

UNIVERSITY OF STRATHCLYDE

DOCTORAL THESIS

---

**Essays in Macroeconomic and  
Financial Forecasting using Big  
Data Econometric Methods**

---

*Author:*

Aristeidis RAFTAPOSTOLOS

*Supervisors:*

Professor Gary KOOP

Professor Stuart MCINTYRE

A thesis submitted in fulfillment of the requirements  
for the degree of Doctor of Philosophy

Department of Economics  
University of Strathclyde

September, 2022

# Declaration of Authorship

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: September 8, 2022  
\_\_\_\_\_

# Statement of Co - Authorship

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

- Candidate: Aristeidis Raftapostolos, University of Strathclyde
- Author 1: Ilias Chronopoulos, University of Essex
- Author 2: George Kapetanios, King’s College London
- Author 3: Gary Koop, University of Strathclyde
- Author 4: Stuart McIntyre, University of Strathclyde
- Author 5: James Mitchell, Federal Reserve Bank of Cleveland

Contribution of work by co-authors for each paper:

PAPER 1: Located in Chapter 3.

Koop, McIntyre, Mitchell and Raftapostolos (2022) Monthly GDP Growth Estimates for the U.S. States (Working Paper).

Author contributions:

- Co-writing, Methodology: Candidate, Author 3, Author 4, Author 5
- Coding, Estimation: Candidate
- Data Curation: Candidate, Author 4, Author 5.

PAPER 2: Located in Chapter 4.

Chronopoulos, Raftapostolos and Kapetanios (2022) Deep Quantile Regression (2<sup>nd</sup> revise and resubmit, Journal of Financial Econometrics).

Author contributions:

- Co-writing, Methodology: Candidate, Author 1, Author 2
- Coding, Estimation, Data Curation: Candidate, Author 1.

Whereas Chapter 2 is solo authored, Chapters 3 and 4 (of my thesis) are joint work. Chapter 3 is co-authored with Gary Koop, Stuart McIntyre and James Mitchell and is a working paper. Chapter 4 is co-authored with Ilias Chronopoulos and George Kapetanios and is under review in the Journal of Financial Econometrics. 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.

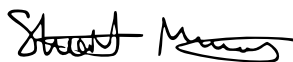
Signed:



Aristeidis Raftapostolos,  
University of Strathclyde

Ilias Chronopoulos,  
University of Essex

George Kapetanios,  
King's College London



Gary Koop,  
University of Strathclyde

Stuart McIntyre,  
University of Strathclyde

James Mitchell,  
Federal Reserve Bank of  
Cleveland

Date: September 8, 2022

---



*“The first and greatest victory is to conquer yourself; to be conquered by yourself is of all things most shameful and vile.”*

Plato

# *Abstract*

This thesis explores several aspects of econometric methods in time series forecasting of both macroeconomic and financial variables. The contribution is provided in three essays.

The first essay (Chapter 2) contributes to the econometric literature and develops models for regional nowcasting. We use Bayesian mixed frequency methods estimated at the common lower frequency. Moreover, we propose a procedure which allows model estimation with stochastic volatility and large datasets. We produce high frequency state-level GDP nowcasts that will assist policymakers in understanding the impact of greater regionalisation on economic growth in the U.S., and evaluate its impact on present and future economic conditions in a more timely fashion. We evaluate the accuracy of point and density forecasts, by making comparisons across models with constant and stochastic volatility. We provide results on the accuracy of nowcasts of real-time economic growth in the U.S. from 2006 to 2018. Empirical results suggest that models with stochastic volatility outperform models with constant volatility at nowcasting.

The second essay (Chapter 3) develops a Mixed Frequency Vector Autoregressive model (MF-VAR) for producing timely monthly nowcasts and historical estimates of GDP growth at the state level in the U.S. economy. The variables in the MF-VAR include GDP growth at the state and country level, as well as additional monthly variables at the state and country level. The variables are observed at different frequencies, leading to a complicated high-dimensional MF-VAR. A computationally-fast approximate Bayesian Markov Chain Monte Carlo (MCMC) algorithm is proposed for estimating the MF-VAR coefficients and nowcasting. Empirical results explore the nature and magnitude of spillover effects among the U.S. states. Further, the proposed model produces historical estimates at monthly frequency for both the U.S. economy and U.S. states.

The third essay (Chapter 4) proposes a deep quantile estimator, using neural networks and their universal approximation property to examine a non-linear association between the conditional quantiles of a dependent variable and predictors. The proposed methodology is versatile and allows both the use of different penalty functions, as well as high dimensional covariates. We present a Monte Carlo exercise where we

examine the finite sample properties of the proposed estimator and show that our approach delivers good finite sample performance. We use the deep quantile estimator to forecast Value-at-Risk and find significant gains over linear quantile regression alternatives, supported by various testing schemes. The Chapter also contributes to the interpretability of neural networks output by making comparisons between the commonly used SHAP values and an alternative method based on partial derivatives.

# *Acknowledgements*

I would like to thank my supervisor, Gary Koop, for giving me the opportunity to work alongside of him, for trusting me with important projects and allowing me to become an independent and responsible researcher by not micro-managing me.

I would also like to thank Stuart McIntyre, who is also my supervisor, for being supportive and quick to provide any advice I may require.

I would also like to thank George Kapetanios for being very kind and supporting as a co-author. He has been a role model and has expanded my research horizons.

Many thanks go to Ilias Chronopoulos, Florian Huber, James Mitchell, Aubrey Poon and participants at various conferences.

This work was supported by the Economic and Social Research Council and I would like to thank them for their financial support.

Finally, I would like to thank my family and friends for their encouragement, understanding and support throughout my studies.

# Contents

<b>Declaration of Authorship</b>	<b>i</b>
<b>Statement of Co - Authorship</b>	<b>ii</b>
<b>Abstract</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>Contents</b>	<b>viii</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background	1
1.2 Contribution	4
<b>2 Regional nowcasting: Evidence from the U.S.</b>	<b>6</b>
2.1 Introduction	6
2.2 Model Setup	9
2.2.1 Homoskedastic MF-VAR	10
2.2.2 MF-VAR with Stochastic Volatility	11
2.2.3 Priors	13
2.2.3.1 Independent Prior	13
2.2.3.2 Stochastic Volatility	16
2.2.4 Stochastic Search Variable Selection (SSVS)	17
2.2.5 Horseshoe	17
2.2.6 Prior Elicitation	19
2.2.7 Curse of dimensionality	20

2.2.8	Entropic Tilting with quarterly national-level GDP releases . . . .	20
2.3	Empirical application . . . . .	24
2.3.1	Data . . . . .	24
2.3.2	Definitions . . . . .	25
2.3.3	Forecast Evaluation . . . . .	26
2.3.4	Robustness check . . . . .	28
2.3.5	Empirical Results . . . . .	29
2.4	Conclusion . . . . .	30
A.1	Appendix (Chapter 2) . . . . .	32
<b>3</b>	<b>Nowcasting Monthly GDP Growth for U.S. States Using a Mixed Frequency VAR</b>	<b>44</b>
3.1	Introduction . . . . .	44
3.2	Econometric Methods . . . . .	46
3.2.1	Notation and Data Observability . . . . .	46
3.2.2	The MF-VAR . . . . .	47
3.2.3	The Prior . . . . .	50
3.2.4	State Space setup . . . . .	52
3.2.5	A Computationally Efficient Approximate MCMC Algorithm . .	54
3.2.5.1	Monte Carlo . . . . .	57
3.3	Empirical Application . . . . .	59
3.3.1	Data . . . . .	59
3.3.2	Historical Estimates of monthly regional growth . . . . .	59
3.3.3	Measuring the connectedness of U.S. states . . . . .	60
3.4	Conclusion . . . . .	65
A.1	Data Appendix (Chapter 3) . . . . .	66
B.1	Appendix (Chapter 3) – Exact Growth Rates . . . . .	69
<b>4</b>	<b>Deep Quantile Regression</b>	<b>79</b>
4.1	Introduction . . . . .	79
4.2	Theory . . . . .	84
4.2.1	Linear Quantile Regression . . . . .	84
4.2.2	Neural Networks . . . . .	86
4.2.3	Non-linear Quantile Regression . . . . .	87
4.2.4	Regularized Non-Linear Quantile Regression . . . . .	88
4.2.4.1	Regularization . . . . .	89
4.2.4.2	Cross Validation . . . . .	89
4.2.4.3	Optimisation . . . . .	91
4.3	Monte Carlo . . . . .	91
4.3.1	Setup . . . . .	91
4.3.2	Results . . . . .	93
4.4	Empirical Setup . . . . .	94
4.4.1	Deep Quantile <i>VaR</i> forecasting . . . . .	94

---

4.4.2	Forecasting Exercise Design . . . . .	96
4.4.3	Forecast Evaluation . . . . .	97
4.4.3.1	Diebold Mariano Test . . . . .	97
4.4.3.2	Giacomini White Test . . . . .	98
4.4.3.3	Conditional Quantile Forecast Encompassing (CQFE) . . . . .	99
4.5	Semi-Structural analysis . . . . .	101
4.5.1	Shapley values . . . . .	102
4.5.2	Partial Derivatives . . . . .	103
4.5.3	Results . . . . .	104
4.6	Conclusion . . . . .	106
A.1	Appendix (Chapter 4) . . . . .	108
<b>5</b>	<b>Conclusion</b>	<b>119</b>
	<b>Bibliography</b>	<b>121</b>

# List of Figures

2.1	Comparison of point forecast accuracy. Each panel describes a US State with quarterly and annual data. The x axis reports the RMSFE for BVAR, the y axis reports the RMSFE for BVAR-SV. Each point corresponds to a different forecast horizon. . . . .	41
2.2	Comparison of point forecast accuracy. Each panel describes a US State with quarterly and annual data. The x axis reports the RMSFE for BVAR, the y axis reports the RMSFE for BVAR-SV. Each point corresponds to a different forecast horizon. . . . .	42
2.3	Comparison of point forecast accuracy. Each panel describes a US State with quarterly and annual data. The x axis reports the RMSFE for BVAR, the y axis reports the RMSFE for BVAR-SV. Each point corresponds to a different forecast horizon. . . . .	43
3.1	Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC) . . . . .	69
3.2	Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC) . . . . .	70
3.3	Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC) . . . . .	71
3.4	Historical Estimates . . . . .	72
3.5	Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC) . . . . .	73
3.6	Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC) . . . . .	74
3.7	Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC) . . . . .	75
3.8	Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC) . . . . .	76
3.9	Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC) . . . . .	77
3.10	Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC) . . . . .	78



- 4.1 Monte Carlo results for Case I. Model:  $y_t = h_\tau(x_t) + u_t$ ,  $h_\tau(x_t) = \sin(2\pi x_t)$ ,  $x_t \sim N(0,1)$ ,  $u_t \sim iidN(-\sigma\Phi^{-1}(\tau), \sigma^2)$ ,  $\sigma = 0.1$  and  $\Phi^{-1}$  is the quantile function of the standard normal distribution. Figure presents the average mean squared error of the estimated residuals,  $AMSE_{\hat{u}_t, pen}$  for the different penalization schemes (Model),  $T = 100, 300, 500, 1000, 2000, 5000$  and different quantiles,  $\tau = (1\%, 2.5\%, 5\%, 10\%, 20\%)$ . . . . . 109
- 4.2 Monte Carlo results for Case II. Model:  $y_t = h_\tau(x_t) + u_t$ ,  $h_\tau(x_t) = \sin(2\pi x_t)$ ,  $x_t = 0.8x_{t-1} + \varepsilon_t$ ,  $\varepsilon_t \sim N(0,1)$ ,  $u_t \sim iidN(-\sigma\Phi^{-1}(\tau), \sigma^2)$ ,  $\sigma = 0.1$  and  $\Phi^{-1}$  is the quantile function of the standard normal distribution. Figure presents the average mean squared error of the estimated residuals,  $AMSE_{\hat{u}_t, pen}$  for the different penalization schemes (Model),  $T = 100, 300, 500, 1000, 2000, 5000$  and different quantiles,  $\tau = (1\%, 2.5\%, 5\%, 10\%, 20\%)$  . . . . . 110
- 4.3 Monte Carlo results for Case III. Model:  $y_t = h_\tau(x_t) + u_t$ ,  $h_\tau(x_t) = \sin(2\pi x_t)$ ,  $x_t = \sigma_t \varepsilon_t$ ,  $\sigma_t^2 = 1 + 0.7x_{t-1}^2 + 0.2\sigma_{t-1}^2$ ,  $\varepsilon_t \sim N(0,1)$ ,  $u_t \sim iidN(-\sigma\Phi^{-1}(\tau), \sigma^2)$ ,  $\sigma = 0.1$  and  $\Phi^{-1}$  is the quantile function of the standard normal distribution. Figure presents the average mean squared error of the estimated residuals,  $AMSE_{\hat{u}_t, pen}$  for the different penalization schemes (Model),  $T = 100, 300, 500, 1000, 2000, 5000$  and different quantiles,  $\tau = (1\%, 2.5\%, 5\%, 10\%, 20\%)$  . . . . . 111
- 4.4 Monte Carlo results for Case IV. Model:  $y_t = h_\tau(x_t) + u_t$ ,  $h_\tau(x_t) = G(x_t, \mathbf{w})$ ,  $x_t \sim N(0,1)$ ,  $w_{i,j} = \delta_{i,j} 1(\delta_{i,j} > 0.1)$ ,  $\delta_{i,j} \sim U(0,1)$ ,  $u_t \sim iidN(-\sigma\Phi^{-1}(\tau), \sigma^2)$ ,  $\sigma = 0.1$  and  $\Phi^{-1}$  is the quantile function of the standard normal distribution. Figure presents the average mean squared error of the estimated residuals,  $AMSE_{\hat{u}_t, pen}$  for the different penalization schemes (Model),  $T = 100, 300, 500, 1000, 2000, 5000$  and different quantiles,  $\tau = (1\%, 2.5\%, 5\%, 10\%, 20\%)$  . . . . . 112
- 4.5 Partial Derivative, SHAP and  $\hat{\beta}(\tau)$  for GARCH and RM models. . . . . 115
- 4.6 Partial Derivative, SHAP and  $\hat{\beta}(\tau)$  for SV model. . . . . 116
- 4.7 Partial Derivative, SHAP and  $\hat{\beta}(\tau)$  for ASV model. . . . . 117
- 4.8 Partial Derivative, SHAP and  $\hat{\beta}(\tau)$  for ASV model. . . . . 118

