}^{-1} + X_u' X_u / \sigma_u^2)^{-1}$$

$$\hat{\rho}^u = D_{\rho u} (V_{\rho u}^{-1} \rho_0^u + X_u' (\tilde{u} - \phi^u \text{ogap}) / \sigma_u^2)$$

where,

$V_{\rho u}^{-1}$ is the prior variance and ρ_0^u is the prior mean,

$$X_u = \begin{pmatrix} \tilde{u}_0 & \tilde{u}_{-1} \\ \tilde{u}_1 & \tilde{u}_0 \\ \vdots & \\ \tilde{u}_{T-1} & \tilde{u}_{T-2} \end{pmatrix}$$

Substep 16.2 Derive the conditional distribution $p(\sigma_u^2 | Y, \bullet)$

$p(\sigma_u^2 | Y, \bullet)$ is a standard inverse-Gamma density,

$$p(\sigma_u^2 | Y, \bullet) \sim IG(\nu_{u0} + \frac{T}{2}, S_{u0} + \frac{1}{2} \sum_{t=1}^T (\tilde{u}_t - \rho_1^u \tilde{u}_{t-1} - \rho_2^u \tilde{u}_{t-2} - \phi^u \text{ogap}_t)^2)$$

Substep 16.3 Derive the conditional distribution $p(\phi^u | Y, \bullet)$

Given the constraint $\phi^u < 0$, the conditional distribution $p(\phi^u | Y, \bullet)$ is a truncated normal density. The candidate draws are sampled from the proposal distribution $N(\hat{\phi}^u, D_{\phi u})$ using the precision based algorithm, and simple Accept-Reject step is applied to the candidate draws.

Rewrite the unemployment rate (gap) measurement equation in matrix notation as

$$Y^\phi = \phi^u \text{ogap} + \varepsilon^u \quad \varepsilon^u \sim N(0, \sigma_u^2) \tag{C.82}$$

where,

$$y_t^\phi = \tilde{u}_t - \rho_1^u \tilde{u}_{t-1} - \rho_2^u \tilde{u}_{t-2}$$

$$Y^\phi = (y_1^\phi, \dots, y_T^\phi)'$$

$$D_{\phi u} = (V_{\phi u}^{-1} + \text{ogap}' \text{ogap} / \sigma_u^2)^{-1}$$

$$\hat{\phi}^u = D_{\phi u} (V_{\phi u}^{-1} \phi_0^u + \text{ogap}' Y^\phi / \sigma_u^2)$$

where,

$V_{\phi_u}^{-1}$ is the prior variance and ϕ_0^u is the prior mean,

Substep 16.4 Derive the conditional distribution $p(\sigma_{u^*}^2|Y, \bullet)$

$p(\sigma_{u^*}^2|Y, \bullet)$ is a non-standard density because U^* is a bounded random walk,

$$\log p(\sigma_{u^*}^2|Y, \bullet) \propto -(\nu_{u^*0} + 1) \log \sigma_{u^*}^2 - \frac{S_{u^*0}}{\sigma_{u^*}^2} - \frac{T-1}{2} \log \sigma_{u^*}^2 - \frac{1}{2\sigma_{u^*}^2} \sum_{t=2}^T (U_t^* - U_{t-1}^*)^2 + g_{u^*}(U^*, \sigma_{u^*}^2)$$

The candidate draws from $p(\sigma_{u^*}^2|Y, \bullet)$ are obtained via the MH step with the proposal density

$$IG(\nu_{u^*0} + \frac{T-1}{2}, S_{u^*0} + \frac{1}{2} \sum_{t=2}^T (U_t^* - U_{t-1}^*)^2)$$

Substep 16.5 Derive the conditional distribution $p(\beta^u|Y, \bullet)$

Candidate draws are sampled from $N(\hat{\beta}^u, D_{\beta_u})$ using the precision based algorithm.

where,

$$D_{\beta_u} = (V_{\beta_u}^{-1} + U^{*\prime} \Omega_{zu}^{-1} U^*)^{-1}$$

$$\hat{\beta}^u = D_{\beta_u} (V_{\beta_u}^{-1} \beta_0^u + U^{*\prime} \Omega_{zu}^{-1} J^{zu})$$

$$j_t^{zu} = Z_t^u - C_t^u$$

$$J^{zu} = (j_1^{zu}, \dots, j_T^{zu})'$$

$V_{\beta_u}^{-1}$ is the prior variance and β_0^u is the prior mean for β^u

Substep 16.6 Derive the conditional distribution $p(\sigma_{zu}^2|Y, \bullet)$

$p(\sigma_{zu}^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{zu}^2|Y, \bullet) \sim IG(\nu_{zu0} + \frac{T}{2}, S_{zu0} + \frac{1}{2} \sum_{t=1}^T (Z_t^u - C_t^u - \beta^u U^*)^2)$$

Substep 16.7 Derive the conditional distribution $p(\sigma_{cu}^2|Y, \bullet)$

$p(\sigma_{cu}^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{cu}^2|Y, \bullet) \sim IG(\nu_{cu0} + \frac{T-1}{2}, S_{cu0} + \frac{1}{2} \sum_{t=2}^T (C_t^u - C_{t-1}^u)^2)$$

Substep 16.8 Derive the conditional distribution $p(\sigma_{gdp^*}^2|Y, \bullet)$

$p(\sigma_{gdp^*}^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{gdp^*}^2|Y, \bullet) \sim IG(\nu_{gdp^*0} + \frac{T-1}{2}, S_{gdp^*0} + (gdp^* - \alpha_{gdp^*})' * H_2 H_2 * (gdp^* - \alpha_{gdp^*})/2)$$

where (although they are defined above but for convenience we redefine them),

$$\alpha_{gdp^*} = \begin{pmatrix} gdp_0^* + \Delta gdp_0^* \\ -gdp_0^* \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}$$

H_2 is a band matrix with unit determinant and hence is invertible.

Substep 16.9 Derive the conditional distribution $p(\rho^g|Y, \bullet)$

Given the stationary constraints, $\rho_1^g + \rho_2^g < 1$, $\rho_2^g - \rho_1^g < 1$, and $|\rho_2^g| < 1$

$\rho^g = (\rho_1^g, \rho_2^g)'$ is a bivariate truncated normal. To obtain draws from this truncated normal distribution, ARMH sampling algorithm is applied to the candidate draws from the proposal density, $N(\hat{\rho}^g, D_{\rho^g})$.

$$D_{\rho g} = (V_{\rho g}^{-1} + X'_{\rho g} X_{\rho g} / \sigma_{ogap}^2)^{-1}$$

$$\hat{\rho}^g = D_{\rho g} (V_{\rho g}^{-1} \rho_0^g + X'_{\rho g} Y_{ogap} / \sigma_{ogap}^2)$$

where,

$V_{\rho g}^{-1}$ is the prior variance and ρ_0^g is the prior mean,

$$X_{\rho g} = \begin{pmatrix} 0 & 0 \\ ogap_1 & 0 \\ ogap_2 & ogap_1 \\ \vdots & \\ ogap_{T-1} & ogap_{T-2} \end{pmatrix}$$

$$y_t^{ogap} = ogap_t - a^r (r_t - r_{t-1}) - \lambda^g \tilde{u}_t$$

$$Y_{ogap} = (y_1^{ogap}, \dots, y_T^{ogap})'$$

Substep 16.10 Derive the conditional distribution $p(a^r|Y, \bullet)$

Candidate draws are sampled from $N(\hat{a}^r, D_{ar})$ using the precision based algorithm.

where,

$$D_{ar} = (V_{ar}^{-1} + X'_{ar} \Omega_{ogap}^{-1} X_{ar})^{-1}$$

$$\hat{a}^r = D_{ar} (V_{ar}^{-1} a_0^r + X'_{ar} \Omega_{ogap}^{-1} J^{ar})$$

$$j_t^{ar} = ogap_t - \rho_1^g ogap_{t-1} - \rho_2^g ogap_{t-2} - \lambda^g \tilde{u}_t$$

$$J^{ar} = (j_1^{ar}, \dots, j_T^{ar})'$$

$$X_{ar} = (\tilde{r}_1, \dots, \tilde{r}_T)'$$

$$\tilde{r}_t = r_t - r_t^*$$

V_{ar}^{-1} is the prior variance and a_0^r is the prior mean for a^r

Substep 16.11 Derive the conditional distribution $p(\lambda^g|Y, \bullet)$

Given the constraint $\lambda^g < 0$, the conditional distribution $p(\lambda^g|Y, \bullet)$ is a truncated normal density. The candidate draws are sampled from the proposal distribution $N(\hat{\lambda}^g, D_{\lambda g})$ using the precision based algorithm, and simple Accept-Reject step is applied to the candidate draws.

where,

$$D_{\lambda_g} = (V_{\lambda_g}^{-1} + X_u' \Omega_{ogap}^{-1} X_u)^{-1}$$

$$\hat{\lambda}^g = D_{\lambda_g} (V_{\lambda_g}^{-1} \lambda_0^g + X_u' \Omega_{ogap}^{-1} B^g)$$

$$b_t^g = ogap_t - \rho_1^g ogap_{t-1} - \rho_2^g ogap_{t-2} - a^r \tilde{r}_t$$

$$B^g = (b_1^g, \dots, b_T^g)'$$

$$X_u = \text{diag}(\tilde{u}_1, \dots, \tilde{u}_T)'$$

$$\tilde{r}_t = r_t - r_t^*$$

$V_{\lambda_g}^{-1}$ is the prior variance and λ_0^g is the prior mean for λ^g

Substep 16.12 Derive the conditional distribution $p(\sigma_{ogap}^2 | Y, \bullet)$

$p(\sigma_{ogap}^2 | Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{ogap}^2 | Y, \bullet) \sim IG(\nu_{ogap0} + \frac{T}{2}, S_{ogap0} + \frac{1}{2} \sum_{t=1}^T (ogap_t - \rho_1^g ogap_{t-1} - \rho_2^g ogap_{t-2} - \lambda^g \tilde{u}_t - a^r \tilde{r}_t)^2)$$

Substep 16.13 Derive the conditional distribution $p(\sigma_{zg}^2 | Y, \bullet)$

$p(\sigma_{zg}^2 | Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{zg}^2 | Y, \bullet) \sim IG(\nu_{zg0} + \frac{T}{2}, S_{zg0} + \frac{1}{2} \sum_{t=1}^T (Z_t^g - C_t^g - \beta^g gdp_{t-1}^* + \beta^g gdp_t^*)^2)$$

Substep 16.14 Derive the conditional distribution $p(\sigma_{cg}^2 | Y, \bullet)$

$p(\sigma_{cg}^2 | Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{cg}^2|Y, \bullet) \sim IG(\nu_{cg0} + \frac{T-1}{2}, S_{cg0} + \frac{1}{2} \sum_{t=2}^T (C_t^g - C_{t-1}^g)^2)$$

Substep 16.15 Derive the conditional distribution $p(\beta^g|Y, \bullet)$

Candidate draws are sampled from $N(\hat{\beta}^g, D_{\beta g})$ using the precision based algorithm.

where,

$$D_{\beta g} = (V_{\beta g}^{-1} + (Hgd p^* - \alpha_g)' \Omega_{zg}^{-1} (Hgd p^* - \alpha_g))^{-1}$$

$$\hat{\beta}^g = D_{\beta g} (V_{\beta g}^{-1} \beta_0^g + (Hgd p^* - \alpha_g) \Omega_{zg}^{-1} J^{zg})$$

$$j_t^{zg} = Z_t^g - C_t^g$$

$$J^{zg} = (j_1^{zg}, \dots, j_T^{zg})'$$

$$\alpha_g = (gd p_0^*, 0, 0, \dots, 0)'$$

$V_{\beta g}^{-1}$ is the prior variance and β_0^g is the prior mean for β^g

Substep 16.16 Derive the conditional distribution $p(\rho^p|Y, \bullet)$

Given the stationary constraint, $|\rho^p| < 1$

ρ^p is a truncated normal. To obtain draws from this truncated normal distribution, AR sampling step is applied to the candidate draws from the proposal density, $N(\hat{\rho}^p, D_{\rho p})$.

$$D_{\rho p} = (V_{\rho p}^{-1} + X'_{prod} \Omega_P^{-1} X_{prod})^{-1}$$

$$\hat{\rho}^p = D_{\rho p} (V_{\rho p}^{-1} \rho_0^p + X'_{prod} \Omega_P^{-1} Y^{prod})$$

where,

$V_{\rho p}^{-1}$ is the prior variance and ρ_0^p is the prior mean,

$$\tilde{p}_t = P_t - P_t^*$$

$$X_{prod} = (\tilde{p}_0, \dots, \tilde{p}_{T-1})'$$

$$y_t^{prod} = \tilde{p}_t - \lambda_t^p \tilde{u}_t$$

$$Y^{prod} = (y_1^{prod}, \dots, y_T^{prod})'$$

Substep 16.17 Derive the conditional distribution $p(\sigma_{hp}^2|Y, \bullet)$

$p(\sigma_{hp}^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{hp}^2|Y, \bullet) \sim IG(\nu_{hp0} + \frac{T-1}{2}, S_{hp0} + \frac{1}{2} \sum_{t=2}^T (h_t^p - h_{t-1}^p)^2)$$

Substep 16.18 Derive the conditional distribution $p(\sigma_{p^*}^2|Y, \bullet)$

$p(\sigma_{p^*}^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{p^*}^2|Y, \bullet) \sim IG(\nu_{p^*0} + \frac{T-1}{2}, S_{p^*0} + \frac{1}{2} \sum_{t=2}^T (P_t^* - P_{t-1}^*)^2)$$

Substep 16.19 Derive the conditional distribution $p(\sigma_{\lambda^\pi}^2|Y, \bullet)$

$p(\sigma_{\lambda^\pi}^2|Y, \bullet)$ is a non-standard density because of the constraints on λ^π ,

$$\log p(\sigma_{\lambda^\pi}^2|Y, \bullet) \propto -(\nu_{\lambda^\pi 0} + 1) \log \sigma_{\lambda^\pi}^2 - \frac{S_{\lambda^\pi 0}}{\sigma_{\lambda^\pi}^2} - \frac{T-1}{2} \log \sigma_{\lambda^\pi}^2 - \frac{1}{2\sigma_{\lambda^\pi}^2} \sum_{t=2}^T (\lambda_t^\pi - \lambda_{t-1}^\pi)^2 + g_{\lambda^\pi}(\lambda^\pi, \sigma_{\lambda^\pi}^2)$$

The candidate draws from $p(\sigma_{\lambda^\pi}^2|Y, \bullet)$ are obtained via the MH step with the proposal density

$$IG(\nu_{\lambda^\pi 0} + \frac{T-1}{2}, S_{\lambda^\pi 0} + \frac{1}{2} \sum_{t=2}^T (\lambda_t^\pi - \lambda_{t-1}^\pi)^2)$$

Substep 16.20 Derive the conditional distribution $p(\sigma_{\rho^\pi}^2|Y, \bullet)$

$p(\sigma_{\rho\pi}^2|Y, \bullet)$ is a non-standard density because of the constraints on ρ^π ,

$$\log p(\sigma_{\rho\pi}^2|Y, \bullet) \propto -(\nu_{\rho\pi 0} + 1)\log \sigma_{\rho\pi}^2 - \frac{S_{\rho\pi 0}}{\sigma_{\rho\pi}^2} - \frac{T-1}{2}\log \sigma_{\rho\pi}^2 - \frac{1}{2\sigma_{\rho\pi}^2} \sum_{t=2}^T (\rho_t^\pi - \rho_{t-1}^\pi)^2 + g_{\rho\pi}(\rho^\pi, \sigma_{\rho\pi}^2)$$

The candidate draws from $p(\sigma_{\rho\pi}^2|Y, \bullet)$ are obtained via the MH step with the proposal density

$$IG(\nu_{\rho\pi 0} + \frac{T-1}{2}, S_{\rho\pi 0} + \frac{1}{2} \sum_{t=2}^T (\rho_t^\pi - \rho_{t-1}^\pi)^2)$$

Substep 16.21 Derive the conditional distribution $p(\sigma_{h\pi}^2|Y, \bullet)$

$p(\sigma_{h\pi}^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{h\pi}^2|Y, \bullet) \sim IG(\nu_{h\pi 0} + \frac{T-1}{2}, S_{h\pi 0} + \frac{1}{2} \sum_{t=2}^T (h_t^\pi - h_{t-1}^\pi)^2)$$

Substep 16.22 Derive the conditional distribution $p(\sigma_{\pi^*}^2|Y, \bullet)$

$p(\sigma_{\pi^*}^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{\pi^*}^2|Y, \bullet) \sim IG(\nu_{\pi^* 0} + \frac{T-1}{2}, S_{\pi^* 0} + \frac{1}{2} \sum_{t=2}^T (\pi_t^* - \pi_{t-1}^*)^2)$$

Substep 16.23 Derive the conditional distribution $p(\sigma_{z\pi}^2|Y, \bullet)$

$p(\sigma_{z\pi}^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{z\pi}^2|Y, \bullet) \sim IG(\nu_{z\pi 0} + \frac{T}{2}, S_{z\pi 0} + \frac{1}{2} \sum_{t=1}^T (Z_t^\pi - C_t^\pi - \beta^\pi \pi^*)^2)$$

Substep 16.24 Derive the conditional distribution $p(\sigma_{c\pi}^2|Y, \bullet)$

$p(\sigma_{c\pi}^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{c\pi}^2|Y, \bullet) \sim IG(\nu_{c\pi 0} + \frac{T-1}{2}, S_{c\pi 0} + \frac{1}{2} \sum_{t=2}^T (C_t^\pi - C_{t-1}^\pi)^2)$$

Substep 16.25 Derive the conditional distribution $p(\beta^\pi|Y, \bullet)$

Candidate draws are sampled from $N(\hat{\beta}^\pi, D_{\beta\pi})$ using the precision based algorithm.

where,

$$D_{\beta\pi} = (V_{\beta\pi}^{-1} + \pi^{*\prime} \Omega_{z\pi}^{-1} \pi^*)^{-1}$$

$$\hat{\beta}^\pi = D_{\beta\pi} (V_{\beta\pi}^{-1} \beta_0^\pi + \pi^{*\prime} \Omega_{z\pi}^{-1} J^{z\pi})$$

$$j_t^{z\pi} = Z_t^\pi - C_t^\pi$$

$$J^{z\pi} = (j_1^{z\pi}, \dots, j_T^{z\pi})'$$

$V_{\beta\pi}^{-1}$ is the prior variance and β_0^π is the prior mean for β^π

Substep 16.26 Derive the conditional distribution $p(\sigma_{w*}^2|Y, \bullet)$

$p(\sigma_{w*}^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{w*}^2|Y, \bullet) \sim IG(\nu_{w*0} + \frac{T-1}{2}, S_{w*0} + \frac{1}{2} \sum_{t=2}^T (w_t^* - \pi_t^* - P_t^*)^2)$$

Substep 16.27 Derive the conditional distribution $p(\sigma_{hw}^2|Y, \bullet)$

$p(\sigma_{hw}^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{hw}^2|Y, \bullet) \sim IG(\nu_{hw0} + \frac{T-1}{2}, S_{hw0} + \frac{1}{2} \sum_{t=2}^T (h_t^w - h_{t-1}^w)^2)$$

Substep 16.28 Derive the conditional distribution $p(\sigma_{\rho w}^2|Y, \bullet)$

$p(\sigma_{\rho w}^2|Y, \bullet)$ is a non-standard density because of the constraints on ρ^w ,

$$\log p(\sigma_{\rho w}^2|Y, \bullet) \propto -(\nu_{\rho w0}+1)\log \sigma_{\rho w}^2 - \frac{S_{\rho w0}}{\sigma_{\rho w}^2} - \frac{T-1}{2}\log \sigma_{\rho w}^2 - \frac{1}{2\sigma_{\rho w}^2} \sum_{t=2}^T (\rho_t^w - \rho_{t-1}^w)^2 + g_{\rho w}(\rho^w, \sigma_{\rho w}^2)$$

The candidate draws from $p(\sigma_{\rho w}^2|Y, \bullet)$ are obtained via the MH step with the proposal density

$$IG(\nu_{\rho w0} + \frac{T-1}{2}, S_{\rho w0} + \frac{1}{2} \sum_{t=2}^T (\rho_t^w - \rho_{t-1}^w)^2)$$

Substep 16.29 Derive the conditional distribution $p(\sigma_{\lambda w}^2|Y, \bullet)$

$p(\sigma_{\lambda w}^2|Y, \bullet)$ is a non-standard density because of the constraints on λ^w ,

$$\log p(\sigma_{\lambda w}^2|Y, \bullet) \propto -(\nu_{\lambda w0}+1)\log \sigma_{\lambda w}^2 - \frac{S_{\lambda w0}}{\sigma_{\lambda w}^2} - \frac{T-1}{2}\log \sigma_{\lambda w}^2 - \frac{1}{2\sigma_{\lambda w}^2} \sum_{t=2}^T (\lambda_t^w - \lambda_{t-1}^w)^2 + g_{\lambda w}(\lambda^w, \sigma_{\lambda w}^2)$$

The candidate draws from $p(\sigma_{\lambda w}^2|Y, \bullet)$ are obtained via the MH step with the proposal density

$$IG(\nu_{\lambda w0} + \frac{T-1}{2}, S_{\lambda w0} + \frac{1}{2} \sum_{t=2}^T (\lambda_t^w - \lambda_{t-1}^w)^2)$$

Substep 16.30 Derive the conditional distribution $p(\sigma_{\kappa w}^2|Y, \bullet)$

The candidate draws are obtained from

$$IG(\nu_{\kappa w0} + \frac{T-1}{2}, S_{\kappa w0} + \frac{1}{2} \sum_{t=2}^T (\kappa_t^w - \kappa_{t-1}^w)^2)$$

Substep 16.31 Derive the conditional distribution $p(\rho^i|Y, \bullet)$

Given the constraint $|\rho^i| < 1$, the conditional distribution $p(\rho^i|Y, \bullet)$ is a truncated normal

density. The candidate draws are sampled from the proposal distribution $N(\hat{\rho}^i, D_{\rho^i})$ using the precision based algorithm, and simple Accept-Reject step is applied to the candidate draws.

where,

$$D_{\rho^i} = (V_{\rho^i}^{-1} + X'_{\rho^i} \Omega_i^{-1} X_{\rho^i})^{-1}$$

$$\hat{\rho}^i = D_{\rho^i} (V_{\rho^i}^{-1} \rho_0^i + X'_{\rho^i} \Omega_i^{-1} M^{\rho^i})$$

$$m_t^{\rho^i} = i_t - \pi_t^* - r_t^* - \lambda^i \tilde{u}_t - \kappa^i \tilde{\pi}_t$$

$$M^{\rho^i} = (m_1^{\rho^i}, \dots, m_T^{\rho^i})'$$

$$X_{\rho^i} = (i_0 - \pi_0^* - r_0^*, \dots, i_{T-1} - \pi_{T-1}^* - r_{T-1}^*)'$$

$V_{\rho^i}^{-1}$ is the prior variance and ρ_0^i is the prior mean for ρ^i

Substep 16.32 Derive the conditional distribution $p(\lambda^i|Y, \bullet)$

The candidate draws are sampled from the proposal distribution $N(\hat{\lambda}^i, D_{\lambda^i})$ using the precision based algorithm.

where,

$$D_{\lambda^i} = (V_{\lambda^i}^{-1} + X'_{\lambda^i} \Omega_i^{-1} X_{\lambda^i})^{-1}$$

$$\hat{\lambda}^i = D_{\lambda^i} (V_{\lambda^i}^{-1} \lambda_0^i + X'_{\lambda^i} \Omega_i^{-1} M^{\lambda^i})$$

$$m_t^{\lambda^i} = i_t - \pi_t^* - r_t^* - \rho^i (i_{t-1} - \pi_{t-1}^* - r_{t-1}^*) - \kappa^i \tilde{\pi}_t$$

$$M^{\lambda^i} = (m_1^{\lambda^i}, \dots, m_T^{\lambda^i})'$$

$$X_{\lambda^i} = (\tilde{u}_1, \dots, \tilde{u}_T)'$$

$V_{\lambda^i}^{-1}$ is the prior variance and λ_0^i is the prior mean for λ^i

Substep 16.33 Derive the conditional distribution $p(\kappa^i|Y, \bullet)$

The candidate draws are sampled from the proposal distribution $N(\hat{\kappa}^i, D_{\kappa^i})$ using the pre-

cision based algorithm.

where,

$$D_{\kappa i} = (V_{\kappa i}^{-1} + X'_{\kappa i} \Omega_i^{-1} X_{\kappa i})^{-1}$$

$$\hat{\kappa}^i = D_{\kappa i} (V_{\kappa i}^{-1} \kappa_0^i + X'_{\kappa i} \Omega_i^{-1} M^{\kappa i})$$

$$m_t^{\kappa i} = i_t - \pi_t^* - r_t^* - \rho^i (i_{t-1} - \pi_{t-1}^* - r_{t-1}^*) - \lambda^i \tilde{u}_t$$

$$M^{\kappa i} = (m_1^{\kappa i}, \dots, m_T^{\kappa i})'$$

$$X_{\kappa i} = (\tilde{\pi}_1, \dots, \tilde{\pi}_T)'$$

$V_{\kappa i}^{-1}$ is the prior variance and κ_0^i is the prior mean for κ^i

Substep 16.34 Derive the conditional distribution $p(\sigma_{hi}^2 | Y, \bullet)$

$p(\sigma_{hi}^2 | Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{hi}^2 | Y, \bullet) \sim IG(\nu_{hi0} + \frac{T-1}{2}, S_{hi0} + \frac{1}{2} \sum_{t=2}^T (h_t^i - h_{t-1}^i)^2)$$

Substep 16.35 Derive the conditional distribution $p(\sigma_{zr}^2 | Y, \bullet)$

$p(\sigma_{zr}^2 | Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{zr}^2 | Y, \bullet) \sim IG(\nu_{zr0} + \frac{T}{2}, S_{zr0} + \frac{1}{2} \sum_{t=1}^T (Z_t^r - C_t^r - \beta^r r_t^*)^2)$$

Substep 16.36 Derive the conditional distribution $p(\sigma_{cr}^2 | Y, \bullet)$

$p(\sigma_{cr}^2 | Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_{cr}^2|Y, \bullet) \sim IG(\nu_{cr0} + \frac{T-1}{2}, S_{cr0} + \frac{1}{2} \sum_{t=2}^T (C_t^r - C_{t-1}^r)^2)$$

Substep 16.37 Derive the conditional distribution $p(\beta^r|Y, \bullet)$

Candidate draws are sampled from $N(\hat{\beta}^r, D_{\beta^r})$ using the precision based algorithm.

where,

$$D_{\beta^r} = (V_{\beta^r}^{-1} + r^{*'} \Omega_{z^r}^{-1} r^*)^{-1}$$

$$\hat{\beta}^r = D_{\beta^r} (V_{\beta^r}^{-1} \beta_0^r + r^{*'} \Omega_{z^r}^{-1} J^{z^r})$$

$$j_t^{z^r} = Z_t^r - C_t^r$$

$$J^{z^r} = (j_1^{z^r}, \dots, j_T^{z^r})'$$

$V_{\beta^r}^{-1}$ is the prior variance and β_0^r is the prior mean for β^r

Substep 16.38 Derive the conditional distribution $p(\sigma_d^2|Y, \bullet)$

$p(\sigma_d^2|Y, \bullet)$ is a standard inverse-Gamma density,

Candidate draws are sampled from

$$p(\sigma_d^2|Y, \bullet) \sim IG(\nu_{d0} + \frac{T-1}{2}, S_{d0} + \frac{1}{2} \sum_{t=2}^T (D_t - D_{t-1})^2)$$

C1.d Marginal likelihood computation

Bayesian model comparison is based on the marginal likelihood (marginal data density) metric. In computing marginal likelihood for various models, we use the approach proposed by CCK, which decomposes the marginal density of the data (e.g., inflation) into the product of predictive likelihoods. This approach allows us to separately compute marginal data density for each variable of interest: inflation, nominal wages, interest rate, real GDP, the unemployment rate, and labor productivity. The variable specific marginal densities prove quite useful because it allows for deeper insights about the source of the deficiencies, which in turn helps differentiate models at a more disaggregated level.

Specifically, marginal data density of the variables of interest is computed as follows,

$$p(y^j | X_i^j, M_i) = \prod_{t=3}^T p(y_t^j | y_{1:t-1}^j, X_{1:t,i}^j, M_i) \quad (\text{C.83})$$

where, j = PCE inflation (π), unemployment rate (ur), real GDP (gdp), labor productivity ($prod$), nominal wage inflation ($wage$), nominal short-term interest rate (int);

$p(y_t^j | y_{1:t-1}^j, X_{1:t,i}^j, M_i)$ is the predictive likelihood for variable j , and X_i^j is set of covariates that influences variable j in model M_i . For example, in the case of short-term interest rate, the covariates in the Base model include ur , π , gdp , and survey data. Whereas, in the Base-NoSurv model, the covariates will not include the survey data.

To compute the overall marginal data density of $Y = (y^\pi, y^{ur}, y^{gdp}, y^{prod}, y^{wage}, y^{int})'$ for model M_i ,

$$\begin{aligned} p(Y | X_i, M_i) &= p(y^\pi | X_i^\pi, M_i) \times p(y^{ur} | X_i^{ur}, M_i) \times p(y^{gdp} | X_i^{gdp}, M_i) \dots \\ &\times p(y^{prod} | X_i^{prod}, M_i) \times p(y^{wage} | X_i^{wage}, M_i) \times p(y^{int} | X_i^{int}, M_i) \end{aligned} \quad (\text{C.84})$$

C2. Prior Sensitivity Analysis

In the chapter, we noted that our prior settings are similar to those used in CKP, CCK, Gonzalez-Astudillo and Laforte (2020). As discussed in CCK, UC models with several unobserved variables, such as the one developed in this chapter, require informative priors. That said, our priors settings for most variables are only slightly informative. The use of inequality restrictions on some parameters such as the Phillips curve, persistence, bounds on u-star could be viewed as additional sources of information that eliminates the need for tight priors, something also noted by CKP. The parameters for which there is a strong agreement in the empirical literature on their values, such as the Taylor-rule equation parameters, we use relatively tight priors, such that prior distributions are centered on prior means with small variance. So besides

the prior on the Taylor-rule equation parameters, all other prior settings are taken from related papers.

Here, we examine the sensitivity of loosening the priors on the variances of the shocks to the pi-star, p-star, u-star, and r-star (i.e., for the process D). Specifically, we double the prior mean of the variances from 0.01 to 0.03. Table C2 reports the posterior estimates. The top panel reports the posterior estimates from the main text to facilitate easy comparison, and the bottom panel reports the posterior estimates of re-running the models with these new prior values. The results for p-star are as expected. In the chapter, we noted that the prior view primarily shapes p-star, and we see that manifest here too; prior ($E(\sigma_{p^*}^2)$) and posterior ($E(\sigma_{p^*}^2|Data)$) are fairly identical. Similar evidence is seen in the case of r-star (i.e., D) for the Base-NoSurv model. For pi-star, u-star, and r-star (in the case of Base), the posterior mean estimates' differences between the two panels are small. In fact, the posterior mean estimates from the runs with looser prior are pushed toward values that are closer to the prior mean estimates used in the main chapter, lending credibility to our default prior settings used in the main chapter.