# List of Tables

2.1	Model Acronyms . . . . .	28
2.2	Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model. . . . .	33
2.3	Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model. . . . .	34
2.4	Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model. . . . .	35
2.5	Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model. . . . .	36
2.6	Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model. . . . .	37
2.7	Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model. . . . .	38

2.8	Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model. . . . .	39
2.9	Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model. . . . .	40
3.1	DGP I: X is i.i.d. . . . .	58
3.2	DGP II: X is Real data . . . . .	58
3.3	Connectedness in percent (normalized to sum to 100) at $h = 1$ months . . . . .	63
3.4	Connectedness in percent (normalized to sum to 100) at $h = 12$ months . . . . .	64
3.5	The tcode denotes the following data transformation for a series $x$ : (1) no transformation; (2) $\Delta x_t$ ; (3) $\Delta^2 x_t$ ; (4) $\log(x_t)$ ; (5) $\Delta \log(x_t)$ ; (6) $\Delta^2 \log(x_t)$ ; (7) $\Delta^2 (x_t/x_{t-1} - 1)$ ; (8) $(x_t/x_{t-1} - 1)$ . The FRED column gives mnemonics in FRED followed by a short description. Non-seasonal adjusted series were transformed to seasonal adjusted by applying X11. XX denotes the short code for U.S. state . . . . .	68
4.1	Comparison of the forecasting methods. Table reports relative RMSFE. The smaller the entry ( $< 1$ ) the better the forecast. *, **, and *** denote results from Diebold and Mariano (1995) test with the Harvey, Leybourne, and Newbold (1997) adjustment for predictive accuracy, indicating rejection of the null hypothesis that the models have the same predictive accuracy at the 10%, 5%, and 1% levels of significance, respectively. †, ††, and ††† denote results from the Giacomini and White (2006) test, indicating rejection of the null hypothesis of equal forecasting accuracy of the models at the 10%, 5%, and 1% levels of significance, respectively. . . . .	113
4.2	Entries of the table present the number of times a block encompasses (wins), does not encompass (losses) and is inconclusive, according the CQFE test for different quantiles $\tau$ . Results are reported for $\eta = 0.005$ . . . . .	114

*To my Parents*

# Chapter 1

## Introduction

### 1.1 Background

Classical econometrics is based on the fact that a data generating process exists, for instance  $Y_1, \dots, Y_T \sim \mathcal{D}(\mu, 1)$ , and the unknown parameter(s) can be estimated; in this example the unknown parameter is the mean  $\mu$ . This analysis can be generalised for any distribution  $f(X; \beta)$  with some parameters  $\beta$  and a data generating process. In this case, classical econometricians seek to estimate these parameters;  $\hat{\beta}$ , retrieve their distribution and perform some hypothesis testing.

In contrast, Bayesian econometricians are interested in the conditional distribution of the unknown parameter(s),  $\beta|X$ , that is considered as a random variable given our sample.

Normally, methods in the science of philosophy can also be used by the statistical science, thus the identified research philosophy of this thesis belongs to the so-called Bayesian Epistemology. Currently, econometrics is related to the study of causal relationships among several variables and to the philosophical propositions that constitute empirical modelling. The main objective of econometric science is to interpret economic phenomena such as the relation between unemployment and inflation. Bayesian econometrics fall under inductive logic principles, since inference is made from observations to parameters, in other words our beliefs about future are bound by past and present beliefs.

The concept of epistemology deals with the knowledge and the conclusions of that, answer questions, which are based on the views about reality; this formulates the traditional epistemology. Bayesian epistemology uses tools from the probability theory and studies the degrees of belief i.e. probabilities. On the other hand, epistemology examines the non-probabilistic notion of knowledge that builds upon rational arguments.

The origin of Bayesian epistemology records back to the Reverend Thomas Bayes (c. 1701 - 1761) who expressed an important probability theorem (regarding conditional probabilities – Bayes Rule Theorem) that laid the foundation of Bayesian statistics and further progress the Bayesian philosophy. In recent years, Bayesian Econometrics found many applications to major fields of economics and related disciplines such as macroeconomics. In Bayesian Econometrics, researchers combine a prior distribution with the observed data and extract a posterior distribution. Any term that is not included in the posterior is irrelevant and can't be used for falsification. In a nutshell, evidence is summarised by the posterior rather than employing confidence intervals and hypothesis tests as frequentists do.

[Hartmann and Sprenger \(2010\)](#) indicate the three pillars of Bayesian Epistemology namely, the Bayesian Conditionalization, the Dutch Book Argument and the Principal Principle. Conditionalisation refers to the way beliefs (priors) can be updated with more information. In Dutch Book there is a link between probabilities and betting. By building upon this link, the Dutch Book Argument says that if the law of probabilities does not hold, then in a bet can exist profits and losses; in other words the assumption that a bet is fair is violated. Finally, the Principal principle states that if an agent has a probability  $p$  for an event and there is no other information, then the probability should be equal to  $p$ .

This thesis focuses on the Bayesian Conditionalization as it examines how different types of prior elicitation can be used to improve the forecasting power for the quarterly U.S. states. Specifically, we use the various prior setups such the independent prior (see e.g. [Koop and Korobilis \(2010\)](#)), the Stochastic Search Variable Selection (SSVS) (see e.g. [George, Sun, and Ni \(2008b\)](#)) and the horseshoe prior (see e.g. [Carvalho, Polson, and Scott \(2009\)](#), [Carvalho, Polson, and Scott \(2010\)](#)) and evaluate their forecasting power, so that we can produce more timely and frequent quarterly U.S. GDP by state forecasts.

Bayesian epistemology is based on the Kolmogorov's probability axioms, in which a probability function should have the properties of non-negativity, normalization (unity measure) and finite additivity. Based on these we use Bayes rule:

$$P(B|A) = \frac{P(A|B) * P(A)}{P(B)} \quad (1.1)$$

$$P(\beta|Y) \propto P(Y|\beta) * P(\beta), \quad (1.2)$$

where  $P(\beta|Y)$  is the posterior function,  $P(Y|\beta)$  is the likelihood function and  $P(\beta)$  is the prior distribution.

In 1992, Jeffrey, argued that Bayesian epistemology has two main advantages, the first is the use of subjective probabilities i.e. prior beliefs may not be bound to rational restrictions and secondly, observations generate a range of probabilities and not a certainty about an event. Firstly, in our empirical application in Chapter 2 we start with the Minnesota (or Litterman) prior, namely the independent Normal – Wishart. The specification for our priors follows the work of [Dieppe, Legrand, and Van Roye \(2016\)](#). Finally, in order to avoid using a subjective prior we calculate the marginal likelihood for the Minnesota prior and then we select the hyperparameters that maximise its value. [Giannone, Lenza, and Primiceri \(2015\)](#) showed that this technique mitigates the subjectiveness, while selecting the prior.

A key concept in Bayesian epistemology is evidence, which provides the tools for confirming or not a theoretical model. [Koop, McIntyre, and Mitchell \(2019\)](#) used a stacked Mixed frequency Vector Autoregression (MF-VAR) to produce more frequent (quarterly instead of annual) and more timely regional gross value added (GVA) nowcasts for 10 UK regions (including UK continental shelf). They combine quarterly UK GVA with annual GVA for the 10 UK regions and deploy both homoskedastic and heteroskedastic MF-VAR models with entropic tilting. Their empirical application produces quarterly GVA by region every quarter works fine and their nowcast estimates indeed improve as they add new information each quarter. [Arias, Gascon, and Rapach \(2016\)](#) deploy a dynamic factor model to produce a monthly economic indicator for the 50 largest metropolitan statistical areas (MSA). U.S. population per area in 2014 was the selection criterion for the 50 MSA. Their dataset consists of 12 variables and spans from 1990 to 2015. Their dynamic factor model embeds the maximum-likelihood method proposed by [Bańbura and Modugno \(2014\)](#) to accommodate mixed

frequency data and differences in data-publication lags. They find that in Great Recession almost all MSA entered a recession.

Bayesian econometrics are relying on building a model and on statistical analysis; both belonging to the empiricist philosophy. Empiricism refers to an epistemological stance and its concept states that knowledge relates to past sense experiences and empirical evidence. Some empiricists claim that sensory knowledge can be achieved by latent data. As expected, unobserved evidence can create bias while interpreting results, so empiricists proposed the use of a hypothetico-deductive approach, under which a hypothesis (in our case the prior) is stated and is being tested (viz the way we update our prior beliefs) in an empirical application. What is more, empiricism in econometrics is interpreted as studies on causal identification and as the tools used to make inference on specific subpopulations. One critic of empiricism is the focus on identification rather than building or developing an underlying theory. Besides, the focus on interpretation and analysis of observed data only greatly restricts an empirical experiment, as it prohibits a generalization.

## 1.2 Contribution

This thesis contributes to both the macroeconomic and financial forecasting literatures using Big Data econometric methods in the following ways: first, we discuss possible prior setups using a Mixed frequency Vector Autoregression (MF-VAR) written at the common lower frequency. Second, we use ideas from recent work in estimating high frequency MF -VAR models to discuss how monthly indicators offer more information on producing one-step ahead U.S. GDP by state forecasts. Finally, we drop the linearity assumption and using machine learning, we propose a deep quantile estimator that is both non-linear and non-parametric estimator and forecast Value-at-Risk.

Chapter 2 contributes to the econometric literature and develops models for regional nowcasting. We use Bayesian mixed frequency methods estimated at the common lower frequency. Moreover, we propose a procedure which allows model estimation with stochastic volatility and large datasets. We produce high frequency state-level GDP nowcasts that will assist policymakers in understanding the impact of greater regionalisation on economic growth in the U.S. and evaluate its impact on present and future economic conditions in a more timely fashion. We evaluate the accuracy of point and density forecasts, by making comparisons across models with constant



and stochastic volatility. We provide results on the accuracy of nowcasts of realtime economic growth in the U.S. from 2006 to 2018. Empirical results suggest that models with stochastic volatility outperform models with constant volatility at nowcasting.

The second essay (Chapter 3) develops a Mixed Frequency Vector Autoregressive model (MF-VAR) for producing timely monthly nowcasts and historical estimates of GDP growth at the state level for the U.S. economy. The variables in the MF-VAR include GDP growth at the state and country level, as well as additional monthly variables at the state and country level. The variables are observed at different frequencies, leading to a complicated high-dimensional MF-VAR. A computationally-fast approximate Bayesian Markov Chain Monte Carlo (MCMC) algorithm is proposed for estimating the MF-VAR coefficients and nowcasting. Empirical results explore the nature and magnitude of spillover effects among the U.S. states. Further, the proposed model produces historical estimates at monthly frequency for both the U.S. economy and U.S. states.

The third essay (Chapter 4) proposes a deep quantile estimator, using neural networks and their universal approximation property to examine a non-linear association between the conditional quantiles of a dependent variable and predictors. The proposed methodology is versatile and allows both the use of different penalty functions, as well as high dimensional covariates. We present a Monte Carlo exercise where we examine the finite sample properties of the proposed estimator and show that our approach delivers good finite sample performance. We use the deep quantile estimator to forecast Value-at-Risk and find significant gains over linear quantile regression alternatives, supported by various testing schemes. The Chapter also contributes to the interpretability of neural networks output by making comparisons between the commonly used SHAP values and an alternative method based on partial derivatives.

## Chapter 2

# Regional nowcasting: Evidence from the U.S.

### 2.1 Introduction

Regional policymakers' decisions are based on evaluating present and future economic conditions using incomplete information sets. Incomplete information arises from the fact that real-time data is released with a considerable publication lag. This research aims to develop models for improving regional nowcasting i.e. forecasting the present and near future of the U.S. state-level economic variables, like Gross Domestic Product (GDP), using mixed frequency 'Big Data' techniques. These estimates will ultimately assist economists at central banks, state governments, as well the financial sector in implementing policy decisions and in evaluating its impact by providing state-level nowcasts in a more timely fashion.

Currently, GDP measures for the U.S. state-level is produced by the Bureau of Economic Analysis (BEA) on an annual basis with a considerable publication lag of around a year. As an example, this means that in November (e.g. 2019) each year, we are provided with estimates of state-level economic growth in the previous year (e.g. 2018). Thus, for long periods, this 'ragged/jagged' edge, as known in the literature, makes the evaluation of the impact of policy interventions very difficult without an accurate estimate of what current GDP is, much less what it will be in the near future. This incomplete information impairs the ability of state policymakers to respond to changing economic conditions. However, many potential predictors for U.S. GDP are

released on a quarterly or monthly frequency. This frequency mismatch can be utilised to update nowcasts or forecasts more regularly than the current practice of producing state-level GDP at annual frequency. Precisely, U.S. state-level GDP forecasts can be produced every quarter.

There exists a substantial literature on mixed frequency econometrics that we plan on extending in this work. The U.S. growth nowcasting context differs from the existing literature in several important ways. First, there are many more low frequency variables (the state-level GDP variables) than is usual in the literature. Second, the fact that GDP for the U.S. states adds up to U.S. GDP provides an extra cross-sectional restriction which should improve forecasting by producing conditional forecasts. To be more precise, cross-sectional restriction is one way to move beyond unconditional forecast and produce "soft" conditional forecasts, see e.g. [Krüger, Clark, and Ravazzolo \(2017\)](#) and [Waggoner and Zha \(1999\)](#). Finally, there is a huge range of variables and various frequencies which could be used to improve GDP forecasts (e.g. regional labour market variables at the annual frequency, U.S. macroeconomic variables at the quarterly frequency and regional survey data at the monthly frequency). Although we plan on starting small (i.e. working with 51 U.S. states and quarterly U.S. GDP), adding other potentially useful predictors leads to an increase in the dimensionality of the problem and introduce issues in the estimation of the model's coefficients. We plan on addressing all these issues in our research. To be more precise, we use, and develop statistical models used in producing estimates of current and future values of economic variables at a national level with the added provision that these models are adapted to ensure that state estimates are consistent with national level forecasts.

We provide a brief review on methodology. [Sims \(1980\)](#) in his seminal work introduced vector autoregressive (VAR) models, which provided a new statistical framework that allows the examination of economic co-movements among several macroeconomic variables. Since then, VARs are used as the workhorse models in macroeconomics mainly for their capacity in data summarizing, forecasting and in capturing rich system dynamics. It is clear that VARs are not parsimonious models and their identification is crucial. However, Bayesian techniques and innovations in the machine learning literature assisted in solving the problem of dimensionality. Furthermore, in their seminal work [Bańbura, Giannone, and Reichlin \(2010\)](#) show that the estimation of fairly large Bayesian VARs (around 130 variables) is feasible and gives sufficient forecasting performance. This work laid the foundations to use large BVARs

in forecasting with many variables. As already noted, this research uses a large number of predictors and methods that allow for good estimation in high-dimensions are essential. Lastly, we experiment with different model specifications in order to produce more accurate nowcasts and forecasts.

Since GDP growth is released at quarterly frequency it has been the norm to estimate the parameters in a VAR model at quarterly basis, but due to the fact that many macroeconomic variables are released at monthly (or even higher frequency) basis we can mix high frequency variables with lower ones to allow for a faster updating of nowcasts. To provide such estimates, we draw on and extend the growing literature on mixed frequency Vector Autoregressions, henceforth MF-VARs. As proposed by [Ghysels and Marcellino \(2018\)](#) there are two main families of MF-VARS used in the literature. The first is called observation driven and consists of the stacked MF-VAR approach, where we write the MF-VARs at the lowest frequency. Pioneering work includes [McCracken, Owyang, and Sekhposyan \(2019\)](#) and the MF-MIDAS, where we write the VARs at low-frequency but we write high frequency indicators using parsimonious distributed lag polynomials, see the seminal work of [Ghysels \(2016\)](#). The latter is called parameter driven viz state space MF-VAR, we write the MF-VAR at the high frequency and use state space methods to fill latent low frequency variables as in the pioneering work of [Schorfheide and Song \(2015\)](#).

We extend the MF-VAR model and use entropic tilting to produce conditional forecasts for the U.S. state-level GDP in a more timely fashion by incorporating information from the quarterly releases of nationwide GDP. The benefit of using entropic tilting is twofold; first it improves the accuracy of point nowcast and second the density nowcast. Empirical results suggest that entropic tilting improves unconditional forecasts, however the inclusion of stochastic volatility seems to have minor improvements. Similar to our approach, [Koop, McIntyre, and Mitchell \(2019\)](#) used a stacked MF-VAR to produce more frequent (quarterly instead of annual) and more timely regional gross value added (GVA) nowcasts for 10 UK regions (including the UK continental shelf). They combine quarterly UK GVA with annual GVA for the 10 UK regions and deploy both homoskedastic and heteroskedastic MF-VAR models with entropic tilting. Their empirical application suggests that producing quarterly GVA by region every quarter works relatively well and nowcast estimates indeed improve as they add new information each quarter.

[Arias, Gascon, and Rapach \(2016\)](#) use a dynamic factor model to produce a monthly economic indicator for the 50 largest Metropolitan Statistical Areas (MSA). Their dataset consists of the following 12 variables and spans from 1990 to 2015. Their dynamic factor model embeds the maximum-likelihood method proposed by [Bańbura and Modugno \(2014\)](#) that allows mixed frequency data and differences in data-publication lags. They find that in the Great Recession almost all MSA entered a recession. In addition, some MSA faced recessions in early 1990s and 2000s (associated with the so-called dot-com bubble).

In summary, this work contributes to the econometric literature in developing models for regional nowcasting and forecasting and in understanding the impact of greater regionalisation of economic growth and productivity in the U.S.. In order to know what the impact of this policy is in a timely fashion that is useful for policy makers, investment in improving the economic data available is essential.

The plan of the remainder of this Chapter is as follows. Section [2.2](#) describes our mixed frequency and entropic tilting methodology. Section [2.3](#) delineates our dataset, empirical application and forecast evaluation. Conclusions are set out in Section [2.4](#).

## 2.2 Model Setup

In this Section we provide a brief explanation of the estimation method used and model specifications, as well as information regarding the priors and the MCMC algorithms. Initially, a homoskedastic Bayesian MF-VAR is used, but because of the financial crisis many macroeconomic and financial variables depart from the classical normality assumption, so we plan to enrich our model with stochastic volatility. Since the pioneering work of [Bańbura, Giannone, and Reichlin \(2010\)](#) it is common to use the Minnesota prior while working with large VARs, but our sample size is only 55 and initially the number of variables in each equation is 56 (i.e. we are in the  $K > T$  world) our results may not be sensible as it considers the variance,  $\Sigma$ , known, so we use the independent prior instead of the traditional Minnesota prior. We also use the Stochastic Search Variable Selection (SSVS) prior proposed by [George, Sun, and Ni \(2008a\)](#) and the horseshoe prior used by [Carvalho, Polson, and Scott \(2009\)](#).

### 2.2.1 Homoskedastic MF-VAR

In their work [McCracken, Owyang, and Sekhposyan \(2019\)](#) based their forecasting model on a stacked MF-VAR model estimated at the common lower frequency. This approach has the advantage that it uses the traditional VAR estimating techniques that generally work well. To be more precise, a MF-VAR written at the common lower frequency is a traditional VAR, where all data is collected at a common low frequency (i.e. there are not latent variables in this specification) e.g. at quarterly frequency, and standard VAR estimation procedures are used, see e.g. [McCracken, Owyang, and Sekhposyan \(2019\)](#) and [Ghysels \(2016\)](#).

Our model in reduced form is written as follows:

$$y_t = \tilde{B}_0 + \tilde{B}_1 y_{t-1} + \dots + \tilde{B}_p y_{t-p} + \tilde{\varepsilon}_t \quad (2.1)$$

where  $y_t = (y_{1,t}, y_{2,t}, \dots, y_{n,t})'$ , is a  $n \times 1$  vector with endogenous variables at time  $t$ ,  $\tilde{B}_0$  a  $n \times 1$  vector of constants,  $\tilde{B}_i$   $n \times n$  parameter matrices,  $p$  is the lag order of VAR, and  $\tilde{\varepsilon}_t \sim N(0, \tilde{\Sigma})$ ;  $\tilde{\Sigma}$  is a full variance covariance matrix.

[Chan \(2019\)](#) and [Carriero, Clark, and Marcellino \(2019\)](#) proposed a reparameterisation of (2.1) in which the VAR model is written in its structural form. Our MFBVAR in structural form is denoted as follows:

$$Ay_t = B_0 + B_1 y_{t-1} + \dots + B_p y_{t-p} + \varepsilon_t, \quad (2.2)$$

where  $\varepsilon_t \sim N(0, \Sigma)$ ,  $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$  and  $A_i$  is  $n \times n$  lower triangular impact matrix with ones in the main diagonal. The parameterisation of  $\Sigma$  as diagonal allows a recursive estimation of the model equation-by-equation without efficiency losses as illustrated in [Carriero, Clark, and Marcellino \(2019\)](#).

For our forecasting exercise we need to recover the reduced form VAR coefficients as follows.  $\tilde{B}_0 = A^{-1}B_0$ ,  $\tilde{B}_i = A^{-1}B_i$  and  $\tilde{\Sigma} = A^{-1}\Sigma A^{-1'}$ .

We provide some extra definitions,  $B_{0,i}$  denotes the  $i^{\text{th}}$  element of  $B_0$ ,  $B_{j,i}$  denotes the  $i^{\text{th}}$  row of  $B_j$ ,  $\beta_i = (B_{0,i}, B_{1,i}, \dots, B_{p,i})'$  and  $a_i$  is the  $i^{\text{th}}$  row of  $A$  and is equal with  $\alpha_i = (A_{i,1}, A_{i,2}, \dots, A_{i,i-1})'$ .

Finally our model is written as:

$$y_{i,t} = w_{i,t}\alpha_i + x_{i,t}\beta_i + \varepsilon_{i,t}, \quad (2.3)$$

where  $\varepsilon_{i,t} \sim N(0, \sigma_i^2)$ ,  $w_{i,t} = (-y_{i,t}, \dots, -y_{i-1,t})'$  and  $x_{i,t} = (1, y_{t-1}, \dots, y_{t-p})'$ .

A more compact representation is:

$$y_{i,t} = x_{i,t}\theta_i + \varepsilon_{i,t}, \quad (2.4)$$

where  $\varepsilon_{i,t} \sim N(0, \sigma_i^2)$ ,  $\theta_i = (\alpha_i', \beta_i')$  and the dimension of  $\theta_i$  is  $k_i = np + i$ .