Table C2: Parameter Estimates: Comparison

Panel A: From the main chapter, where prior $E(\sigma_{\pi^*}^2) = E(\sigma_{u^*}^2) = E(\sigma_d^2) = 0.1^2$ and $E(\sigma_{p^*}^2) = 0.14^2$

Parameter	Parameter description	Posterior estimates					
		Base			Base-NoSurv		
		Mean	5%	95%	Mean	5%	95%
$\sigma_{\pi^*}^2$	Variance of the shocks to π^*	0.121 ²	0.100 ²	0.141 ²	0.127 ²	0.084 ²	0.182 ²
$\sigma_{p^*}^2$	Variance of the shocks to p^*	0.145 ²	0.111 ²	0.183 ²	0.141 ²	0.109 ²	0.176 ²
$\sigma_{u^*}^2$	Variance of the shocks to u^*	0.075 ²	0.064 ²	0.089 ²	0.084 ²	0.071 ²	0.100 ²
σ_d^2	Variance of the shocks to d	0.093 ²	0.077 ²	0.110 ²	0.114 ²	0.084 ²	0.148 ²

Panel B: Prior sensitivity, where prior $E(\sigma_{\pi^*}^2) = E(\sigma_{u^*}^2) = E(\sigma_d^2) = E(\sigma_{p^*}^2) = 0.173^2$

Parameter	Parameter description	Posterior estimates					
		Base			Base-NoSurv		
		Mean	5%	95%	Mean	5%	95%
$\sigma_{\pi^*}^2$	Variance of the shocks to π^*	0.143 ²	0.124 ²	0.163 ²	0.190 ²	0.145 ²	0.236 ²
$\sigma_{p^*}^2$	Variance of the shocks to p^*	0.172 ²	0.134 ²	0.214 ²	0.166 ²	0.130 ²	0.207 ²
$\sigma_{u^*}^2$	Variance of the shocks to u^*	0.102 ²	0.090 ²	0.115 ²	0.121 ²	0.103 ²	0.140 ²
σ_d^2	Variance of the shocks to d	0.122 ²	0.106 ²	0.140 ²	0.175 ²	0.136 ²	0.218 ²

C3. MCMC Convergence Diagnostics

In this section, we document the diagnostic properties of our MCMC algorithm in the Base and Base-NoSurv models (see figures C1 and C2). Following Primiceri (2005), Koop, Leon-Gonzalez, and Strachan (2010), and Korobilis (2017), we report the autocorrelation functions of the posterior draws (10th and 50th order sample autocorrelation), inefficiency factors (IFs), and convergence diagnostic (CD) of Geweke (1992).¹

One of the most common metrics examined to assess the efficiency of the MCMC sampler is to look at the autocorrelation function of the draws, which indicates how well the chain is mixing. Low autocorrelations are preferred to higher because the lower the autocorrelation, the closer the draws are to being independent and higher the efficiency of the algorithm. The plots shown in the top panel of the figures correspond to 10th and 50th order autocorrelations in the draws, and as can be seen, they indicate very low autocorrelation. In the case of 50th order autocorrelation, all of them indicate close to zero, and in the case of 10th order except for a couple most indicate correlation below 0.2.

The inefficiency factor related to the autocorrelation functions is the inverse of Geweke’s (1992) relative numerical efficiency measure (RNE). It is computed using the following formula, $(1 + 2 \sum_{i=1}^{\infty} \rho_i)$, where ρ_i refers to the $k - th$ order autocorrelation of the chain. The middle panel in the figures plots the IF for each of the parameters. The values lower than or close to 20 are considered desirable. As shown, in the case of the Base model, all the IFs are below 20, and most at below 10. Similarly, in the case of Base-NoSurv, except one, for all other parameters, IFs are below 20. (Note: IFs are computed using the default setting in LeSage’s toolbox: estimation of spectral density at frequency zero uses a tapered window of 4%)

As discussed in Koop, Leon-Gonzalez, and Strachan (2010), to assess whether the MCMC sampler has converged, a rough rule of thumb is to look at the CDs and see whether 95% of them are less than 2. If they are, then convergence is likely achieved. Based on the plots in figures C1 and C2 (third-panel), most CDs are within ± 2 . The very few that exceed 2 are only slightly larger than 2.

We also note that the results are fairly identical to the different initial conditions of the chain (picked randomly) and to a significantly lower number of simulations (and burn-in). For example, a run using 320K posterior draws with a burn-in of the first 300K and retaining all the remaining draws provide similar inference. However, the MCMC convergence properties favor higher simulations because it allows for a greater degree of thinning.

¹In computing some of these metrics, we have benefitted from the Matlab toolbox developed by James P. LeSage. Detailed explanation including intuition for these convergence diagnostics are provided in Koop (2006; page 67-68) and Chan, Koop, Poirier, and Tobias (2019; page 209).

Overall, these diagnostic measures provide us confidence in the good convergence properties of our MCMC algorithm in both the Base and Base-NoSurv models.

Figure C1: MCMC Diagnostics of Base Model

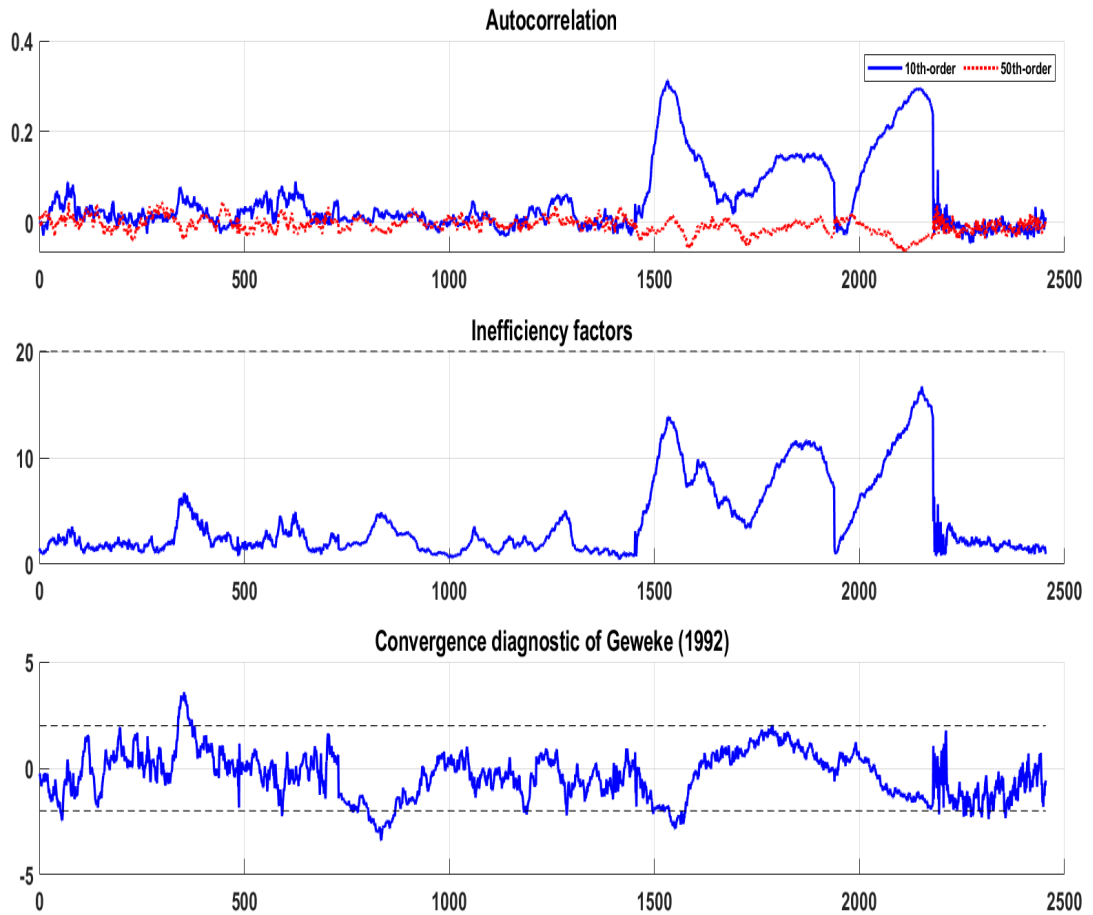
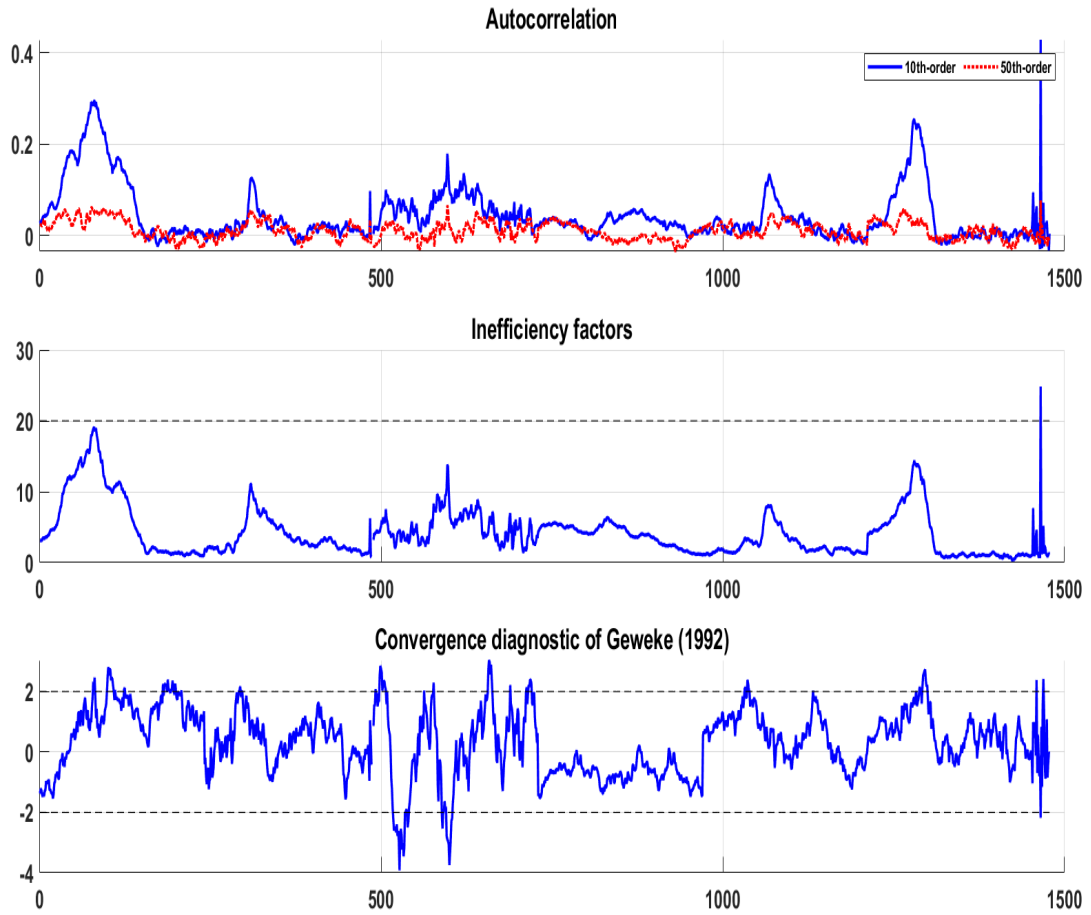


Figure C2: MCMC Diagnostics of Base-NoSurv Model



C4. Additional Forecasting Results: Base vs. Benchmarks

In this section we compare the real-time forecasting performance of our Base model to the outside benchmark models, which forecasting literature has shown to be useful forecasting devices. Specifically, we compare the accuracy of the inflation forecasts from our Base model to the following three models, UCSV of Stock and Watson (2007) [UCSV], Chan, Koop, and Potter (2016) [CKP], and Chan, Clark, and Koop (2018) [CCK]. We compare the accuracy of the unemployment rate forecasts from our Base model to the CKP, and the accuracy of the nominal wage inflation from the Base model to the UCSV model applied to the nominal wage inflation – motivated by Knotek (2015).

Table C3 presents the forecast evaluation results for headline PCE inflation, nominal wage inflation, and the unemployment rate. These results indicate following three observations. First, in terms of point forecast accuracy, inflation forecasts from all the four models considered are competitive to each other. There is some statistically significant evidence that the Base model is more accurate than UCSV at $h=12Q$. Regarding the density forecast accuracy, the Base model is more accurate than the UCSV but inferior to CCK, as the latter produces more precise intervals than the Base model. Second, in the case of nominal wage inflation, the Base model generates more accurate forecasts (both point and density) than UCSV, and the gains are statistically significant for the most part.

Third, the accuracy of the unemployment forecasts from the Base model is competitive to the CKP model statistically speaking, even though the relative numbers favor CKP. A closer inspection of the forecast errors reveals that the Base model, which incorporates survey forecasts of the unemployment rate, experienced significantly bigger misses than the CKP model around the Great Recession period. Outside of this period, the Base model is slightly more accurate than the CKP, and when combined with the Great Recession period, on the net, the much bigger misses of the Base model results in overall higher RMSE.

As illustrated in Tallman and Zaman (2020), just before and at the onset of the Great Recession, the survey participants projected relatively upbeat long-run forecasts of unemployment, which indicated a declining natural rate of unemployment. It was not until few months into the recession that survey participants recognized the extent of the labor market damage and began to revise their estimates of the long-run unemployment higher. Hence, models such as the Base model that take signals from the survey forecasts experienced big misses.

To sum up, we view these forecasting results as providing evidence supporting our Base model's competitive forecasting properties.

Table C3: Out-of-Sample Forecasting Performance: **Base vs. Benchmarks**

Full Sample (Recursive evaluation: 1999.Q1-2019.Q4)									
Point forecasting					Density forecasting				
	4Q	8Q	12Q	20Q		4Q	8Q	12Q	20Q
PCE Inflation									
Relative RMSE					Relative Log Score				
Base/UCSV	0.95	0.97	0.93*	0.96	Base - UCSV	0.013*	0.023*	0.028*	0.041*
Base/CCK	1.01	1.04	1.01	1.04*	Base - CCK	-0.018*	-0.030*	-0.046*	-0.058*
Base/CKP	0.98	0.99	0.97	1.02	Base - CKP	0.002	0.001	-0.003*	-0.008*
Nominal Wage									
Relative RMSE					Relative Log Score				
Base/UCSV	0.89*	0.87*	0.92	0.64	Base - UCSV	0.012	0.027*	0.037*	0.041*
Unemployment Rate									
Relative MSE					Relative Log Score				
Base/CKP	1.08	1.12	1.15	1.24	Base - CKP	0.001	0.000	-0.004	-0.007

Notes for Table: For variables PCE inflation and nominal wage (i.e., average hourly earnings), the forecasts and associated accuracy correspond to the quarterly annualized rate. Base forecast is defined as the Steady-State (SS) VAR forecast in which the steady states are assumed to be the estimates of the stars from the Base model. UCSV forecast corresponds to the forecast from the univariate unobserved component stochastic volatility model similar to Stock and Watson (2007). The model is used to construct forecasts of PCE inflation and nominal wage inflation. CCK forecast corresponds to the forecast from the bivariate unobserved component stochastic volatility model of Chan, Clark and Koop (2018). CKP forecast corresponds to the forecast from the bivariate unobserved component stochastic volatility model of Chan, Koop and Potter (2016), with the bounds on u-star fixed to values identical to the Base model. The left panel reports results for the point forecast accuracy (relative root mean squared errors) and the right panel reports the corresponding density forecast accuracy (mean of the relative log predictive score). The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *up to 10% significance level. The test statistics use two-sided standard normal critical values for horizons less than equal to 8 quarters, and two-sided t-statistics for horizons greater than 8 quarters.

C5. Additional Forecasting Results: SSBVAR, Base stars vs. Survey

In macroeconomic forecasting, research by Wright (2013) and Tallman and Zaman (2020), among others, show using workhorse Bayesian VAR models that the predictive performance boils down to good starting conditions (i.e., nowcasts) and terminal conditions (i.e., steady-states proxied by stars). Survey forecasts provide both nowcasts and long-run projections, whose accuracy has been shown by past research to be quite good. Wright (2019) emphasizes the desirable forecasting properties of the survey forecasts and highlights that econometric approaches utilizing survey projections are at the forecasting frontier, especially in inflation forecasting. Most empirical research on forecasting has focused on proposing methods to improve the accuracy of the nowcast estimates relative to survey nowcasts' accuracy, but only little effort has been dedicated to improving estimates of long-run projections. Hence, this chapter raises the natural curiosity in the usefulness of the stars' estimates from our modeling framework for macroeconomic forecasting using Bayesian VARs (via the imposition of steady-states).

To assess the efficacy of our star's estimates for the external VAR models, we perform a real-time out-of-sample forecasting evaluation similar to Wright (2013) and Tallman and Zaman (2020). These studies informed the time-varying steady-states for the steady-state (SS) BVAR using the long-run survey projections and found that doing so leads to significant accuracy gains. Accordingly, the design of our forecasting examination is as follows. We take the SSBVAR from Tallman and Zaman (2020) and perform three sets of recursive real-time out-of-sample forecasting runs. In the first run, we inform the steady-states for real GDP growth, PCE inflation, core PCE inflation, the unemployment rate, nominal wage inflation, and labor productivity growth using the long-run survey projections. For the latter two variables, we use the survey expectations from the SPF.² The forecasts from this run are denoted 'Survey' in table C4. In the second run, we repeat the exercise, but this time inform the steady-states using the real-time estimates of the stars from the Base-NoSurv model, denoted 'BaseNoSurv'. In the third run, we inform the steady-states using the real-time estimates of stars from the Base model, denoted 'Base'.

Each of the three forecasting runs is based on estimating the SSBVAR with a recursively expanding sample, i.e., the recursive execution uses an additional quarterly data point in the estimation. The SSBVAR is estimated with quarterly data beginning 1959Q2. The model consists of ten variables: (1) real GDP growth; (2) real consumption expenditures; (3) headline PCE inflation; (4) core PCE inflation; (5) labor productivity growth; (6) growth in average hourly earnings; (7) growth in payroll employment; (8) the unemployment rate; (9) the shadow federal funds rate; (10) and the risk spread, defined as the difference between the yield on 10-year Treasury Bond and yield on BAA Bond. The out-of-sample forecasting period spans

²In the case of nominal wage inflation, we construct an implied survey projection by adding the survey expectation of PCE inflation and productivity, both of which are obtained from the SPF.

1999Q1 through 2019Q4. The forecast accuracy (point and density) is computed from one-quarter ahead to 20 quarters out. Partly due to focus on the medium-term horizon and partly in the interest of space, we report accuracy metrics for four, eight, twelve, and twenty quarters ahead.

We evaluate the forecast accuracy using real-time data; specifically, we treat the “actual” as the third quarterly estimate. For instance, in the case of real GDP, the third estimate for 2018Q4 corresponds to the GDP data available in late 2019Q1. The point forecast accuracy is assessed using the root mean squared error (RMSE) metric, and the density forecast accuracy is assessed using either the continuous ranked probability score (CRPS). Forecasts with lower RMSE and CRPS are preferred. The statistical significance of the point and density forecast accuracy is gauged using the Diebold-Mariano and West test. The description of these tests is listed in the notes accompanying the tables reporting forecast accuracy.

Table C4 reports forecast evaluation results corresponding to this exercise. The left panel reports the point forecast accuracy results, while the right panel results for the density forecast accuracy. We evaluate and compare the point, and density forecast accuracies among the Base, BaseNoSurv, and Survey forecasts in a pairwise fashion. For each variable, the three rows report the relative RMSE (for point forecast accuracy) and the relative CRPS (for density forecast accuracy). The first row reports the RMSE of the Base relative to Survey, the second row reports the RMSE of BaseNoSurv to Survey, and the third row reports the RMSE of BaseNoSurv relative to Base. A model with a lower values of RMSE and CRPS is preferred to a model with higher values. These relative metrics indicate the following. First, for real GDP growth, statistically speaking, Survey outperforms both Base and BaseNoSurv. A closer inspection of the errors reveals that most of the gains of Survey over Base and BaseNoSurv are achieved over the post-Great Recession period.

As indicated in the figures plotting real-time estimates (see figure C5), starting in 2011 onwards, while both Base and Base-NoSurv have g -star falling sharply in the vicinity of 1.0%, the Survey has g -star falling only a little to 2.0%. This more rapid deceleration in g -star inferred by our models hurts the forecast accuracy of real GDP forecasts. This particular forecasting result suggests that our models misleadingly attribute a higher portion of the low GDP growth realizations in the post-Great Recession period to a trend decline in real GDP growth instead of cyclical fluctuations.

For headline PCE inflation, all three are competitive to each other, with some statistically significant gains in the density forecast accuracy of Base and Base-NoSurv over Survey. In the case of nominal wage inflation, both Base and Base-NoSurv generate forecasts that are substantially more accurate than Survey on average. The gains are statistically significant for the most part. In the case of labor productivity, while Base is more accurate than BaseNoSurv, both are inferior to the Survey. This result suggests that maybe bringing in survey information about productivity in the Base model may improve the econometric estimation of p -star.

For the unemployment rate, both Base and BaseNoSurv are inferior to the Survey, but

the gains are not statistically significant for the most part. The SSBVAR with steady-states informed from the Base model generates more accurate unemployment forecasts than Base-NoSurv, but the forecast gains are statistically significant only for the very long horizons. In the case of the shadow federal funds rate, both Base and Survey are competitive but are inferior to BaseNoSurv for $h=4Q$ and $h=8Q$.

Overall, these forecasting results lend credibility to our stars' estimates (except for g-star) in their use to inform steady-states for VAR forecasting models. We also note the results of this section lend support to the survey projections in their use as proxies for stars, something also documented by Tallman and Zaman (2020), among others.

The fact that the estimates of stars from our models are generally competitive to survey long-run projections we believe is a good outcome. It has been well-established that survey expectations are at the frontier of forecasting (e.g., Wright, 2019). However, the preference is for forecasts (or estimates of stars) obtained using a single multivariate model because the resulting forecasts will be coherent and allow for a credible narrative in a systematic manner.

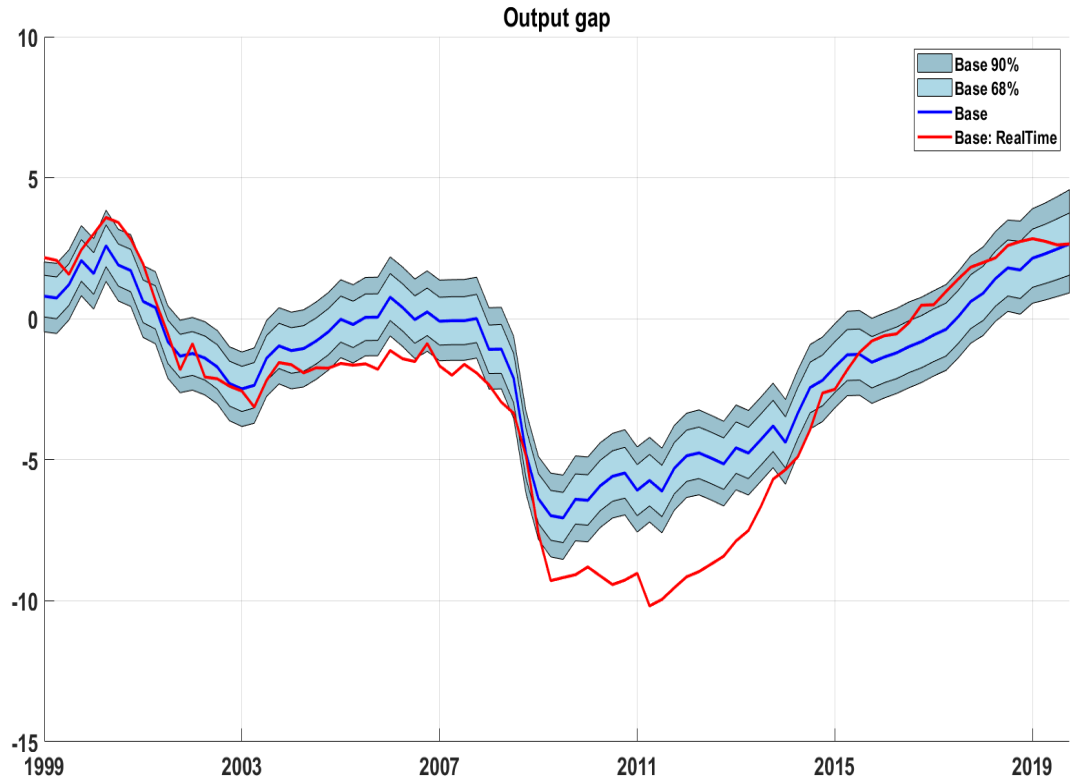
Table C4: Out-of-Sample Forecasting Performance: **Steady-State BVAR**

Full Sample (Recursive evaluation: 1999.Q1-2019.Q4)									
Point forecasting					Density forecasting				
	4Q	8Q	12Q	20Q		4Q	8Q	12Q	20Q
Real GDP									
Relative RMSE					Relative CRPS				
Base/Survey	1.05*	1.07*	1.06*	1.01	Base - Survey	0.09*	0.09*	0.07*	0.01
BaseNoSurv/Survey	1.04	1.09*	1.07*	1.03*	BaseNoSurv - Survey	0.06	0.12*	0.09*	0.03
BaseNoSurv/Base	0.98	1.02	1.01	1.02*	BaseNoSurv - Base	-0.02	0.03	0.01	0.02
PCE Inflation									
Relative RMSE					Relative CRPS				
Base/Survey	0.99*	0.98	1.00	1.05	Base - Survey	-0.02*	-0.02*	-0.01	0.04
BaseNoSurv/Survey	0.97	1.00	1.04	1.06	BaseNoSurv - Survey	-0.03*	-0.01	0.02	0.04
BaseNoSurv/Base	0.99	1.02	1.04	1.01	BaseNoSurv - Base	-0.02	0.01	0.03	0.00
Productivity									
Relative RMSE					Relative CRPS				
Base/Survey	1.04*	1.08*	1.05*	1.00	Base - Survey	0.04*	0.08*	0.06*	0.00
BaseNoSurv/Survey	1.06*	1.13*	1.12*	1.05	BaseNoSurv - Survey	0.07*	0.13*	0.12*	0.05
BaseNoSurv/Base	1.02	1.05*	1.06*	1.05*	BaseNoSurv - Base	0.02	0.05*	0.06*	0.05*
Nominal Wage									
Relative RMSE					Relative CRPS				
Base/Survey	0.73*	0.77*	0.84*	0.92*	Base - Survey	-0.08*	-0.09*	-0.09*	-0.08*
BaseNoSurv/Survey	0.72*	0.76*	0.93*	1.06	BaseNoSurv - Survey	-0.08*	-0.09*	-0.05*	0.03
BaseNoSurv/Base	0.98	0.99	1.10	1.16	BaseNoSurv - Base	0.00	0.00	0.04	0.11
Unemployment Rate									
Relative MSE					Relative CRPS				
Base/Survey	1.05	1.08*	1.09	1.11	Base - Survey	0.03	0.09*	0.13*	0.18*
BaseNoSurv/Survey	1.07	1.13	1.19	1.27*	BaseNoSurv - Survey	-0.08	-0.15	-0.10	0.20*
BaseNoSurv/Base	1.02	1.05	1.09	1.14*	BaseNoSurv - Base	0.02	0.10	0.19	0.31
Shadow FFR									
Relative RMSE					Relative CRPS				
Base/Survey	0.98	0.99	1.01	1.06	Base - Survey	-0.02	-0.02	0.02	0.18
BaseNoSurv/Survey	0.91*	0.92	0.96	1.07	BaseNoSurv - Survey	-0.08*	-0.15	-0.10	0.20
BaseNoSurv/Base	0.93*	0.93*	0.95	1.01	BaseNoSurv - Base	-0.06*	-0.13*	-0.12	0.02

Notes for Table: For the variables real GDP, PCE inflation, productivity, nominal wage (i.e., average hourly earnings), the forecasts and the associated accuracy correspond to the quarterly annualized rate. Base forecast is defined as the Steady-State (SS) VAR forecast in which the steady states are assumed to be the estimates of the stars from the Base model. BaseNoSurv forecast is defined as the SS-VAR forecast in which the steady states are taken from the Base-NoSurv model. The left panel reports results for the point forecast accuracy (relative root mean squared errors) and the right panel reports the corresponding density forecast accuracy (mean of the relative continuous ranked probability score). The table reports statistical significance based on the Diebold-Mariano and West test with the lag $h - 1$ truncation parameter of the HAC variance estimator and adjusts the test statistic for the finite sample correction proposed by Harvey, Leybourne, and Newbold (1997); *up to 10% significance level. The test statistics use two-sided standard normal critical values for horizons less than equal to 8 quarters, and two-sided t-statistics for horizons greater than 8 quarters.

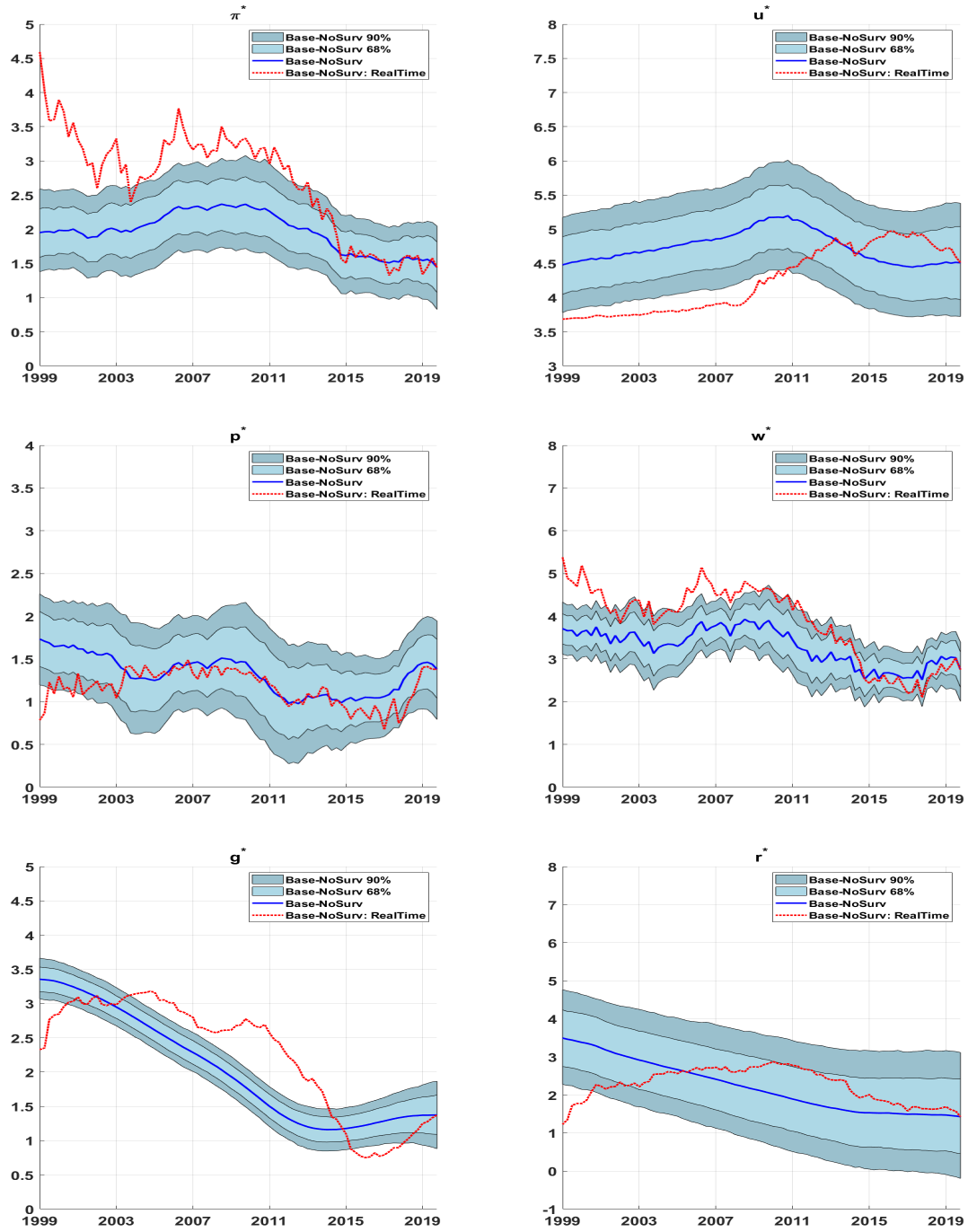
C6. Additional Real-time Estimates Stars