Finally, we stack over t and we get:

$$y_i = x_i\theta_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_i^2 \mathbf{I}_T) \quad (2.5)$$

Having specified our model we need to derive the corresponding likelihood function of eq. (2.5)

$$f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\sigma}^2) = \prod_{i=1}^n f(\mathbf{y}_i|\boldsymbol{\theta}_i, \sigma_i^2) = \prod_{i=1}^n (2\pi\sigma_i^2)^{-T/2} e^{-\frac{1}{2\sigma_i^2}(\mathbf{y}_i - x_i\boldsymbol{\theta}_i)'(\mathbf{y}_i - x_i\boldsymbol{\theta}_i)}. \quad (2.6)$$

## 2.2.2 MF-VAR with Stochastic Volatility

In their empirical application [Clark and Ravazzolo \(2012\)](#) find that modelling stochastic volatility performs better than any alternative volatility specifications. To that end, the general framework of a MFBVAR with stochastic volatility (SV) model can be described by the following equations:

$$Y_t = B_0 + B_1 Y_{t-1} + \dots + B_p Y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \tilde{\Sigma}) \quad (2.7)$$

$$h_t = h_{t-1} + \varepsilon_t^h, \quad \varepsilon_t^h \sim \mathcal{N}(0, \sigma_h^2), \quad (2.8)$$

where  $\varepsilon_t$  and  $\varepsilon_t^h$  are mutually and serially uncorrelated. State  $h_t$  is the log-volatility and  $\text{var}(Y_t|h_t) = e^{h_t}$ .  $\tilde{\Sigma}$  is set to  $\text{diag}(e^{h_{1,t}}, e^{h_{2,t}}, \dots, e^{h_{n,t}})$ .

In a similar way as in eq. (2.5) we rewrite our model as follows.

$$y_i = x_i\theta_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, e^{h_{i,t}} I_T) \quad (2.9)$$

$$h_{i,t} = h_{i,t-1} + \epsilon_{i,t}^h, \quad \epsilon_{i,t}^h \sim \mathcal{N}(0, \Omega). \quad (2.10)$$

The state equation for the latent log-volatilities  $h_{i,t}$  follows a random walk process as in [Cogley and Sargent \(2005\)](#). Moreover, the initial condition  $h_0$  is estimated through the model.  $\Omega$  is set to  $\text{diag}(\sigma_{h_1}^2, \sigma_{h_2}^2, \dots, \sigma_{h_n}^2)$ , where contrary to [Primiceri \(2005\)](#) we do not assume time variation in the variance matrix.

We describe the conditional posterior for the stochastic volatility term, which follows the auxiliary mixture sampler proposed by [Kim, Shephard, and Chib \(1998\)](#). In their paper, they estimate a non-linear stochastic volatility model with the following equations:

$$y_t = \exp\left(\frac{1}{2}h_t\right)\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, 1) \quad (A.1)$$

$$\text{var}(y_t|h_t) = \exp(h_t)$$

$$h_t = h_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_h^2). \quad (A.2)$$

We take the logarithm of A.1 and get:

$$\log y_t^2 = h_t + \log \epsilon_t \quad (A.3)$$

$$y_t^* = h_t + \epsilon_t^*, \quad y_t^* = \log(y_t^2 + c) \quad \text{and} \quad \epsilon_t^* \sim X^2. \quad (A.4)$$

Offset variable  $c$  is set to 0.001 to ensure that the quasi-maximum likelihood estimator is very small. We then approximate the density of  $\epsilon_t^*$  by using the following  $n$ -component Gaussian mixture:

$$f(\epsilon_t^*) \approx \sum_{i=1}^n p_i \phi(\epsilon_t; \mu_i, \sigma_i^2), \quad (A.5)$$

where  $\phi(x; \mu, \sigma^2)$  represents a Gaussian density with mean  $\mu$  and variance  $\sigma^2$  and  $p_i$  the probability of  $i^{\text{th}}$  mixture component. Let's assume that  $s_t \in [1, \dots, n]$  is an auxiliary random variable, with the following mixture density:  $\epsilon_t^*|s_t = i \sim \mathcal{N}(\mu_i - 1.2704, \sigma_i^2)$  and  $\mathbb{P}(s_t = 1) = p_1$ . [Kim, Shephard, and Chib \(1998\)](#) propose a seven component Gaussian mixture as follows:

$$f(x) \approx \sum_{i=1}^7 p_i \phi(x; \mu_i - 1.2704, \sigma_i^2) \quad (A.6)$$



To finalise the step for SV we select the mean, variance and component probability of the mixing distribution to be  $\log \mathcal{X}^2$  as in Table 4 p.371 in Kim, Shephard, and Chib (1998) and run the Gibbs Sampler to get the posterior density. We draw the mixture component indicator for the  $i^{th}$  equation.

The algorithm we use for SV is in line with Kim, Shephard, and Chib (1998) and the precision sampling methods described in Chan and Jeliazkov (2009) and Chan, Hsiao, et al. (2014). To be in line with Koop, McIntyre, and Mitchell (2019), we set the lag length to 1, despite the fact that this assumption is not in line with the literature, see e.g. Koop (2013).

### 2.2.3 Priors

#### 2.2.3.1 Independent Prior

For our empirical application we use the Independent Normal – Wishart prior. In this framework the variance-covariance matrix  $\Sigma$  is assumed to be unknown. In total we have to estimate  $n * (np + 1) = 55 * 56$  parameters, which is a fairly large BVAR. We assume  $\theta_i \sim \mathcal{N}(\theta_{i,0}, \mathbf{V}_{\theta_i})$ .

The prior mean vector  $\theta_{i,0}$  is set to zero. We set to zero the first own lag coefficients as we work with growth rates. The prior covariance matrix  $\mathbf{V}_{\theta_i}$  is a diagonal matrix with elements specified as follows:

$$\begin{aligned} V_{ii} &= \left( \frac{\lambda_1}{l\lambda_3} \right)^2 \\ V_{ij} &= \left( \frac{s_i^2}{s_j^2} \right) \left( \frac{\lambda_1 \lambda_2}{l\lambda_3} \right)^2 \\ V_{c_i} &= s_i^2 (\lambda_1 \lambda_4)^2 \end{aligned}$$

where  $\lambda_1$  is the overall tightness parameter,  $l$  is the lag operator, and  $\lambda_3$  is the scaling coefficient controlling the speed at which coefficients for lags greater than 1 converge to 0,  $\lambda_2$  represents the cross-variable specific variance parameter and finally  $\lambda_4$  controls the tightness of constants. Lastly,  $s_i^2, s_j^2$  is the residual variance from a univariate AR(1) for the  $i^{th}, j^{th}$  equation of the VAR, respectively.

For the unknown  $\Sigma$  matrix we assume the following:

$$\sigma_i^2 \sim \mathcal{IW}(v_{i,0}, \mathbf{S}_{i,0}).$$

We choose a relative uninformative prior for the shape and set  $v_{i,0} = n + 2$  (minimum possible degrees of freedom so that we obtain well-defined mean and variance for  $\sigma_i^2$ ). The scale is selected to match  $\mathbb{E} \sigma_i^2 = (v_{i,0} - n - 1) \mathbf{S}_{i,0}$ , where  $\mathbf{S}_{i,0}$  is a diagonal matrix with its elements set to the residual variance from a univariate AR(1) for the  $i^{\text{th}}$  equation of the VAR.

The prior densities can be written as:

$$\pi(\boldsymbol{\theta}_i) = \prod_{i=1}^n (2\pi)^{-ki/2} |\mathbf{V}_{\theta_i}|^{-1/2} e^{-\frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0})' \mathbf{V}_{\theta_i}^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0})} \quad (2.11)$$

$$\pi(\boldsymbol{\sigma}_i) = \prod_{i=1}^n \frac{|\mathbf{S}_{i,0}|^{v_{i,0}/2}}{2^{v_{i,0}/2} \Gamma(v_{i,0}/2)} (\sigma_i^2)^{-(v_{i,0}+1+ki/2)} e^{-\frac{1}{2} \text{tr}(\frac{\mathbf{S}_{i,0}}{\sigma_i^2})}. \quad (2.12)$$

Our conditional posterior are formulated as follows:

$$\begin{aligned} \pi(\boldsymbol{\theta}_i | y_i, \sigma_i^2) &= f(\mathbf{y}_i | \boldsymbol{\theta}_i, \boldsymbol{\sigma}_i^2) \pi(\boldsymbol{\theta}_i) \\ &= (2\pi\sigma_i^2)^{-T/2} (2\pi)^{-ki/2} |\mathbf{V}_{\theta_i}|^{-1/2} e^{-\frac{1}{2\sigma_i^2} (\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\theta}_i)' (\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\theta}_i)} e^{-\frac{1}{2} (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0})' \mathbf{V}_{\theta_i}^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0})} \\ &\propto e^{-\frac{1}{2\sigma_i^2} (\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\theta}_i)' (\mathbf{y}_i - \mathbf{x}_i \boldsymbol{\theta}_i) - \frac{1}{2} (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0})' \mathbf{V}_{\theta_i}^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i,0})} \\ &= e^{-\frac{1}{2} \boldsymbol{\theta}_i' \mathbf{V}_{\theta_i}^{-1} \boldsymbol{\theta}_i - 2\boldsymbol{\theta}_i' \mathbf{V}_{\theta_i}^{-1} \boldsymbol{\theta}_{i,0} + \boldsymbol{\theta}_{i,0}' \mathbf{V}_{\theta_i}^{-1} \boldsymbol{\theta}_{i,0} + \mathbf{y}_i' \mathbf{y}_i - 2(\sigma_i^2)^{-1} \boldsymbol{\theta}_i' \mathbf{x}_i' \mathbf{y}_i + (\sigma_i^2)^{-1} \boldsymbol{\theta}_i' \mathbf{x}_i' \mathbf{x}_i \boldsymbol{\theta}_i} \\ &= e^{-\frac{1}{2} \boldsymbol{\theta}_i' (\mathbf{V}_{\theta_i}^{-1} + (\sigma_i^2)^{-1} \mathbf{x}_i' \mathbf{x}_i) \boldsymbol{\theta}_i - 2\boldsymbol{\theta}_i' (\mathbf{V}_{\theta_i}^{-1} \boldsymbol{\theta}_{i,0} + (\sigma_i^2)^{-1} \mathbf{x}_i' \mathbf{y}_i) + \boldsymbol{\theta}_{i,0}' \mathbf{V}_{\theta_i}^{-1} \boldsymbol{\theta}_{i,0} + (\sigma_i^2)^{-1} \mathbf{y}_i' \mathbf{y}_i} \\ &= e^{-\frac{1}{2} \boldsymbol{\theta}_i' \mathbf{K}_{\theta_i} \boldsymbol{\theta}_i - 2\boldsymbol{\theta}_i' \mathbf{K}_{\theta_i} \hat{\boldsymbol{\theta}}_i + \boldsymbol{\theta}_{i,0}' \mathbf{V}_{\theta_i}^{-1} \boldsymbol{\theta}_{i,0} + (\sigma_i^2)^{-1} \mathbf{y}_i' \mathbf{y}_i} \\ &= e^{-\frac{1}{2} (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)' \mathbf{K}_{\theta_i} (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i) - \hat{\boldsymbol{\theta}}_i' \mathbf{K}_{\theta_i} \hat{\boldsymbol{\theta}}_i + \boldsymbol{\theta}_{i,0}' \mathbf{V}_{\theta_i}^{-1} \boldsymbol{\theta}_{i,0} + (\sigma_i^2)^{-1} \mathbf{y}_i' \mathbf{y}_i} \\ &\propto e^{-\frac{1}{2} (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)' \mathbf{K}_{\theta_i} (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)}, \end{aligned} \quad (2.13)$$

where  $K_{\theta_i} = V_{\theta_i}^{-1} + (\sigma_i^2)^{-1} x_i' x_i$  and  $\hat{\theta}_i = K_{\theta_i}^{-1} (V_{\theta_i}^{-1} \theta_{i,0} + (\sigma_i^2)^{-1} x_i' y_i)$ .

$$\begin{aligned}
\pi(\sigma_i^2 | y_i, \theta_i) &= f(y_i | \theta_i, \sigma_i^2) \pi(\sigma_i^2) \\
&= (2\pi\sigma_i^2)^{-T/2} (2\pi)^{-ki/2} |V_{\theta_i}|^{-1/2} e^{-\frac{1}{2\sigma_i^2} (y_i - x_i \theta_i)' (y_i - x_i \theta_i)} \frac{|S_{i,0}|^{v_{i,0}/2}}{2^{v_{i,0}/2} \Gamma(v_{i,0}/2)} \\
&\times (\sigma_i^2)^{-(v_{i,0}+1+ki/2)} e^{-\frac{1}{2} \text{tr}(\frac{S_{i,0}}{\sigma_i^2})} \\
&\propto (\sigma_i^2)^{-(v_{i,0}+1+ki/2+T/2)} e^{-\frac{1}{2\sigma_i^2} (y_i - x_i \theta_i)' (y_i - x_i \theta_i)} e^{-\frac{1}{2} \text{tr}(\frac{S_{i,0}}{\sigma_i^2})} \\
&= (\sigma_i^2)^{-(v_{i,0}+1+ki/2+T/2)} e^{-\frac{1}{2} \text{tr}(\frac{S_{i,0} + (y_i - x_i \theta_i)' (y_i - x_i \theta_i)}{\sigma_i^2})}.
\end{aligned} \tag{2.14}$$

Our posterior analysis is summarised as follows. We use a two-block Gibbs Sampler to simulate draws from the posterior:  $\pi(\theta_i, \sigma_i^2 | y)$

1.  $\pi(\theta_i | \sigma_i^2, y)$
2.  $\pi(\sigma_i^2 | \theta_i, y)$

Steps 1-2 are summarised as follows.

$$\theta_i \sim \mathcal{N}(\hat{\theta}_i, K_{\theta_i}^{-1}) \tag{2.15}$$

$$\sigma_i^2 \sim \mathcal{IW}(v_{i,0} + T + ki, S_{i,0} + (y_i - x_i \theta_i)' (y_i - x_i \theta_i)). \tag{2.16}$$

Our econometric model faces the following challenges: i) the cross-sectional restriction i.e. state-level GDP adds up to national GDP; this restriction is the basis to improve our nowcasts, ii) we have many low frequency variables viz 51 annual GDP by state and one high frequency variable, namely quarterly GDP (Koop, McIntyre, and Mitchell (2019) faced the same issue in their nowcasting exercise and their model works relatively well) and iii) the timing issue i.e. February, May, August and November are the months we chose to observe the quarterly GDP so we get our posterior estimates these months and use the proposed methodology to get the unconditional forecasts.