Figure C3: Real-time Recursive Estimates of Output Gap: Base model



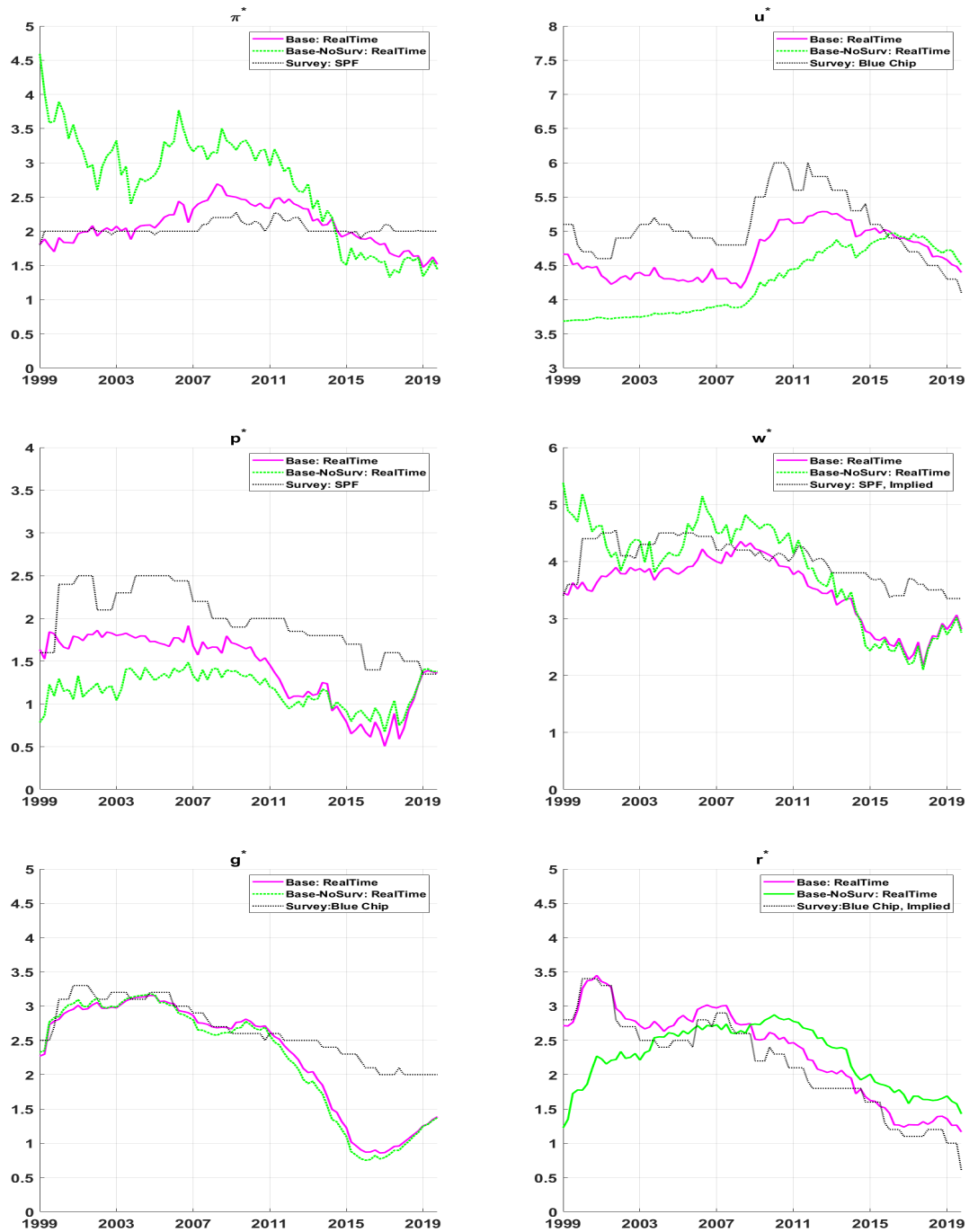
Notes: The plot denoted Base corresponds to smoothed (posterior mean) estimates which are based on the full sample information, i.e., 1959.Q4 through 2019.Q4. The plot denoted Base: RealTime corresponds to real-time recursive (posterior mean) estimate generated by estimating Base model at different points in time, specifically 1999.Q1 through 2019.Q4. The credible intervals reflect the uncertainty around the posterior mean smoothed estimates.

Figure C4: Real-time Recursive Estimates of Stars: Base-NoSurv model



Notes: The plots denoted Base-NoSurv correspond to smoothed estimates which are based on the full sample information, i.e., 1959.Q4 through 2019.Q4. The plots denoted Base-NoSurv: RealTime correspond to real-time recursive estimates generated by estimating Base-NoSurv model at different points in time, specifically 1999.Q1 through 2019.Q4.

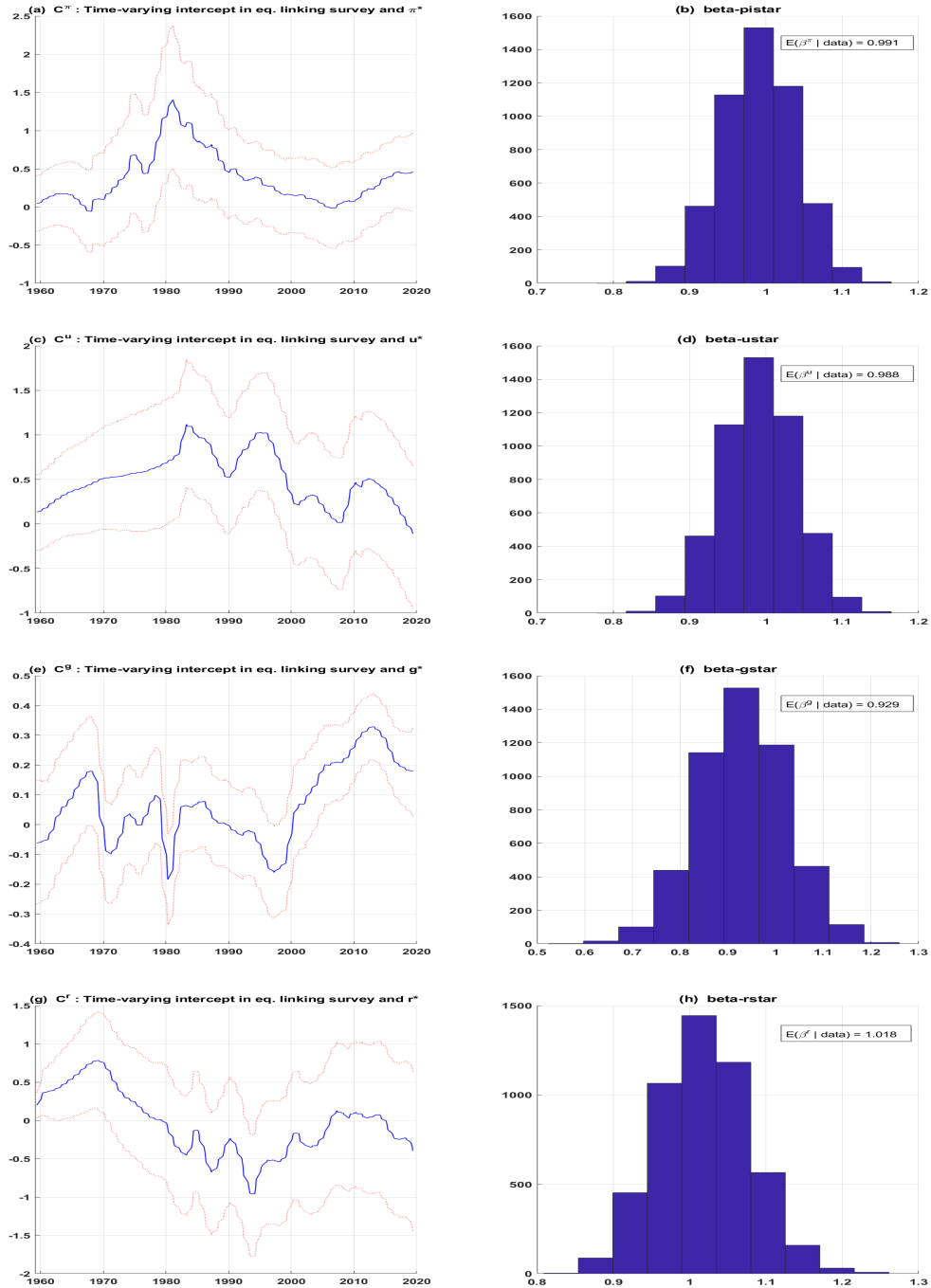
Figure C5: Real-time Recursive Estimates of Stars: Base model vs. Base-NoSurv model



Notes: The plots correspond to real-time recursive estimates generated by estimating Base and Base-NoSurv models at different points in time, specifically 1999.Q1 through 2019.Q4. To facilitate comparison, real-time estimates from the survey either Blue Chip or Survey of Professional Forecasters (SPF) are also plotted.

C7. Estimated relationship between Survey and Stars

Figure C6: Estimated Link Between Survey Forecasts and Stars



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

C8. Additional COVID-19 Pandemic Results

Figure C7 presents posterior estimates of u -star, g -star, and r -star from Base and Base-NoSurv models based on estimating data through 2020Q3. Also plotted to facilitate comparison are the corresponding posterior estimates based on estimation through 2019Q4. Figure C8, similarly, provides estimates of π -star, p -star, and w -star. A visual inspection of the plots suggests the following four observations. First, estimates appear reasonable, indicating the model isn't blowing up. Second, adding pandemic data to the estimation sample has small effects on the historical estimates of stars in the Base model and, for the most part, also applies to the Base-NoSurv model. For u -star, there are some notable revisions in the estimates obtained from the Base-NoSurv model comparing between estimation pre-and post-pandemic recession. The considerable revision in the posterior mean of u -star is associated with decreased precision, as evidenced by the larger width of the 90% credible intervals; however, in the Base model, the estimation with pandemic data is associated with increased precision of u -star.

Third, in the case of g -star, estimation using pandemic data yields posterior mean estimates of g -star that are revised four-tenths higher starting 2000 onwards compared to estimation using pre-pandemic data. Fourth, as would be expected (see Carriero et al., 2021), the precision plots indicate an uptick in uncertainty towards the end of the sample period associated with pandemic data. Though except for p -star and w -star, the uptick in uncertainty is small. The Base model generally held up better because the survey forecasts help anchor the econometric estimates of stars to a reasonable range. Without it, extreme data movements in the unemployment rate profoundly influenced the econometric estimates of u -star in the Base-NoSurv model. In light of the discussion in the preceding paragraph, we view the uptick in uncertainty around p -star as a reasonable result.

We believe that the rich features of our models, which includes: (1) modeling the changing economic relationships via the implementation of time-varying parameters; (2) allowing for changing variance of the innovations to various equations (i.e., SV); (3) imposing bounds on some of the random walk processes; (4) joint modeling of the output gap and unemployment gap in particular; (5) and the use of survey forecasts; helped position our models to handle the pandemic data better.

Carriero et al. (2021) using monthly Bayesian VARs show models that allow for SV better handle pandemic observations than those without SV. But, even models with SV have a drawback in the context of pandemic data. This drawback arises from the standard approach to modeling SV, which assumes a random walk process or a very persistent AR process. So in the face of a temporary spike in volatility, the model will attribute this spike incorrectly to a persistent increase in volatility. Inspired by the outlier treatment method of Stock and Watson (2016) for UCSV models, Carriero et al. (2021) propose an outlier-adjusted SV method that models the VAR residuals as a combination of persistent and transitory changes in volatility.

We believe that this drawback of standard SV applies more to monthly VARs and to a lesser extent in quarterly models, as is the case here. However, we stress that Stock and Wat-

son outlier treatment method can be conveniently implemented in our modeling framework. It would also require introducing SV in both the output gap and the unemployment gap equations. To keep the length of the chapter manageable, we leave this extension for future research.

The COVID-19 pandemic provides an excellent real-time illustration of the importance of using survey expectations data in the econometric estimation of the stars. The unprecedented nature of the pandemic crisis and the extreme movements in the data induced by the pandemic are too volatile to provide a timely and credible signal about the long-run macroeconomic consequences. Complicating the signal extraction problem from the data during the pandemic period is that consensus has been developing (perhaps rightly so) to treat macroeconomic data for the periods 2020Q2 and 2020Q3 as outliers in estimating the macro-econometric models; see Schorfhedie and Song (2020), Lenza and Primiceri (2020), Carriero et al. (2021), among others.

On the other hand, judgment assessment informed from past event studies and understanding of many decades of economics research indicates that the COVID pandemic is likely going to have implications on the long-run productivity growth (p-star), the growth rate of potential output (g-star), the natural rate of unemployment (u-star), the long-run real rate of interest (r-star); see Jorda, Singh, and Taylor (2020). As time rolls forward, and more is revealed about the possible long-term macroeconomic impacts of the pandemic on the underlying trends, the survey participants would judgmentally adjust their estimates of long-run projections in a more timely manner. And by extension, our Base model, which incorporates the long-run survey projections.

Base model vs. external sources: post-pandemic Recession

We next compare our Base model estimates with those produced by external sources (and or models) to assess further the reliability of our Base model estimates post-pandemic recession. Figure C9 compares the estimates of the output gap (panel a), r-star (panel b), u-star (panel c), and pi-star (panel d) from the Base model to the outside estimates.³ The estimates are based on data through 2020Q3 (specifically vintage of data corresponding to late November 2020). In the case of CBO, the projections correspond to an update as of late July 2020.

The plots in panel (a) indicate remarkable similarity between the posterior mean estimate of the base model's output gap and the CBO output gap. Compared with Morley and Wong (2020), even though before the pandemic, the base model's output gap estimates indicated less

³Morley and Wong (2020) estimates are based on their updated work Berger, Morley, and Wong (forthcoming) and are available to download from outputgapnow.com. The estimates were downloaded in the last week of November, which included the nowcast estimate for 2020Q4 that we do not plot. Thank Murat Tasci, for providing the estimates of the u-star from the Tasci (2012) model. And also, thank Benjamin Johannsen for providing the r-star estimates from Johannsen and Mertens (2019). The LW estimates of r-star were downloaded from the New York Fed website. Del Negro et al. (2017) estimates of r-star were downloaded from github.com/FRBNY-DSGE/rstarBrookings2017. Lubik and Matthes estimates were downloaded from the Richmond Fed website in late November 2020.

tight resource utilization, for 2020, they are quite similar. Morley and Wong (2020), based on a BVAR approach, could be viewed more flexibly than ours because it explicitly considers the possible error correlation across model equations. However, at the same time, their approach could be deemed less flexible than ours because it does not explicitly model time-variation in parameters and stochastic volatility – i.e., abstracts from the issue of “changing economic environment.” Both the Base model and Morley and Wong (2020) estimates the output gap at -3.5% for 2020Q3, with CBO just a tenth higher at -3.6%.

Panel (b) plots the estimates of the r-star from various sources. Except for Laubach and Williams (2003) [LW], all others are based on information available as of late November 2020. LW estimate reflects information through August 2020. Comparing between 2019Q4 and 2020Q3, the Base model, Johanssen and Mertens (2019), and Del Negro et al. (2017), all three estimate r-star to have changed only a little; Base model: from 1.36% to 1.26%, Del Negro et al. (2017): from 1.11% to 1.08%, Johanssen and Mertens (2019): from 1.48% to 1.47%. In contrast, Lubik and Matthes (2015) have r-star increasing from 0.64% to 1.0%. However, in their estimate, r-star first falls from 0.64% to -0.68% and then bounces back to 1.0% in 2020Q3. Their estimate of r-star displays considerable volatility compared to others.

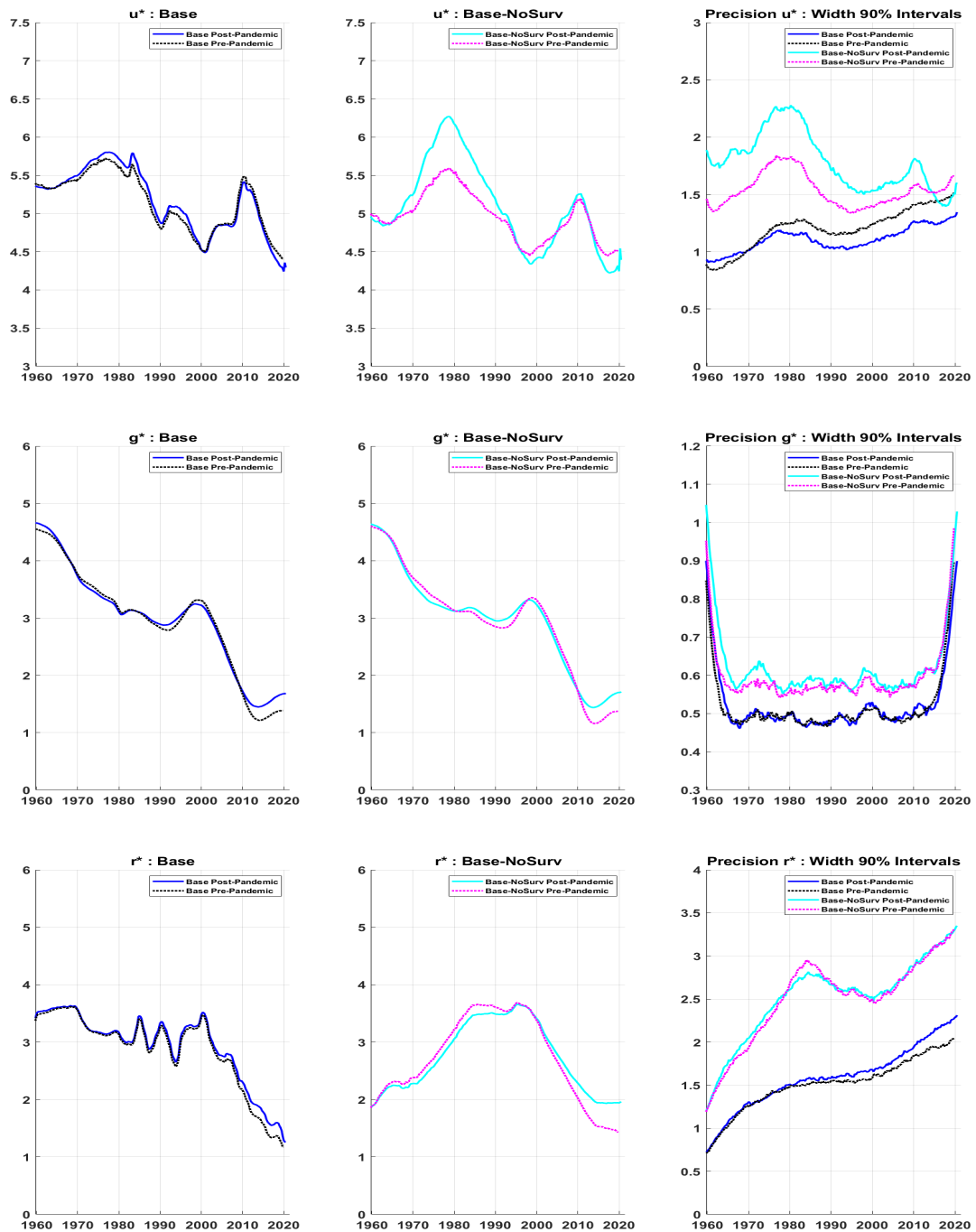
Panel (c) plots the estimates of the u-star from four sources, Base model, CBO, Tasci (2012), and Chan, Koop, and Potter (2016). Comparing between the Base model and CBO, the contours of the u-star plots are quite similar. But, the levels through 2010 are notably different, with CBO higher than the Base model. From mid-2013 onwards, the levels are quite similar, and in 2020Q3, both indicate u-star at 4.3% (Base) and 4.4% (CBO). Interestingly, both CBO and the Base model have u-star remaining mostly stable between 2019Q4 and 2020Q3, suggesting that they attribute most of the increase in the pandemic’s unemployment rate to the cyclical component. It is worth highlighting that the (median) estimate of u-star reported in the September 2020 Summary of Economic Projections, which the Federal Reserve compiles, also indicated a stable u-star (at 4.1%) between 2019Q4 and 2020Q3.

Broadly speaking, the contour of the u-star implied by the CKP (bivariate Phillips curve) is similar to the Base model and CBO. But, the estimated level of u-star is significantly higher. According to the CKP model, the estimated u-star in 2020Q3 is 5.7%, just a tenth higher than 2019Q4. The Tasci (2012) model, which is based on the flow rates in-and-out of unemployment, is significantly impacted by the pandemic data, as the u-star is estimated to have increased from 4.7% in late 2019 to 5.2% in 2020Q3. Part of the explanation of more significant movements in u-star seen in the Tasci model in response to pandemic data is that the model is estimated using maximum likelihood methods, which are relatively known to have done a less well job in handling extreme pandemic induced movements in variables. More generally, Tasci (2019) document the challenges of estimating u-star in real-time with these models during crisis periods.

Panel (d) presents pi-star estimates from three sources: the Base model, CCK model, and CKP model. All three models indicate pi-star to have remained stable between 2019Q4 and 2020Q3. However, the pi-star estimates differ slightly across models, with the Base model at

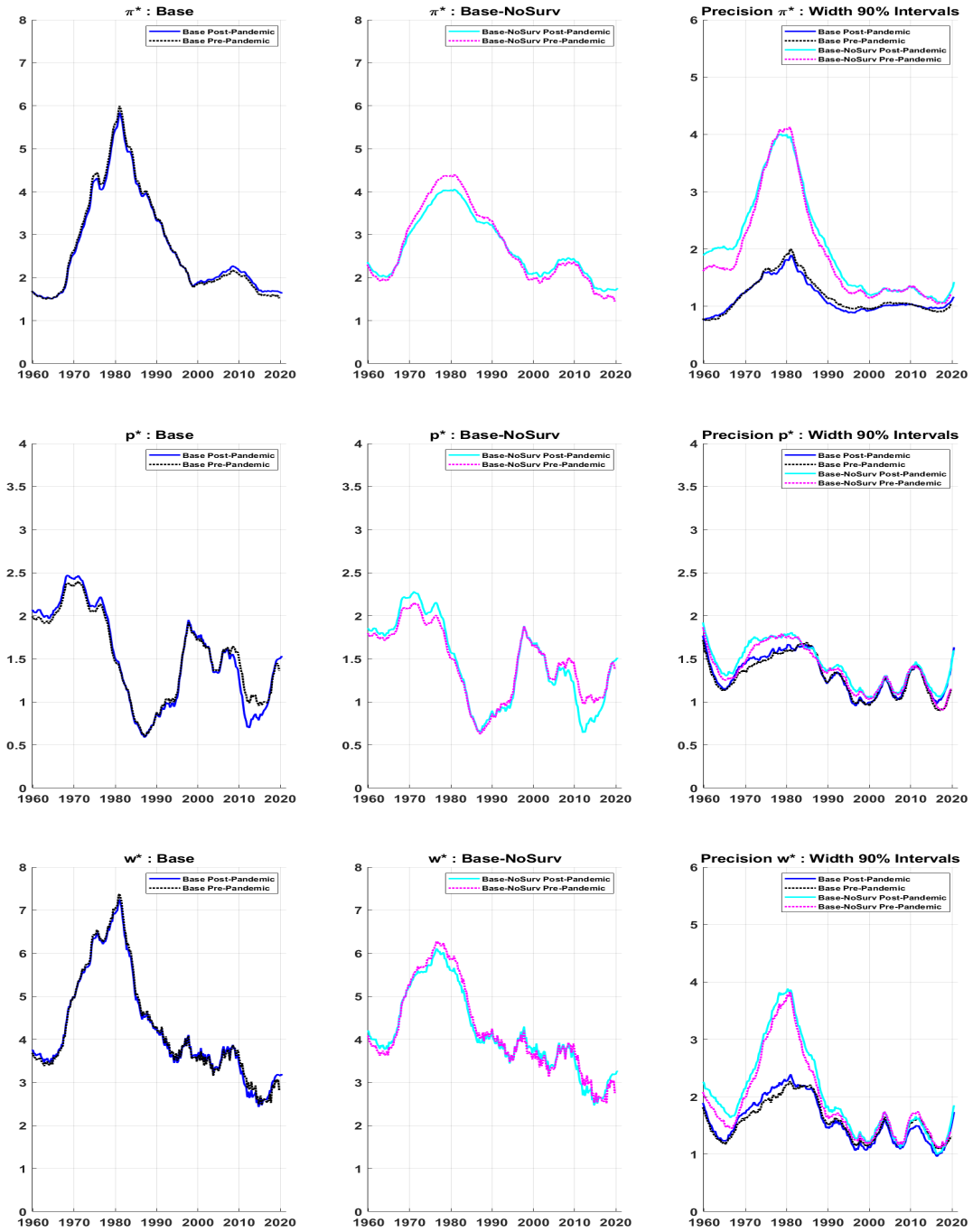
1.65%, CCK at 1.50%, and CKP at 1.44% (in 2020Q3).

Figure C7: Estimates of Stars pre- vs. post-COVID Recession



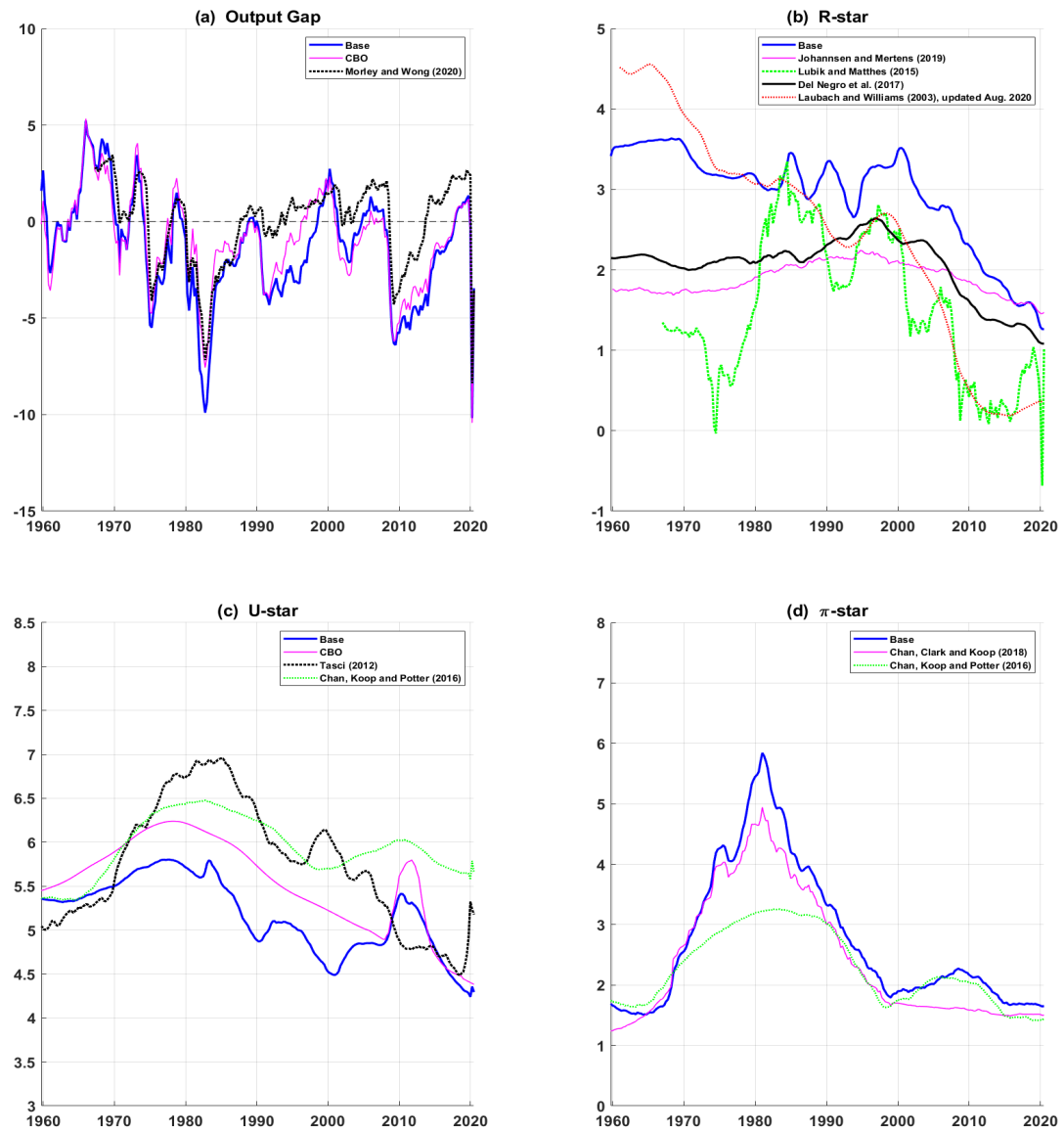
Notes: The plots labeled Pre Pandemic reflect posterior estimates based on information in the sample 1959Q4 through 2019Q4, and plots labeled Post Pandemic reflect posterior estimates based on the sample 1959Q4 through 2020Q3.

Figure C8: Estimates of Stars pre- vs. post-COVID Recession (more)



Notes: The plots labeled Pre Pandemic reflect posterior estimates based on information in the sample 1959Q4 through 2019Q4, and plots labeled Post Pandemic reflect posterior estimates based on the sample 1959Q4 through 2020Q3.

Figure C9: Estimates of Stars post-COVID Recession: Base vs. Outside



Notes: In the case of Johanssen and Mertens (2019), Del Negro et al. (2017), and Lubik and Matthes (2015), the estimates plotted are the posterior median, for all others it is the (posterior) mean estimate.

C9. R*: Backcast Survey R* from 1959-1982

The survey estimates of g-star, u-star, and pi-star are direct reads from the survey. In contrast, the r-star survey estimate is not a direct estimate. Instead, it is inferred from the Blue Chip survey long-run estimates of GDP deflator and short-term interest rates (3-month Treasury bill) using the long-run Fisher equation. Specifically, the long-run forecast of 3-month Treasury bill less the long-run forecast of GDP deflator. To this differential, we add +0.3 to reflect the average differential between the federal funds rate and the 3-month Treasury bill. (r-star refers to the long-run equilibrium federal funds rate)

Survey projections are not available before 1983Q1. To fill in estimates for the survey variables between 1959Q4 and 1982Q4, we use CBO long-run projections in the case of real GDP growth and the unemployment rate. In the case of inflation, we use the PTR series available from the Federal Reserve Board website; this series is used in many studies employing long-run expectations of inflation (e.g., CCK, Tallman and Zaman, 2020). We do not have a readily available historical source for long-run forecasts for interest rates (and r-star). So we backcast a particular time series of implied r-star using CBO's long-run projections of g-star. Specifically, we first fit a simple linear regression model over the post-1983 period that regresses survey r-star on a constant, its lags (2 lags), and a one-period lag "gap," defined as the difference between survey r-star and survey g-star. We use the estimated model and the CBO long-run projections of g-star over the sample 1959Q4 through 1982Q4 to backcast the implied survey r-star estimates. (When backcasting, the initial values of r-star for 1959Q2 and 1959Q3 are assumed identical to CBO g-star)

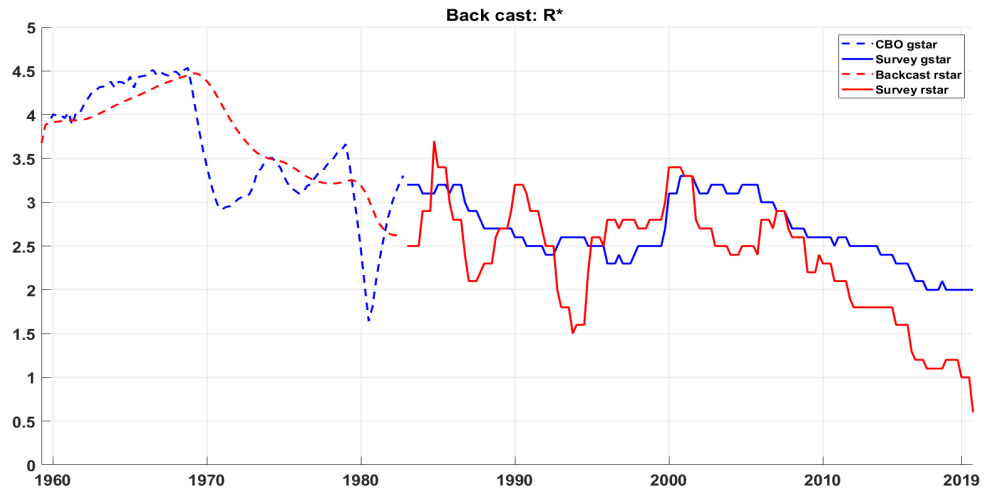
$$r_t^{*,Surv} = c + \beta_1 gap_{t-1}^{r^*,g^*,Surv} + \beta_2 r_{t-1}^{*,Surv} + \beta_3 r_{t-2}^{*,Surv} + \varepsilon_t^{*,Surv}, \quad \varepsilon_t^{*,Surv} \sim N(0, \sigma_{*,Surv}^2) \quad (C.85)$$

$$\text{where, } gap_t^{r^*,g^*,Surv} = g_t^{*,Surv} - r_t^{*,Surv}$$

The OLS estimation yields, $c = -0.0745$; $\beta_1 = 0.06$; $\beta_2 = 1.167$; $\beta_3 = -0.148$

Figure C10 plots the survey g-star and r-star estimates in solid, and the CBO g-star and the backcast r-star in dashed lines.

Figure C10: Survey r^* and g^*



C10. R* : Additional Full Sample Results

C10.a. Role of data vs. prior in shaping r-star

Kiley (2020), using a model in which r-star follows a RW process, documents an essential finding that data provide very little information in shaping the r-star process. Hence, the model-based r-star estimate is mainly driven by the modeler's prior views. Our results generally confirm Kiley's findings. However, in our Base specification, where the variance of the g-star process influences both the prior and the posterior for the r-star process, the data does influence the r-star estimate; because we find evidence that data provide information about the g-star process. This latter evidence of data's influence on the identification of g-star is also noted by Kiley (2020).