### 2.2.3.2 Stochastic Volatility

We assume the following priors:

$$\begin{aligned}\theta_i &\sim \mathcal{N}(\theta_{i,0}, V_{\theta_i}) \\ h_t &\sim \text{AR}(1) \\ h_0 &\sim \mathcal{N}(\alpha_h, V_h) \\ \sigma_h^2 &\sim \text{IG}(v_h, S_h).\end{aligned}$$

Our posterior analysis is summarised as follows. We use a five-block Gibbs Sampler to simulate draws from the posterior:  $\pi(\boldsymbol{\theta}_i, \mathbf{h}_i, \mathbf{h}_0, \sigma_h^2 | \mathbf{y})$ , using the following steps:

1.  $\pi(\boldsymbol{\theta}_i | \mathbf{h}_i, \mathbf{h}_0, \sigma_h^2, \mathbf{y})$
2.  $\pi(\mathbf{h}_i | \boldsymbol{\theta}_i, \mathbf{h}_0, \sigma_h^2, \mathbf{y})$
3.  $\pi(\mathbf{h}_0 | \boldsymbol{\theta}_i, \mathbf{h}_i, \sigma_h^2, \mathbf{y})$
4.  $\pi(\sigma_h^2 | \boldsymbol{\theta}_i, \mathbf{h}_0, \mathbf{h}_i, \mathbf{y})$ .

Step 1 is summarised as follows:

$$\theta_i \sim \mathcal{N}(\hat{\theta}_i, K_{\theta_i}^{-1}),$$

where  $K_{\theta_i} = V_{\theta_i}^{-1} + (e^{-h_{i,t}} x_i' x_i)$  and  $\hat{\theta}_i = K_{\theta_i}^{-1}(V_{\theta_i}^{-1} \theta_{i,0} + e^{-h_{i,t}} x_i' y_i)$ .

We describe step 2 in Section 2.2.2. Finally, for steps 3 and 4 their conditional distributions are standard and summarised as follows:

- $(\mathbf{h}_0 | \boldsymbol{\theta}_i, \mathbf{h}_i, \sigma_h^2, \mathbf{y}) \sim \mathcal{N}(\hat{h}_0, K_{h_0}^{-1})$
- $(\sigma_h^2 | \boldsymbol{\theta}_i, \mathbf{h}_0, \mathbf{h}_i, \mathbf{y}) \sim \text{IG}(v_h + T/2, \hat{S}_h)$

where  $K_{h_0} = 1/\sigma_h^2 + 1/V_h$ ,  $\hat{h}_0 = K_{h_0}^{-1}(V_h^{-1} \alpha_h + h_0/\sigma_h^2)$  and  $\hat{S}_h = S_h + [h_0^2 + \sum_{t=2}^T (h_t - h_{t-1})^2]/2$ .

For SV we use the algorithm of [Kim, Shephard, and Chib \(1998\)](#) and the precision sampling methods as described in [Chan and Jeliazkov \(2009\)](#) and [Chan, Hsiao, et al. \(2014\)](#).

### 2.2.4 Stochastic Search Variable Selection (SSVS)

A conventional SSVS prior, for a VAR coefficient,  $\beta_{i,j}$ , is a mixture of two Normal distributions:

$$\theta_i | \gamma_i \sim (1 - \gamma_i) N(0, \underline{c} \times \tau_i^2) + \gamma_i N(0, \tau_i^2), \quad (2.17)$$

where  $\gamma_{i,j} \in \{0, 1\}$  is an unknown parameter,  $\underline{c}$  is a number close to zero,  $\mathcal{P}(\gamma_i = 1) = \pi_i$  and  $\mathcal{P}(\gamma_i = 0) = 1 - \pi_i$ . Thus,  $\gamma_i$  serves as a variable selection indicator. If  $\gamma_i = 0$  then the coefficient is shrunk to be very close to zero, whereas if  $\gamma_i = 1$  it is not.

The prior we use in this Chapter takes a similar form, except that the mixture probability  $\pi_i$ , and the variance  $\tau_i^2$  are assumed to be the parameters to be estimated. We follow [Cross, Hou, and Poon \(2020\)](#) and specify an uninformative prior  $\pi_i \sim \mathcal{U}(0, 1)$ . Finally, we set  $\tau_i^2 \sim \mathcal{IG}(1, 1)$ .

Our posterior analysis is summarised as follows:

1.  $\pi(\boldsymbol{\theta}_i | \boldsymbol{\sigma}_i^2, \gamma_i, \boldsymbol{\pi}_i, \tau_i^2, \mathbf{y})$
2.  $\pi(\boldsymbol{\sigma}_i^2 | \boldsymbol{\theta}_i, \gamma_i, \boldsymbol{\pi}_i, \tau_i^2, \mathbf{y})$
3.  $\pi(\gamma_i | \boldsymbol{\theta}_i, \boldsymbol{\sigma}_i^2, \boldsymbol{\pi}_i, \tau_i^2, \mathbf{y})$
4.  $\pi(\boldsymbol{\pi}_i | \boldsymbol{\theta}_i, \boldsymbol{\sigma}_i^2, \gamma_i, \tau_i^2, \mathbf{y})$
5.  $\pi(\tau_i^2 | \boldsymbol{\theta}_i, \boldsymbol{\sigma}_i^2, \gamma_i, \boldsymbol{\pi}_i, \mathbf{y})$ .

### 2.2.5 Horseshoe

In this Section, we adopt the Global-Local Hierarchical Shrinkage Priors, specifically we adopt the horseshoe prior proposed by [Carvalho, Polson, and Scott \(2010\)](#) as a method for shrinkage in the presence of sparsity. The horseshoe prior belongs to the family of global-local (GL) shrinkage procedure in which  $\lambda_i$  determines the local shrinkage, while the overall one is determined by  $\tau$ . The horseshoe prior has the advantage of aggressively penalising small coefficients (i.e. noise) but applies no shrinkage to large coefficients (i.e. signal). This property offers a non-uniform shrinkage across all coefficients and contrast already existing penalisation methods such as Bayesian LASSO, in which shrinkage is achieved uniformly across all coefficients. Further, another interesting advantage is that it doesn't require any hyperparameters.

A conventional horseshoe prior, for the  $i^{\text{th}}$  VAR equation,  $\theta_i$ , is the following:

$$\theta_i | \Lambda_i^2, \tau_i^2, \sigma_i^2 \sim N(0, \Lambda_i^2 \tau_i^2 \sigma_i^2), \quad \Lambda_i^2 = \text{diag}(\lambda_{i,1}^2, \dots, \lambda_{i,k_i}^2) \quad (2.18)$$

$$\sigma_i^2 \sim \mathcal{IG}(v_{i,0}, S_{i,0}), \quad (2.19)$$

$$\tau_i^2 \sim C^+(0, 1), \quad (2.20)$$

$$\lambda_{i,j}^2 \sim C^+(0, 1), \quad j = 1, \dots, k_i. \quad (2.21)$$

Despite its advantages, this specification results in a non-standard conditional distribution, which makes standard Gibbs sampler intractable. There are various specifications of the horseshoe prior and associated MCMC sampling schemes. [Neal \(2003\)](#) and [Polson, Scott, and Windle \(2014\)](#) proposed the use of a slice sampler that allows  $\tau_i^2$  and  $\lambda_{i,j}^2$  to be updated efficiently in high-dimensions which we follow.

The above hierarchy makes Gibbs sampling from the posterior distribution straightforward.

1. The conditional posterior distribution of the regression coefficients  $\theta_i$  is drawn using the following steps:

$$\theta_i | \Lambda_i^2, \tau_i^2, \sigma_i^2 \sim N(A_i^{-1} X_i' y_i, \sigma_i^2 A_i^{-1}), \quad (2.22)$$

where  $A_i = (X_i' X_i + \Lambda_i^{-1})$ ,  $\Lambda_i = \tau_i^2 \Lambda_i^2$ . We use the efficient algorithm of [Bhattacharya, Chakraborty, and Mallick \(2016\)](#) to sample them.

2. The conditional posterior distribution of  $\sigma_i^2$  is drawn using the following steps:

$$\sigma_i^2 \sim \mathcal{IG}(v_{i,0}^-, S_{i,0}^-), \quad (2.23)$$

where  $v_{i,0}^- = (T + K_i)/2$  and  $S_{i,0}^- = (y_i - X_i \theta_i)' (y_i - X_i \theta_i) / 2 + S_{i,0}$ .

3. The conditional posterior distribution of  $\tau_i^2$  is drawn using the following steps:

- set  $\eta_i = 1/\tau_i^2$
- sample  $u$  from

$$u | \eta_i \sim \mathcal{U}\left(0, \frac{1}{1 + \eta_i}\right), \quad (2.24)$$

- sample  $\eta_i$  from

$$\eta_i \sim \gamma \left( (k_i + T)/2, u \frac{2\sigma_i^2}{\sum (\frac{\theta_i}{\lambda_{i,j}})^2} \right), \quad (2.25)$$

where  $\gamma$  is the lower incomplete gamma function

- set  $\tau_i = \frac{1}{\sqrt{\eta_i}}$ .

4. The conditional posterior distribution of  $\lambda_{i,j}^2$  is drawn using the following steps:

- set  $\eta_{i,j} = 1/\lambda_{i,j}^2$
- sample  $u$  from

$$u|\eta_{i,j} \sim \mathcal{U} \left( 0, \frac{1}{1 + \eta_{i,j}} \right), \quad (2.26)$$

- sample  $\eta_{i,j}$  from

$$\eta_{i,j} \sim e^{\frac{\theta_i^2}{2\sigma_i^2} \eta_{i,j}} I \left( \frac{u}{1-u} > \eta_{i,j} \right) \quad (2.27)$$

- set  $\lambda_{i,j} = \frac{1}{\sqrt{\eta_{i,j}}}$ .

## 2.2.6 Prior Elicitation

In this Subsection, we define the prior hyperparameters as follows. The values we select for the MFBVAR-SV model imply diffuse priors. To allow comparisons between different forecast evaluations, we keep the hyperparameters the same, across all models.

For the independent prior we estimate the shrinkage parameters as in [Giannone, Lenza, and Primiceri \(2015\)](#). We use the following grid for  $\lambda_1$  in the interval [0.05, 0.3] with a step size of 0.01, for  $\lambda_2$  we use [0.1, 3] with a step size of 0.05,  $\lambda_3$  is in [1, 2] with step size 0.2 and finally for  $\lambda_4$  we use a grid over the interval [100, 1000] with step size 100. In this case the derivation of the marginal likelihood is not possible and in order to calculate it we resort to the methodology proposed by [Chib \(1995\)](#), who proposes a simple technique based on Gibbs sampling outputs.

For the SSVS we set  $\theta_{i,0} = 0$ ,  $\gamma_i = 0.5$ ,  $\tau_i^2 = 1$  and  $\underline{c} = 0.0001$ .

For the horseshoe we set  $\sigma_i^2 = 1$ ,  $\lambda_{i,j} = 1$  and  $\tau_i = 1$ .

Finally for the BVAR with stochastic volatility we set  $\alpha_h = 0$ ,  $V_h = 10$ ,  $\sigma_h^2 = 0.1$ ,  $\nu_{h_0} = 5$ ,  $S_h = 0.1 * (\nu_{h_0} - 1)$ , and  $h_0 = \log(\text{var}(y_i) * 0.8)$ .