We begin by comparing the prior and posterior estimates of the parameter $\sigma_{r^*}^2$, which governs the shock variance of the r-star process, in the Base-R*RW and Base-NoSurvR*RW – both these specs model r-star as a RW similar to Kiley (2020). We set prior for $E(\sigma_{r^*}^2) = 0.1^2$, which is the same as in Astudillo and Laforte (2020) but tighter than 0.25^2 used by Kiley.⁴ (Our choice of tighter prior than Kiley is due to a significantly more complex model). Our model estimation yields posterior estimates of 0.09^2 (with 90% credible intervals 0.07^2 to 0.11^2) in Base-R*RW and 0.1^2 (with 90% interval 0.08^2 to 0.13^2) in Base-NoSurvR*RW, respectively. It appears that in the case of Base-NoSurv-R*RW, the prior setting of the r-star process is driving the trajectory as evidenced by the posterior mean of the parameter $\sigma_{r^*}^2$ identical to the prior. But in the case of Base, the posterior mean of the parameter $\sigma_{r^*}^2$ is slightly different from the prior mean, suggesting that by bringing survey data in the estimation, the data does play a role in shaping r-star.

We next confirm our finding by re-doing our exercise setting a looser prior for $E(\sigma_{r^*}^2) = 0.25^2$, same as in Kiley (2020). The updated model estimation yields posterior estimates of $E(\sigma_{r^*}^2) = 0.15^2$ (with 90% credible intervals 0.13^2 to 0.17^2) in Base-R*RW and $E(\sigma_{r^*}^2) = 0.22^2$ (with 90% interval 0.18^2 to 0.27^2 in Base-NoSurvR*RW, respectively. The fit of these models to interest rate data (and other model data) is significantly worse compared to Base and Base-NoSurv.

We explored the impact on the r-star estimates of even more looser priors on the shock process governing r-star. We find that as the prior on the r-star process loosens, data becomes more informative in shaping the r-star estimate (echoing Lewis and Vazques-Grande, 2019). But it comes at the cost of worsening model fit, higher volatility in the r-star estimate, and worsening precision of r-star.

⁴We also explore a model specification in which prior variance is set at 0.25^2 , the fit of this specification was significantly inferior, and the r-star estimate was quite volatile.

C10.b. Base vs. External models

In figure C11, the left panel plots r -star from Base (solid line) and two external models, the seminal model of LW (dashed line) and the more recent model developed in Del Negro et al. (2017) (dotted line). As is the case with most r -star estimates presented in the literature, the LW estimate shows a marked decline in r -star from 2000 and beyond. As shown in the figure, compared to the r -star estimate from the Base, the LW estimate is lower over this period. Part of the explanation of this difference in the estimates comes from the different estimates of g -star (not shown).

In the LW model, the mechanical reason for this steadily declining trajectory of r -star is coming from the fact that their model estimate of g -star has been steadily declining over the same period. Over this period, GDP grew just slightly above their estimate of g -star, even though the real short-term rate is significantly below zero over this period. The model explains the combination of moderate growth in GDP (suggesting a small positive output gap) and negative real short-term interest through a low level of r -star estimate so to obtain a negative real interest rate gap (see LW, 2016). In our Base (and Base-NoSurv) model, because the estimate of g -star is even lower than LW, which implies a more positive output gap (than LW), a less negative real interest rate gap (than LW) is required to explain the output gap. The less negative real interest rate gap (i.e., a smaller interest rate gap) implies a higher level of r -star than LW.

The r -star estimate from Del Negro et al. is stable around 2% from 1960 through early 1980 and then slowly move up, reaching 2.5% by late 1990. From thereon, it begins a gradual decline ending 2019 at 1.2%, identical to Base, and two-tenths lower than Base-NoSurv. It is worth noting that Del Negro et al. also utilize survey expectations on r -star to estimate r -star but their approach in how they model the link between the two is very different than ours.⁵ They also assume a relationship between g -star (in their case, long-run productivity growth) and r -star. However, their model structure is different compared to ours. Shortly, we show an r -star estimate from our model specification with the tighter prior assumption for the r -star process, which is remarkably similar to Del Negro et al.

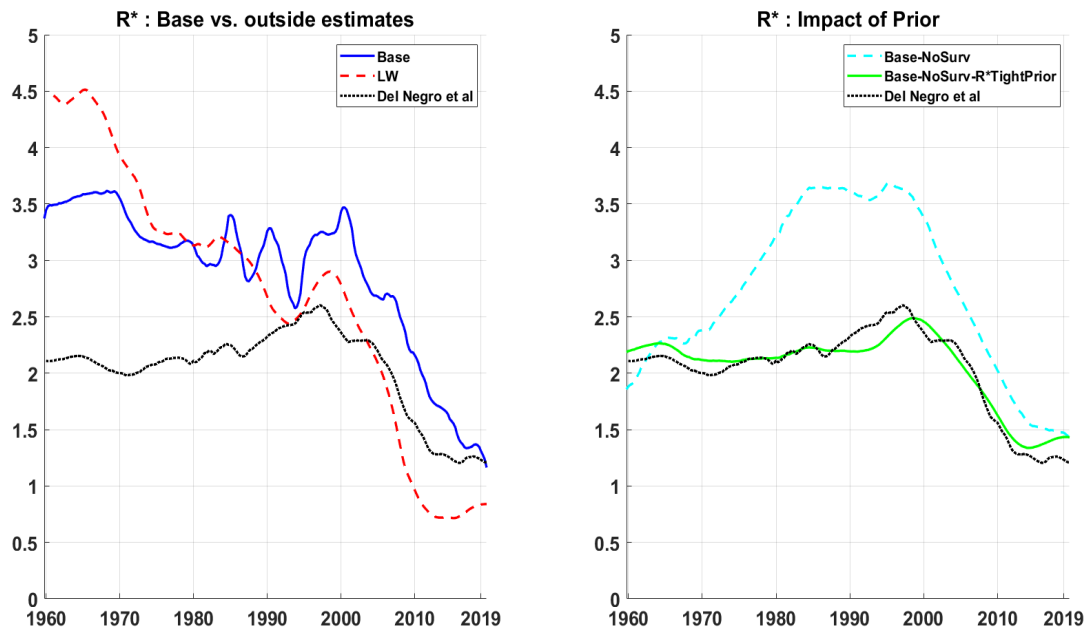
C10.c. Sensitivity of r -star to the prior setting

As just shown and noted by others (e.g., Kiley (2020)), the prior elicitation for the variance parameter of the shock process governing r -star has a notable influence on the dynamics of

⁵Del Negro et al. use survey expectations from Survey of Professional Forecasters, which start from 1992 onwards. In addition, in their framework survey expectations is one of the several financial indicators that they use to extract a common trend. So arguably, in their approach, the survey expectations of r -star will be less influential in driving r -star than in our approach in which a direct connection between r -star and survey data is assumed.

r-star. We briefly show another illustration highlighting the sensitivity of r-star to the prior setting. In figure C11, the right panel plots the posterior mean r-star obtained from model specification, Base-NoSurv-R*TightPrior, which is Base-NoSurv but with a tighter prior value for the parameter σ_d^2 (0.01^2 instead of 0.1^2). The parameter σ_d^2 refers to the variance of the shock process defining the “catch-all” component D. Also plotted are the posterior estimates of r-star from Base-NoSurv and Del Negro et al. (2017) model. Three things immediately stand out. First, imposing a tighter prior has a notable impact on r-star, as shown by comparing dashed and solid lines in the figure. Second, the model specification Base-NoSurv-R*TightPrior has the posterior mean of r-star near 2% from early 1960 through mid-1980, which is similar to the r-star estimate reported in Kiley (2020). Third, the entire trajectory of r-star from the Base-NoSurv-R*TightPrior is remarkably similar to the median estimate of r-star from the Del Negro et al. model. These results indicate that very different approaches could provide similar estimates, yet somewhat related approaches could yield very different estimates.

Figure C11: R* estimates

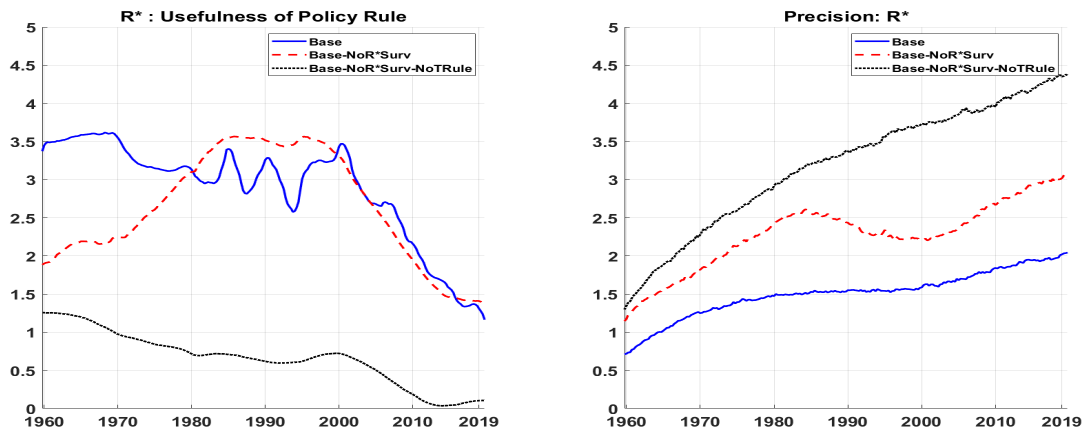


C10.d. The usefulness of the Taylor-rule equation and the equation linking r^* to survey

In recent studies on estimating r^* , a Taylor-type rule equation is added to the model structure to improve the econometric estimation. Our Base model also includes a Taylor-type rule. As we now illustrate, this addition is crucial to improve precision and the plausibility of the r^* estimates significantly. The left panel in figure C12 plots three estimates of r^* obtained from model specifications Base (solid line), Base-No R^* Surv (dashed line), and Base-No R^* Surv-NoTRule (dotted line). The right panel plots the corresponding precision of the r^* estimates. The specification Base-No R^* Surv excludes the equation linking r^* to survey expectations from the model (but keeps equations relating other stars to survey). Doing so produces a trajectory of r^* similar to the Base-NoSurv spec; and not surprisingly, the precision of r^* is reduced relative to the Base spec, as evidenced by the plot corresponding to Base-No R^* Surv lying above the Base.

The specification Base-No R^* Surv-NoTRule excludes the equation linking r^* to survey expectations and the Taylor-rule equation. So in this spec, r^* is identified from the IS-curve equation, and the equation relating r^* to g^* . As expected, shrinking the model's structure further by excluding the Taylor-rule equation reduces the r^* estimate's precision dramatically, as evidenced by Base-No R^* Surv-NoTRule plot located above all the others in the left panel. Besides the impact on the precision, as would be expected, changes in the system's structure result in notable differences in the estimated level of the r^* . The posterior mean estimate of r^* , which has the r^* declining steadily over the sample, is substantially lower than both Base and Base-No R^* Surv. However, the uncertainty around the posterior mean is enormous complicating inference with a reasonable degree of certainty.

Figure C12: The Usefulness of Taylor Rule equation



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

C11. π^* : Additional Full Sample Results

C11.a. Pi-star comparison Base vs. outside models

In figure C13, panel (a) plots posterior mean estimates of pi-star from some related (smaller size) models from the literature alongside Base to facilitate comparison. In particular, estimates are shown for CKP, CCK, and the celebrated UCSV model of Stock and Watson (2007).⁶ Panel (b) plots the corresponding precision estimates of pi-star.

There are some interesting similarities and differences across the pi-star estimates. Whereas UCSV displays very volatile and erratic estimates of pi-star, others show smoother evolution of pi-star. CKP indicates a lower estimate of pi-star than others from the early 1970s through the late 1980s. The primary factor contributing to lower pi-star in CKP is the model assumption of a bounded random walk for pi-star. As discussed in CKP, the addition of bounds on pi-star leads the model to attribute a substantial share of the high observed inflation of the 1970s to the increased persistence of the inflation gap and only a small increase in the pi-star. Hence, pi-star is estimated to have risen less than implied by other models. For instance, CCK model had pi-star peaking at 4.9%, Base at 6.0% and CKP at 3.2%. As alluded in CKP, this small rise in pi-star is consistent with a specific narrative that during the Great Inflation period, the Fed had a low implicit target for inflation but was either unable to or unwilling to correct large deviations of inflation from the target.

The contours of pi-star from Base is similar to CCK through 2000, but from 2000 to 2012, Base is identical to CKP, with CCK a touch lower. It is interesting to note that from the early 2000s through 2010, both Base and CKP indicate pi-star at 2%. From 2012 through 2019, both Base and CKP gradually drift lower to 1.5% (same as CCK) and 1.3%, respectively.

Panel (b), which plots the corresponding precision of pi-star, reveals some interesting patterns. First, the precision of the pi-star evolved generally with the level of pi-star. As pi-star increased during the Great Inflation, pi-star became more uncertain, i.e., more imprecise. Subsequently, as pi-star trended lower during the Volcker disinflation, so did the uncertainty about it (i.e., precision increased). Second, comparing across models, there is significant heterogeneity in the precision of pi-star. From 1960 through the mid-1970s, the Base model indicates the most precise pi-star, followed by CCK and CKP. UCSV model shows volatile estimates of precision, sharply fluctuating between the most precise to least precise. From the mid-1970s through 2019, the CCK model indicates the most precise (least uncertain) pi-star, followed by Base, CKP, and UCSV. CCK had the uncertainty of pi-star gradually trending down starting in the mid-1970s. In contrast, in others, the uncertainty continued to trend higher until peaking in the early 1980s.

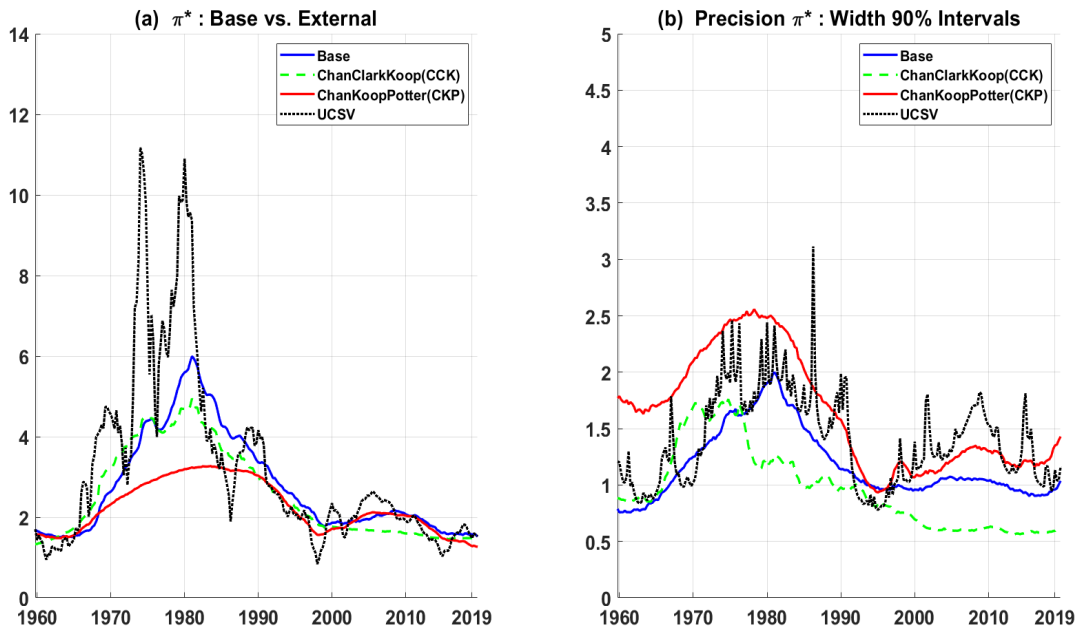
Third, between 2000 and 2019, the uncertainty around pi-star implied by CCK and Base has been reasonably stable, an artifact of the use of survey data. During this period, the precision of

⁶Whereas in estimating the UCSV model, Stock and Watson (2007) fix the parameters governing the smoothness of the SV processes, we estimate them.

pi-star implied by CCK is on average 40 basis points higher (i.e., uncertainty is lower) compared to Base. This improved precision of CCK is interesting because both CCK and Base utilize information from survey expectations of inflation. However, at the same time, compared to Base, which has a rich structure (hence more parameters), the CCK model is parsimonious, as it uses information from inflation and survey only to estimate pi-star.

An additional factor that could contribute to the differential in precision is that, unlike Base, CCK allows SV in the pi-star equation. A more in-depth inspection of the estimation results reveals the primary factor driving the superior precision of the CCK estimate of pi-star compared to Base is tighter priors on the assumed relationship between survey forecast and pi-star. And that translates into a posterior estimate implying a stronger connection between survey forecast and pi-star in CCK than Base.

Figure C13: Pi* estimates: Base vs. External models



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4). In all cases, the inflation measure is the PCE inflation.

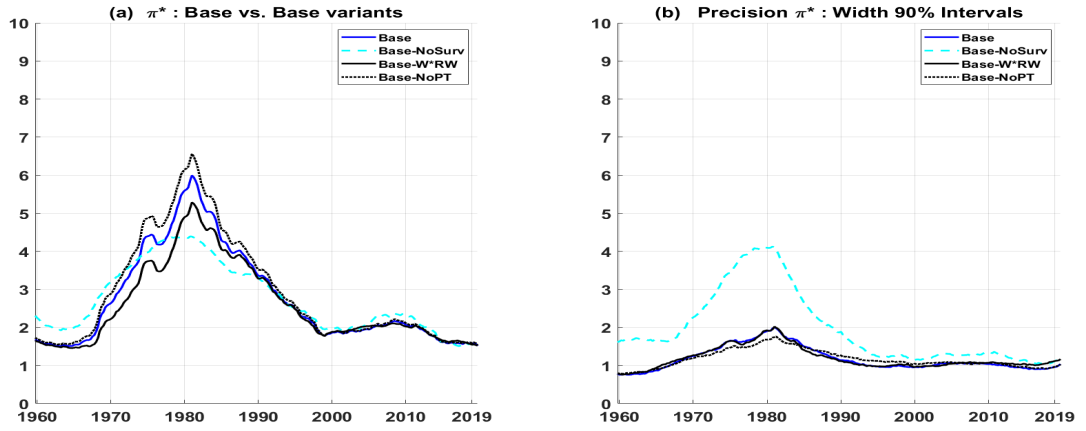
C11.b. Sensitivity of pi-star to modeling assumptions

Figure C14, panel (a) indicates the sensitivity of the pi-star estimates to modeling assumptions. The plot labeled Base-W*RW is the variant of the Base model that removes the theoretical restriction imposed by equation (4.28) and instead assumes a random walk assumption for w-star. Comparing the Base and Base-W*RW plots indicate the effects of the theoretical restriction on pi-star. As shown, the posterior mean estimate of pi-star from Base-W*RW is marginally lower than Base in the period 1970 through the early 1980s (Great Inflation period). However, from thereon, estimates of pi-star are identical. During the high-inflation period, compared to the Base model, the Base-W*RW allocates a higher share of the increase in inflation to the persistence component than pi-star (i.e., the random walk component); see figure C14. Hence, the lower level of pi-star in Base-W*RW than Base.

The plot labeled Base-NoPT is the variant of the Base model that eliminates the passthrough from prices to wages, modeled via equation (29b)—doing so results in a slightly higher pi-star (Base-NoPT) from 1970 through the early 1980s. However, thereafter, estimates of pi-star are identical between Base and Base-NoPT. During the high-inflation period, compared to the Base model, the Base-NoPT allocates a lower share of the increase in inflation to the persistence component than pi-star; hence, the higher level of pi-star in Base-NoPT than Base. Based on the model comparison, Base-W*RW model's fit to the inflation data and other data is inferior compared to Base. In the case of Base-NoPT, the fit to the inflation data is slightly higher than Base. However, the overall fit of the Base-NoPT is significantly worse than Base. The Base-NoPT model's reduced fit is the net effect of its reduced ability to fit wages and its improved ability to fit prices.

We also explored a variant of the Base model that allowed the passthrough from wages to prices in the price inflation equation, denoted Base-PT-Wage-to-Prices in table 1 (see main chapter). The estimates of pi-star (and of other parameters) are identical to those of the Base; hence, they are not shown. Therefore, not surprisingly, as reported in table 8 (see main chapter), both models' ability to fit inflation data are very similar. We also highlight that allowing SV in the inflation equation is very important, as evidenced by a significantly reduced fit of the Base-NoSV model, which is the Base model variant that does not feature SV in any model equations.

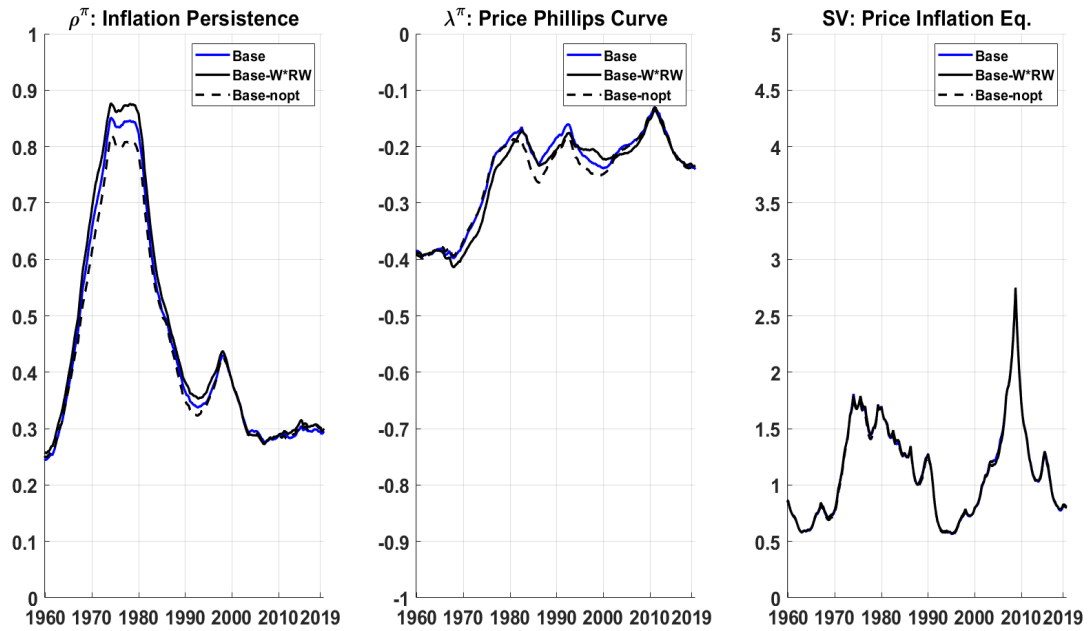
Figure C14: More Estimates for Price inflation block



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

C11.c. Pi-star estimates for some variants of the Base model

Figure C15: Pi* estimates: Base vs. Base model variants



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

C12. P*: Base comparison with Kahn and Rich (2007)

In this section, we compare our model-based estimates of p-star with the narrative about p-star implied from the two-regime Markov switching model of Kahn and Rich (2007).⁷ A regime-switching framework (as in Kahn and Rich) allows for deterministic values of p-star, where the number of deterministic values equals the number of possible regimes. Accordingly, in a 2-regime setup, the estimated p-star would periodically alternate from one-regime (e.g., low productivity regime) to the other regime (e.g., high productivity regime). In contrast, the random walk assumption for p-star adopted in this chapter (and in others such as Roberts, 2001; Edge et al., 2007; Benati, 2007) allows for the possibility that p-star may be (slowly) changing in every period. This latter assumption implies that the possible values of p-star could equal the number of periods in the estimation sample. The differences in the stochastic conception between the two frameworks complicate direct comparison in p-star.

One possible albeit imperfect approach to comparing the implied p-star from two frameworks is to use the regime-switching model's identified regimes to assess how well those corroborate with p-star estimates implied from the RW assumption model. Specifically, for the RW model, compute the "average" p-star over the specific periods (identified regimes). Then assess the following: (1) whether the "average" rates imply characterization of regimes that corroborate with the identified regimes; (2) how close the "average" rates of p-star are to the deterministic values of p-star estimated in the regime-switching model. We use this approach to compare the estimates of p-star from our models to the p-star estimated by the Kahn and Rich model.

Figure C16 presents the comparison of p-star. Panel (a) compares the Base model with Kahn and Rich model, and panel (b) compares the Base-W*RW model with Kahn and Rich model. In the panels, the shaded areas refer to the two regimes identified by the Kahn and Rich model using the same vintage of data as our models. The lighter shaded area corresponds to the "high productivity regime," and the darker shaded area "low productivity regime." Their model identifies two subperiods of high productivity regimes: the beginning of our sample through 1974Q4 and 1996Q3 through 2004Q4. Similarly, two subperiods of low productivity regime: 1975Q1 through 1996Q2 and 2005Q1 through the end of the sample, 2019Q4. Based on the "average" rates of p-star computed for the specific two regimes from our models, if we assume a cutoff of 1.5%, with "average" rate of p-star $\leq 1.5\%$ as defining low productivity regime, and "average" rate $> 1.5\%$ as defining high productivity regime, then the characterization of regimes (and in-turn the narrative) aligns perfectly with Kahn and Rich.

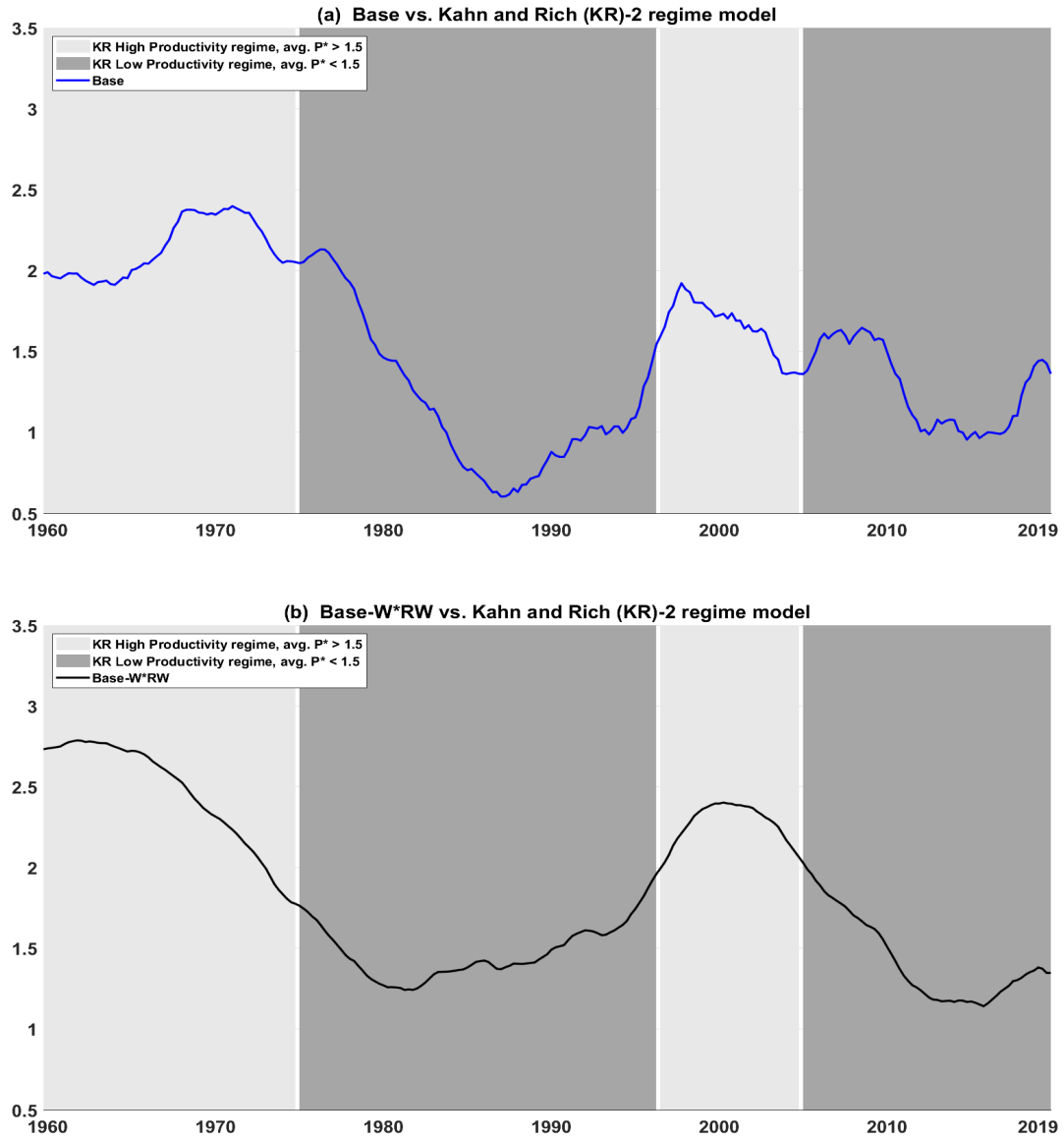
Next, we compare the "average" rates for the two-regimes implied by our models to Kahn and Rich model. The Base model implies for a low productivity regime an "average" rate of 1.3% (for both subperiods) and for a high productivity regime an "average" rate of 2.1% (subperiod beginning of our sample through 1974Q4) and 1.7% (subperiod 1996Q3 through

⁷The estimates of p-star implied by the Kahn and Rich (2007) model are routinely updated and made available for download at James A. Kahn's website: <http://sites.google.com/view/james-a-kahn-economics/home/trend-productivity-update>

2004Q4). The Base-W*RW model implies for a low productivity regime an “average” rate of 1.5% (for both subperiods) and for a high productivity regime an “average” rate of 2.5% (in the first subperiod) and 2.3% (in the second subperiod). In comparison, Kahn and Rich’s model implies a p-star of 1.33% for a low productivity regime for both subperiods – p-star are equal across subperiods by construction; and 2.96% p-star for a high productivity regime. For the low productivity regime, the implied p-star is similar between our models and Kahn and Rich, but for the high productivity regime, Kahn and Rich’s model is on the higher side than our models.

Overall, this illustration suggests that the two approaches provide generally similar inferences about developments in p-star, and we view this as a useful result for macroeconomists tasked with modeling and tracking productivity developments.

Figure C16: P^* consistent with narrative from 2-Regime Markov-Switching Model

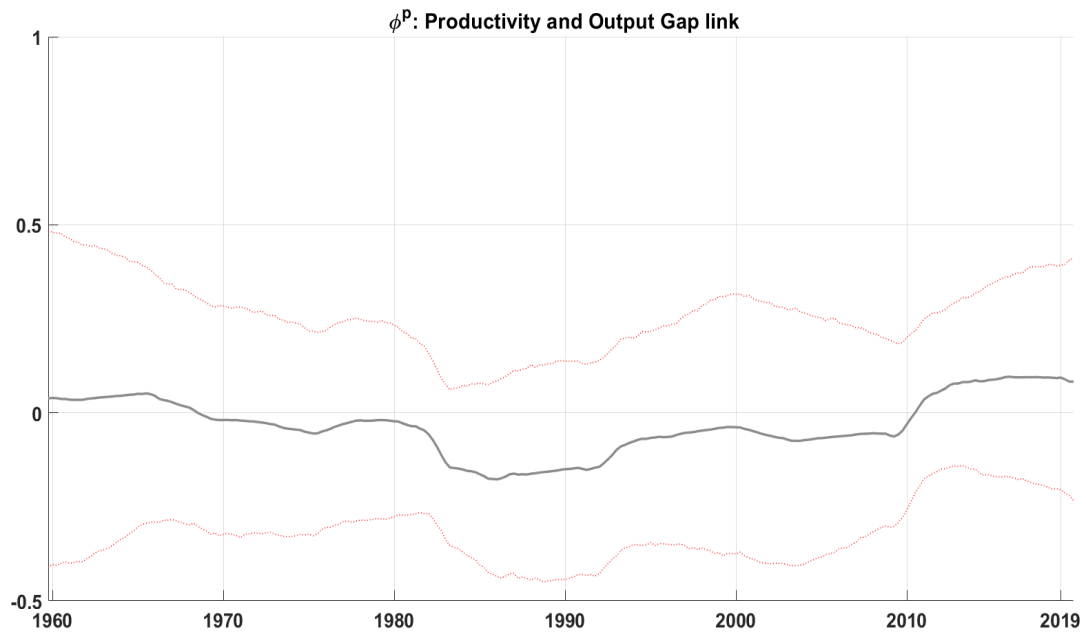


Notes: The shaded areas refer to the two regimes identified by the Kahn and Rich model using the same vintage of data as our models. The lighter shaded area corresponds to the “high productivity regime,” and the darker shaded area “low productivity regime.” The plots labeled Base and Base-W*RW are the posterior mean estimates based on the full sample (from 1959Q4 through 2019Q4).

C13. P* : Additional Full Sample Results

C13.a. Cyclical Productivity based on Output gap

Figure C17: Base-P*CycOutputGap model



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4). The solid line represents the posterior mean and the dotted lines represent the 90% credible intervals.