Our posterior estimates are based on 25,000 draws from the MCMC algorithm and we discard 23,000. We select the lag length to be 1 for the homoskedastic BVAR<sup>1</sup> and 6 for the BVAR with stochastic volatility.

### 2.2.7 Curse of dimensionality

In this Subsection we present common challenges of the proposed estimators. Normally, BVARs using the independent prior are computationally complex. To be more precise, we drop the  $i$  subscript in eq. (2.15) and compute the BVAR using traditional techniques. The dimension of our variance covariance matrix  $K_\theta$  is  $n(np + 1)$  and consequently this involves the following operations: i) invert  $K_\theta$ , ii) calculate the Cholesky factor of  $K_\theta$ , and iii) multiply matrices from i) and ii) i.e the calculation complexity is  $3\mathcal{O}(N^6)$ . However, if we estimate the model equation-by-equation the dimension of  $K_{\theta_i}$  is  $(np + 1)$  which means that we have a complexity of  $\mathcal{O}(N^3)$ <sup>2</sup>. For instance, suppose  $n = 10$  the traditional algorithm needs 3 million calculations for each MCMC draw, while the equation-by-equation algorithm only needs 1,000 per draw.

As noted above, the reparameterization of eq. (2.1) requires writing our model in its structural form, which can lead to identification issues. Since we used the cholesky decomposition as a means to estimate the model and not for identifying the structural shocks. For this reason, the order of our variables does not matter and ultimately the posterior would not be affected.

Lastly, we note that the proposed model has been estimated in an Intel i5-6500 @ 3.2 GHz with 8 GB of RAM. The computation time for 2,000 posterior draws with one lag and 55 variables is about 30 seconds using the equation-by-equation method, contrary to the traditional algorithm, which required about 7,000 seconds.

### 2.2.8 Entropic Tilting with quarterly national-level GDP releases

The proposed entropic tilting method is based on the work of [Robertson, Tallman, and Whiteman \(2005\)](#). The way to incorporate new information in an existing predictive

<sup>1</sup>As in [Koop, McIntyre, and Mitchell \(2019\)](#)

<sup>2</sup>In [Carriero, Clark, and Marcellino \(2019\)](#) the complexity is  $\mathcal{O}(N^4)$ , however we further improve our computational gains by calculating  $K_{\theta_i} = C_{K_{\theta_i}}' C_{K_{\theta_i}}$ ; see [Chan \(2019\)](#) for more details.



density is described in [Altavilla, Giacomini, and Ragusa \(2017\)](#). Furthermore, [Giacomini and Ragusa \(2014\)](#) illustrate that entropic tilting improves (or at least it should not deteriorate) conditional forecasts, when the moment condition is properly specified. We estimate the proposed models and produce current state-level nowcasts of GDP with data available on the second release of nationwide GDP i.e. the predictive density for our nowcasts is  $\pi(Y_{t+h}|Y_{1:t})$ . The entropic tilting methodology is summarised as:

Suppose we have a predictive density with a sample with  $R$  unconditional forecast draws  $y_i$  and some weights,  $w_i$  for each observation which ensure that each observation receives a weight in the sample according to the predictive distribution. Then assume we have more information regarding  $\pi(Y_{t+h}|Y_{1:t})$  that we call  $g(y)$  and impose the following moment condition:

$$\mathbb{E} g(y) = \bar{g}.$$

Then the following minimization problem is the so-called entropic tilting

$$\begin{aligned} \min_{w_i} \quad & KLIC(f^*, f) \\ \text{s.t.} \quad & \mathbb{E} g(y) = \bar{g}. \end{aligned}$$

We find a set of  $w_i$  representing the new predictive density  $\pi^*(Y_{t+h}|Y_{1:t})$ . KLIC denotes the Kullback-Leibler information criterion.

In the empirical application we use a Gaussian approximation, so that we have to find a new forecast density;  $f^*$  which is close to the old normal density and satisfies the restrictions. To be more precise,

$$\pi(Y_{t+h}|Y_{1:t}) \sim \mathcal{N}(\mu, V) \tag{2.28}$$

where  $\mu = (\mu_{US} \quad \mu_S)'$ <sup>3</sup> and  $V = \begin{pmatrix} V_{US} & V_{US,S} \\ V_{S,US} & V_S \end{pmatrix}$ .

Since eq. (2.5) and (2.9) contain lags of the nationwide U.S. GDP (measured at quarterly frequency) and the contemporaneous effects of state-level GDP and U.S. GDP are captured by the variance-covariance matrix, the state-level nowcasts are updated as new quarterly GDP information for each quarter is released. To be more precise, the

<sup>3</sup> $\mu_{US}$  denotes the U.S. mean and  $\mu_S$  is the state-level mean

way that quarterly U.S. GDP releases can be used to shed light on what is happening in the U.S. states. We call this method cross-sectional restriction. To that end, the restriction that state-level GDP adds up to nationwide U.S. GDP should help improve our nowcasts. Since we work with exact growth rates, our cross-sectional restriction holds approximately. Thus, we have:

$$\begin{aligned} y_t^{US} &= \frac{Y_t^{US} - Y_{t-1}^{US}}{Y_{t-1}^{US}} = \sum_{q=1}^4 \frac{Y_{t-1,q}^{US}}{\sum_{q=1}^4 Y_{t-1,q}^{US}} y_{t,q}^{US} \\ &= \sum_{s=1}^{51} \frac{Y_{t-1,q}^{s,A}}{\sum_{s=1}^{51} Y_{t-1,q}^{s,A}} y_t^{s,A} \\ &= \sum_{q=1}^4 w_{q,t-1}^{US} y_{t,q}^{US} = \sum_{s=1}^{51} w_{s,t-1}^{US} y_t^{s,A} \end{aligned}$$

To add the cross-sectional restriction, we extend the aforementioned result by defining a new variable  $z = Ay_{t+1}$ . Matrix  $A$  shows how we use new quarterly information to update our nowcasts. Since  $z$  is a linear transformation of  $y_t$ , it holds:

$$\pi(z|Y_{1:t}) \sim \mathcal{N}(A\mu, AVA').$$

We then tilt the multivariate predictive density eq. (2.28), so that the mean of  $Y_{t+h}^{US}$  is equal to  $\mu_{US}^*$ . As we add more information and move one quarter into the year, new data on  $Y_{t+h,1}$  is released, thus the new mean i.e  $\mu_{US}^*$  reflects information that is available at period  $t+h$ . In general we try to keep the predictive density to be as close to  $\pi(Y_{t+h}|Y_{1:t})$  as possible. It is proven in [Altavilla, Giacomini, and Ragusa \(2017\)](#) that the new predictive density is:

$$\pi^*(Y_{t+h}|Y_{1:t}) \sim \mathcal{N}(\mu^*, V), \quad \mu^* = \begin{pmatrix} \mu_{US}^* \\ \mu_S^* \end{pmatrix} = \begin{pmatrix} \mu_{US}^* \\ \mu_S - V_{US,S} V_{US}^{-1} (\mu_{US} - \mu_{US}^*) \end{pmatrix}.$$

Note that in [Altavilla, Giacomini, and Ragusa \(2017\)](#) tilting does not change the predictive variance. Furthermore, it may appear that entropic tilting does not directly impact on the state-level growth nowcasts as U.S. nationwide variables are being released throughout the year. Since  $V_{US,S} = 0$  and the quarterly U.S. nowcasts are uncorrelated with the state-level nowcasts, the updating of quarterly U.S. GDP nowcasts spill over into the state-level nowcasts<sup>4</sup>.

<sup>4</sup>The cross-state dependences are captured via  $V_S$

So the new predictive density is as follows:

$$\pi^*(z|Y_{1:t}) \sim \mathcal{N}(\bar{\mu}, \bar{V}), \quad \bar{\mu} = \begin{pmatrix} \mu_{US}^* \\ \bar{\mu}_S \end{pmatrix} = \begin{pmatrix} \mu_{US}^* \\ \bar{\mu}_S - V_{US,S}^- V_{US}^{-1} (\mu_{US}^- - \mu_{US}^*) \end{pmatrix},$$

$$\bar{V} = \begin{pmatrix} V_{US}^- & V_{US,S}^- \\ V_{S,US}^- & \bar{V}_S \end{pmatrix}.$$

Note that  $\bar{V}$  is a singular matrix, but this causes no problem for our derivations as they only involve inverting the non-singular  $V_{US}^-$  and we use the again non-singular  $\bar{V}_S$ .

Following the release calendar of BEA, annual state-level GDP data for year  $t$  is not available before the release of the fourth-quarter data for year at  $t + 1$ . To respect this publication lag and to produce genuinely real-time nowcasts, we condition quarter 1 (Q1) up to quarter 3 (Q3) nowcasts on two-step ahead forecast rather than a one-step-ahead forecasts from the VARs, so in the empirical application we have  $h = 2$ . Lastly, suppose we are at  $t + 1$  and the first quarterly GDP is released, we want to use this information to entropically tilt our nowcasts, but quarter 2 (Q2) — quarter 4 (Q4) are not released yet. We therefore assume, while using cross-sectional restriction, that the values of U.S. GDP for quarter 2 — quarter 4 is the value of quarter 1 (Q1). This assumption holds, since we have seasonally adjusted GDP growth rates.

Before closing this Subsection we show the form of matrix  $z$ ,  $A$ ,  $y$  and  $\bar{\mu}$ .

$$z = Ay_{t+1}$$

$$z = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{t-1} & 0_{1 \times S} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0_{1 \times S} & 0_{1 \times S} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0_{1 \times S} & 0_{1 \times S} \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0_{1 \times S} & 0_{1 \times S} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0_{1 \times S} & 0_{1 \times S} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0_{1 \times S} & w_t \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0_{1 \times S} & 0_{1 \times S} \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0_{1 \times S} & 0_{1 \times S} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0_{1 \times S} & 0_{1 \times S} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0_{1 \times S} & 0_{1 \times S} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_S & 0_S \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0_S & I_S \end{pmatrix} \begin{pmatrix} y_{t,1}^{US} \\ y_{t,2}^{US} \\ y_{t,3}^{US} \\ y_{t,4}^{US} \\ y_{t+1,1}^{US} \\ y_{t+1,2}^{US} \\ y_{t+1,3}^{US} \\ y_{t+1,4}^{US} \\ y_{t,1}^S \\ y_{t+1,1}^S \end{pmatrix} \text{ and } \bar{\mu} = \begin{pmatrix} y_t^{US} \\ y_{t,1}^{US} \\ y_{t,2}^{US} \\ y_{t,3}^{US} \\ y_{t,4}^{US} \\ y_{t+1}^{US} \\ y_{t+1,1}^{US} \\ y_{t+1,2}^{US} \\ y_{t+1,3}^{US} \\ y_{t+1,4}^{US} \\ y_{t,1}^S \\ y_{t+1,1}^S \end{pmatrix}.$$

The following summarise how we update our nowcasts using entropic tilting as new national (i.e. the quarterly GDP) data (i.e. Q1 to Q4) is released:

1. After the release of Q1 U.S. GDP growth (in May of each year) set:

$$\mu_1^* = (y_{\tau+1,1}^{US}, y_{\tau+1,1}^{US})$$

2. After Q2 release (in August of each year) set:

$$\mu_1^* = ((w_{1,\tau}^{US} y_{\tau+1,1}^{US} + 3(w_{2,\tau}^{US} y_{\tau+1,2}^{US})), y_{\tau+1,1}^{US}, y_{\tau+1,2}^{US})$$

3. After Q3 release (in November of each year) set:

$$\mu_1^* = ((w_{1,\tau}^{US} y_{\tau+1,1}^{US} + w_{2,\tau}^{US} y_{\tau+1,2}^{US} + 2w_{3,\tau}^{US} y_{\tau+1,3}^{US}), y_{\tau+1,1}^{US}, y_{\tau+1,2}^{US}, y_{\tau+1,3}^{US})$$

4. After Q4 release (in February of each year) set:

$$\mu_1^* = (y_{\tau+1}^{US}, y_{\tau+1,1}^{US}, y_{\tau+1,2}^{US}, y_{\tau+1,3}^{US}, y_{\tau+1,4}^{US})$$

## 2.3 Empirical application

### 2.3.1 Data

The dataset consists of the following U.S. nominal gross domestic product (GDP) at quarterly frequency sampled from 1947Q1 to 2018Q4. U.S. Annual GDP by state for 50 states plus the District of Columbia is available at the U.S. Bureau of Economic Analysis (BEA)<sup>5</sup>.

The Release Schedule for U.S. GDP can be found on <https://www.bea.gov/news/schedule> and is summarized below:

- U.S. Quarterly GDP has 3 estimates for each quarter. Flash GDP is released one month after the end of each quarter while the last one comes 3 months after the end of each quarter
- U.S. Annual GDP by state is released twice per year. The preliminary release is 5 months after the end of each year, while the last one is released after 11 months.

Our sample spans from 1963 to 2018. We use the full sample 2018 vintage release, which ensures U.S. Quarterly GDP includes its second release for the 4th quarter of 2018.

<sup>5</sup><https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1&acrdn=1>

To delve deep in the 'Big data' world and to perform a robustness check we enrich our dataset with monthly Employment (series ID: CES0000000001) measured at national level available at the Bureau of Labor Statistics(BLS)<sup>6</sup>. All series are seasonally adjusted and their transformation is delineated in the following Subsection.

### 2.3.2 Definitions

In this Subsection, we delineate the details of the econometric model that we use to produce the new U.S. state-level estimates and explain its properties. We use the following notational conventions:

- $t = 1, \dots, T$  runs at the annual frequency.
- $Y_t^{s,A}$  is annual GDP for state  $s$ , for  $s = 1, \dots, 51$ .
- $y_t^{s,A} = \frac{Y_t^{s,A} - Y_{t-1}^{s,A}}{Y_{t-1}^{s,A}}$  is annual GDP growth in state  $s$ .
- $Y_{t,i}^{US}$  is US GDP in the  $i^{th}$  quarter of the year for  $i = 1, \dots, 4$ .
- $y_{t,i}^{US} = \frac{Y_{t,i}^{US} - Y_{t-1,i}^{US}}{Y_{t-1,i}^{US}}$  is annual GDP growth relative to the same quarter in the previous year.
- $y_t = \left( y_{t,1}^{US} \ y_{t,2}^{US} \ y_{t,3}^{US} \ y_{t,4}^{US} \ y_t^A \right)'$ , where  $y_t^A = (y_t^{1,A}, \dots, y_t^{51,A})'$ . All the annual variables are stacked into vectors.
- $emp_{t,i}^{US} = \frac{EMP_{t,i}^{US} - EMP_{t-1,i}^{US}}{EMP_{t-1,i}^{US}}$  is annual GDP growth relative to the same month in the previous year.
- $y_t = \left( emp_{t,m}^{US} \ y_{t,1}^{US} \ y_{t,2}^{US} \ y_{t,3}^{US} \ y_{t,4}^{US} \ y_t^A \right)'$ , where  $m = 1, \dots, 12$
- $Y_{t_0:t_1}$  denotes a sequence of observations, that is abbreviated as  $Y_{1:T}$ .

Note that the dimension of the 'small' model is 55, while the larger model (i.e. the model with monthly employment) has 67 variables.

<sup>6</sup><https://data.bls.gov/timeseries/CES0000000001>

### 2.3.3 Forecast Evaluation

We perform a two-step-ahead recursive out-of-sample forecast exercise for 13 years. Our estimation sample starts from 1964 to 2006 and ends to 1964 to 2018, thus we estimate a total of 13 sets of two-steps-ahead forecasts. The reason why we use a two year-ahead forecast is to address the two year publication lag of the annual GDP by state.

We consider 6 models. The first three are a constant volatility MF-VAR with annual state-level GDP and quarterly GDP and three different prior families, namely the independent prior, SSVS and HS. The last three have the same sets of variables, but have stochastic volatility and the aforementioned prior families. Table 2.1 summarizes the models we estimate.

In addition, we calculate the unconditional forecasts and use entropic tilting methods as in [Altavilla, Giacomini, and Ragusa \(2017\)](#) to add new information. As already explained, the relative entropy method does not require new MCMC draws as it simply re-weights an existing forecast distribution.

Root mean squared forecast error (RMSFE) is used to measure the accuracy of point estimates, while log predictive score (logS) and continuous ranked probability score (CRPS) evaluates the accuracy of the predictive density.

RMSFE is defined as follows:

$$RMSFE_{i,h} = \sqrt{\frac{\sum_{t=1}^P (y_{i,t+h} - \hat{y}_{i,t+h})^2}{P}}$$

logS is defined as follows:

$$\log S_{i,h} = \frac{\sum_{t=1}^P \log \pi(y_{i,t+h} = \hat{y}_{i,t+h} | Y_{1:T})}{P}$$

CRPS is defined as follows:

$$CRPS_{i,h} = \int_{-\infty}^{\infty} [F_{i,t,h}(y) - \mathbb{1}(y_{i,t+h} \leq y)]^2 dy$$

Since the predictive density is Gaussian with mean  $\mu$  and variance  $\sigma^2$ , [Gneiting and Raftery \(2007\)](#) show that CRPS can be approximated as follows:

$$\begin{aligned} CRPS_{i,h} &= \sigma \left[ \frac{1}{\sqrt{\pi}} - 2\phi\left(\frac{x-\mu}{\sigma}\right) - \frac{x-\mu}{\sigma} \left(2\Phi\left(\frac{x-\mu}{\sigma}\right) - 1\right) \right] \\ &= var(y_{i,t+h})^{0.5} \left[ \frac{1}{\sqrt{\pi}} - 2\phi\left(\frac{y_{i,t+h} - y_{i,t+h}}{var(y_{i,t+h})^{0.5}}\right) - \frac{y_{i,t+h} - y_{i,t+h}}{var(y_{i,t+h})^{0.5}} \left(2\Phi\left(\frac{y_{i,t+h} - y_{i,t+h}}{var(y_{i,t+h})^{0.5}}\right) - 1\right) \right] \end{aligned}$$

where  $y_{i,t+h}$  the realised value,  $y_{i,t+h}$  the unconditional or the entropically tilted forecast,  $i$  the variable,  $h$  the forecast horizon,  $F_{i,t,h}(y)$  the cumulative distribution function,  $\phi$  the probability density function and  $\Phi$  the cumulative density function.

For RMSFE and CRPS the smaller the value the better the forecast, while for logS the larger value the better. In general, our entropic tilting methodology follows the one performed in [Koop, McIntyre, and Mitchell \(2019\)](#).

A priori, we expect the inclusion of stochastic volatility to improve density forecasts see [Clark \(2011\)](#), while the use of a large dataset should improve point forecasts.

Tables 2.2 – 2.9 display the root mean squared forecast error (RMSFE) and the continuous ranked probability score (CRPS) relative (ratio) to the benchmark AR(1) so that a value below 1 denotes a model outperforming the benchmark. The log predictive score (logS) relative (difference) to the benchmark AR(1) values are depicted on Tables 2.2 – 2.9 so that a value above 0 denotes a model outperforming the benchmark. The large heteroskedastic model outperforms the large homoskedastic model for more than half states at almost all horizons, suggesting that the inclusion of SV improves the specification of the conditional means and consequently the point forecasts.

[Koop, McIntyre, and Mitchell \(2019\)](#) find little evidence on adding stochastic volatility, but in their application they use UK data, while in our application we use U.S. data.

In Appendix A.1, the x-axes in Figures 2.1 – 2.3 measure the RMSFE obtained by the large homoskedastic BVAR, while the y-axes measures the SV counterpart. Each point corresponds to a different forecast horizon, and when a point is above the 45 degree line it indicates that the inclusion of SV improves point forecasts.

TABLE 2.1: Model Acronyms

Name	Description
1 BVARIP	Mixed Frequency Bayesian VAR with independent prior (IP)
2 BVARIP-SV	Mixed Frequency Bayesian VAR with IP and Stochastic Volatility (SV)
3 BVARSSVS	Mixed Frequency Bayesian VAR with SSVS prior
4 BVARSSVS-SV	Mixed Frequency Bayesian VAR with SSVS prior and SV
5 BVARHS	Mixed Frequency Bayesian VAR with horseshoe(HS) prior
6 BVARHS-SV	Mixed Frequency Bayesian VAR with HS prior and SV
7 BVARIP-M	Mixed Frequency Bayesian VAR with IP prior and monthly employment
8 BVARIP-SV-M	Mixed Frequency Bayesian VAR with IP prior, SV and monthly employment
9 Bench	Autoregressive Model of order 1

### 2.3.4 Robustness check

To check the validity of our model we run the following robustness check, i.e. we add monthly U.S. employment in our BVAR specification, and run the same forecast evaluation. We estimate the following models i) a homoskedastic MF-VAR that also includes monthly national-level employment and ii) its stochastic volatility counterpart, in both models we use the independent prior.

We update the state-level GDP with monthly employment every quarter as follows:

1. After the release of Q1 U.S. GDP growth (in May of each year) set:

$$\mu_1^* = (y_{\tau+1,1}^{US}, emp_{\tau+1,1}^{US,m1-m3}, y_{\tau+1,1}^{US})$$

2. After Q2 release (in August of each year) set:

$$\mu_1^* = ((w_{1,\tau}^{US} y_{\tau+1,1}^{US} + 3(w_{2,\tau}^{US} y_{\tau+1,2}^{US}), y_{\tau+1,1}^{US}, emp_{\tau+1,1}^{US,m1-m3}, y_{\tau+1,2}^{US}, emp_{\tau+1,1}^{US,m4-m6})$$

3. After Q3 release (in November of each year) set:

$$\mu_1^* = ((w_{1,\tau}^{US} y_{\tau+1,1}^{US} + w_{2,\tau}^{US} y_{\tau+1,2}^{US} + 2w_{3,\tau}^{US} y_{\tau+1,3}^{US}), y_{\tau+1,1}^{US}, emp_{\tau+1,1}^{US,m1-m3}, y_{\tau+1,2}^{US}, emp_{\tau+1,1}^{US,m4-m6}, y_{\tau+1,3}^{US}, emp_{\tau+1,1}^{US,m7-m9})$$

4. After Q4 release (in February of each year) set:

$$\mu_1^* = (y_{\tau+1,1}^{US}, y_{\tau+1,1,1}^{US}, emp_{\tau+1,1}^{US,m1-m3}, y_{\tau+1,2}^{US}, emp_{\tau+1,1}^{US,m4-m6}, y_{\tau+1,3}^{US}, emp_{\tau+1,1}^{US,m7-m9}, y_{\tau+1,4}^{US}, emp_{\tau+1,1}^{US,m10-m12})$$



### 2.3.5 Empirical Results

In this Section we present the results of our nowcasting exercise from 1963 to 2018. We start the evaluation period in 2006 and apply an iterated forecasting method that re-estimates the parameters of our model. Specifically, we do a real-time out-of-the-sample (OOS) nowcasting exercise from 2006 up 2018.

In order to evaluate the forecasting performance of the different methods, presented in Table 2.1, we use two-year-ahead-forecast from each method and calculate three different evaluation criteria, RMSFE, logS and CRPS where the benchmark model is an AR(1) with normal errors.

In terms of point estimate forecasts we find that the proposed MF-VAR model outperforms the benchmark in almost all states and at different horizons. There is a small number of states for which the benchmark outperforms other models based on RMSFE. These include (where the second placed model is noted in brackets after the state name): Louisiana (BVARIP-M), North Carolina (BVARHS), Utah (BVARHS, BVARSSVS), California (BVARSSVS), Nebraska (BVARIP-M), Alabama (BVARIP-SV-M), Alaska (BVARIP-SV-M), Florida (BVARHS), Nevada (BVARHS), New Jersey (BVARSSVS-SV) and New York (BVARIP).

Regarding the evaluation of density forecasts, we use the logS criterion, which illustrates that the competing models in Table 2.1 outperform the AR(1) benchmark in almost all states and at almost all horizon. BVARSSVS-SV and BVARHS-SV models underperform relative to the benchmark in almost all 43 states. Finally, based on CRPS we find that the AR(1) benchmark model outperforms the following models: BVARIP-M, BVARIP-SV-M, BVARSSVS-SV, BVARHS, BVARHS-SV and BVARIP-SV in Alaska, Hawaii, Illinois, Kansas, Missouri, Nebraska, New Jersey, New York, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Utah, Vermont, Virginia, Washington, West Virginia and Wisconsin.

As we entropically tilt the new information from Q1, the following models: BVARIP-M, BVARIP-SV-M, BVARSSVS-SV, BVARHS, BVARHS-SV and BVARIP-SV underperform relative to the benchmark in states: Alabama, Arkansas, Florida, Hawaii, Illinois, Kansas, Kentucky, Maine, Maryland, Massachusetts, Minnesota, Mississippi, Missouri, Nebraska, New Jersey, New York, North Carolina, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia and Wisconsin.

While tilting information in Q2, BVARIP-SV-M, BVARIP-M, BVARSSVS-SV, BVARHS, BVARHS-SV, BVARIP-SV perform worse than the AR(1) in the following states Alabama, Alaska, Arkansas, Florida, Hawaii, Illinois, Kansas, Maine, Maryland, Massachusetts, Minnesota, Mississippi, Missouri, New Hampshire, New Jersey, New York, North Carolina, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia and Wisconsin. The same pattern is observed as we tilt new information in Q3.

Finally, as we use the information from Q4, the benchmark outperforms the following models: BVARHS, BVARIP, BVARSSVS, BVARSSVS-SV, BVARHS-SV and BVARIP-SV in Alabama, Arizona, Arkansas, California, Florida, Georgia, Hawaii, Illinois, Indiana, Iowa, Kansas, Kentucky, Maryland, Mississippi, Montana, Nevada, New Jersey, New York, North Carolina, Oregon, Pennsylvania, Rhode Island, Tennessee, Utah, Virginia, Washington, West Virginia and Wisconsin.

In Appendix [A.1](#) we report the best performing model for each criterion in **bold**.

To summarize, in the majority of the cases, the best model according the RMSFE is the BVARHS-SV, while based on the forecast density evaluation metrics the best model is the BVARIP. It is clear that the RMSFE favours models with stochastic volatility. Further, the last row in Tables [2.2](#) – [2.9](#) shows a comparison between the BVARIP and BVARIP-SV, in this case models with stochastic volatility prevail 27 times over homoskedastic ones at all horizons, while 6 times models with homoskedasticity wins over ones with stochastic volatility at all horizons. Finally, there are 18 times that either models with homoskedasticity win at certain horizons or models with stochastic volatility win at certain horizons.

## 2.4 Conclusion

In this Chapter, we start by using both a homoskedastic and stochastic volatility MF-VAR with three different priors. We shift the real out-of-sample annual regional forecast towards the quarterly U.S.-wide GDP and produce quarterly GDP by state. We show out-of-sample forecasting gains and provide evidence towards a MF-VAR with stochastic volatility.

Producing real-time U.S. state-level GDP in a more frequent and timely fashion is of high importance for policymakers since it is a key metric to measure economic activity.

In that spirit, the initiative of Federal Reserve Bank of Atlanta to produce GDPnow and New York Federal Reserve Bank to publish a nowcast report and BEA to measure quarterly GDP by state show the need to construct machinery to produce U.S.-wide and state-level GDP nowcasts, respectively. We hope the output of this chapter is found useful by economists and regional economic policymakers alike.

## **A.1 Appendix (Chapter 2)**

	Tilting using new information														
	Point Forecasting - MRSFE					Density Forecasting - logS					Density Forecasting - CRPS				
	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4
Model vs bench	<b>Alabama</b>														
BVARIP	0.135	0.352	0.336	0.323	0.362	1.297	<b>1.016</b>	<b>1.026</b>	<b>1.032</b>	0.746	0.217	<b>0.362</b>	<b>0.351</b>	<b>0.349</b>	<b>0.408</b>
BVARIP-SV	0.176	<b>0.197</b>	<b>0.188</b>	<b>0.175</b>	0.400	0.650	0.449	0.436	0.439	0.445	0.399	0.534	0.547	0.546	0.508
BVARSSVS	0.334	0.394	0.420	0.412	0.604	0.865	0.702	0.689	0.691	-0.09	0.379	0.465	0.477	0.477	0.841
BVARSSVS-SV	0.210	0.298	0.311	0.296	0.467	0.152	-0.055	-0.071	-0.067	0.041	0.643	0.881	0.907	0.906	0.715
BVARHS	0.331	0.510	0.537	0.518	0.873	0.470	0.259	0.246	0.250	-0.370	0.501	0.694	0.711	0.708	1.140
BVARHS-SV	0.147	0.228	0.235	0.228	<b>0.350</b>	0.118	-0.091	-0.107	-0.104	-0.115	0.656	0.903	0.931	0.931	0.789
BVARIP-M	0.861	1.194	0.995	0.762	0.690	0.167	0.001	0.099	0.293	9.323	0.838	1.048	0.996	0.752	0.581
BVARIP-SV-M	0.792	1.393	1.151	0.937	0.643	-0.008	-0.728	-0.673	-0.547	<b>11.420</b>	0.883	1.664	1.631	1.320	0.460
BVARIP vs BVARIP-SV	0.767	1.786	1.789	1.838	0.904	0.647	0.567	0.590	0.593	0.301	0.545	0.678	0.642	0.638	0.802
Model vs bench	<b>Alaska</b>														
BVARIP	0.221	0.227	0.220	0.214	0.749	1.211	1.215	1.219	1.223	0.284	0.278	0.273	0.272	0.269	0.776
BVARIP-SV	0.316	0.262	0.267	0.273	0.259	0.536	0.556	0.556	0.557	0.832	0.525	0.493	0.497	0.495	<b>0.342</b>
BVARSSVS	0.269	0.223	0.222	0.222	0.371	1.428	1.459	1.460	1.461	0.600	0.249	<b>0.228</b>	<b>0.229</b>	<b>0.228</b>	0.449
BVARSSVS-SV	0.196	0.166	0.164	0.163	0.247	1.103	1.123	1.124	1.126	0.795	0.301	0.282	0.283	0.282	0.348
BVARHS	0.427	0.341	0.341	0.344	0.572	0.692	0.719	0.720	0.720	-0.140	0.480	0.441	0.443	0.442	0.868
BVARHS-SV	0.156	<b>0.138</b>	<b>0.139</b>	<b>0.139</b>	<b>0.218</b>	1.191	1.209	1.209	1.211	0.779	0.272	0.256	0.258	0.257	0.343
BVARIP-M	0.992	0.956	0.988	1.051	0.974	2.425	1.554	0.554	-0.323	3.097	1.053	0.974	1.003	1.047	0.962
BVARIP-SV-M	0.971	0.985	1.026	1.092	0.982	8.822	<b>7.755</b>	<b>6.572</b>	<b>5.707</b>	<b>4.862</b>	0.848	0.835	0.867	0.907	0.838
BVARIP vs BVARIP-SV	0.699	0.865	0.824	0.783	2.896	0.675	0.659	0.663	0.666	-0.548	0.529	0.554	0.548	0.543	2.267
Model vs bench	<b>Arizona</b>														
BVARIP	0.220	0.511	0.489	0.470	0.794	1.359	<b>0.896</b>	<b>0.903</b>	<b>0.911</b>	0.132	0.250	0.449	0.439	0.436	0.819
BVARIP-SV	0.156	0.375	0.370	0.352	0.687	0.759	0.367	0.374	0.379	<b>0.357</b>	0.379	0.599	0.588	0.586	<b>0.671</b>
BVARSSVS	0.179	<b>0.231</b>	<b>0.210</b>	<b>0.191</b>	<b>0.296</b>	1.028	0.668	0.679	0.685	-0.027	0.302	<b>0.434</b>	<b>0.422</b>	<b>0.419</b>	0.697
BVARSSVS-SV	0.197	0.379	0.365	0.350	0.611	0.461	0.080	0.089	0.094	0.167	0.508	0.778	0.760	0.759	0.688
BVARHS	0.322	0.618	0.573	0.562	0.838	0.477	0.094	0.106	0.109	-0.546	0.526	0.811	0.787	0.789	1.254
BVARHS-SV	0.224	0.426	0.399	0.387	0.652	0.558	0.177	0.187	0.191	0.142	0.467	0.717	0.698	0.698	0.717
BVARIP-M	0.414	0.636	0.616	0.594	0.888	0.910	0.529	0.552	0.597	-1.014	0.437	0.549	0.543	0.520	0.854
BVARIP-SV-M	0.477	0.514	0.487	0.471	0.905	0.285	0.014	0.028	0.055	0.101	0.651	0.744	0.738	0.708	0.868
BVARIP vs BVARIP-SV	1.409	1.363	1.321	1.337	1.157	0.600	0.529	0.529	0.531	-0.225	0.661	0.750	0.746	0.744	1.222
Model vs bench	<b>Arkansas</b>														
BVARIP	0.277	0.641	0.644	0.608	<b>0.419</b>	1.169	<b>0.948</b>	<b>0.944</b>	<b>0.945</b>	<b>0.702</b>	0.291	<b>0.417</b>	<b>0.424</b>	<b>0.425</b>	<b>0.450</b>
BVARIP-SV	0.294	0.724	0.678	0.625	0.579	0.597	0.346	0.352	0.356	0.389	0.449	0.693	0.689	0.687	0.637
BVARSSVS	0.377	0.862	0.82	0.764	0.65	0.838	0.615	0.621	0.625	-0.056	0.402	0.581	0.578	0.576	0.866
BVARSSVS-SV	0.378	0.979	0.939	0.879	0.753	0.198	-0.059	-0.054	-0.051	0.028	0.654	1.025	1.024	1.024	0.868
BVARHS	0.376	0.888	0.835	0.785	0.638	0.412	0.164	0.170	0.172	-0.379	0.543	0.830	0.826	0.826	1.108
BVARHS-SV	0.254	0.697	0.666	0.626	0.565	0.231	-0.029	-0.025	-0.022	0.050	0.611	0.966	0.965	0.965	0.771
BVARIP-M	0.258	0.445	0.508	0.517	0.558	1.210	0.763	0.593	0.565	-0.05	0.270	0.439	0.510	0.527	0.632
BVARIP-SV-M	0.324	<b>0.277</b>	<b>0.300</b>	<b>0.306</b>	0.590	0.248	0.341	0.229	0.204	0.362	0.581	0.486	0.542	0.559	0.652
BVARIP vs BVARIP-SV	0.941	0.886	0.950	0.973	0.723	0.572	0.602	0.592	0.589	0.313	0.648	0.603	0.616	0.618	0.706
Model vs bench	<b>California</b>														
BVARIP	0.113	0.564	0.618	0.676	0.852	1.468	0.884	0.847	0.811	-0.213	0.183	0.485	0.521	0.545	1.125
BVARIP-SV	0.211	0.282	0.284	0.276	<b>0.235</b>	0.641	0.328	0.313	0.307	0.891	0.407	0.610	0.626	0.623	0.407
BVARSSVS	0.367	0.874	0.956	1.021	1.045	0.765	0.349	0.322	0.307	0.037	0.424	0.790	0.835	0.849	1.212
BVARSSVS-SV	0.151	<b>0.248</b>	<b>0.247</b>	<b>0.236</b>	0.268	0.214	-0.111	-0.127	-0.133	0.453	0.594	0.925	0.952	0.950	0.604
BVARHS	0.277	0.555	0.672	0.717	0.803	0.466	0.133	0.105	0.096	-0.053	0.489	0.778	0.829	0.835	1.119
BVARHS-SV	0.088	0.453	0.450	0.415	0.334	0.266	-0.074	-0.089	-0.093	0.409	0.557	0.913	0.936	0.926	0.645
BVARIP-M	0.407	0.314	0.394	0.462	0.330	0.919	<b>0.996</b>	<b>0.923</b>	<b>0.826</b>	<b>1.177</b>	0.428	<b>0.335</b>	<b>0.382</b>	<b>0.433</b>	<b>0.330</b>
BVARIP-SV-M	0.397	0.427	0.365	0.405	0.375	0.352	0.020	0.081	0.029	0.845	0.574	0.749	0.686	0.741	0.425
BVARIP vs BVARIP-SV	0.533	2.002	2.179	2.445	3.630	0.827	0.555	0.534	0.504	-1.104	0.449	0.795	0.833	0.875	2.763
Model vs bench	<b>Colorado</b>														
BVARIP	0.214	0.235	0.235	0.233	0.268	1.465	<b>1.359</b>	<b>1.357</b>	<b>1.362</b>	<b>1.852</b>	0.224	<b>0.249</b>	<b>0.249</b>	<b>0.247</b>	<b>0.264</b>
BVARIP-SV	0.213	0.247	0.251	0.25	0.218	0.693	0.545	0.535	0.538	1.581	0.415	0.511	0.521	0.520	0.293
BVARSSVS	0.422	0.459	0.481	0.485	0.418	0.867	0.774	0.767	0.770	0.967	0.422	0.458	0.465	0.463	0.549
BVARSSVS-SV	0.172	0.183	0.188	0.181	0.202	0.311	0.159	0.148	0.151	1.185	0.589	0.736	0.753	0.752	0.406
BVARHS	0.354	0.362	0.369	0.375	0.389	0.510	0.376	0.367	0.369	0.667	0.517	0.613	0.624	0.624	0.692
BVARHS-SV	0.107	<b>0.156</b>	<b>0.158</b>	<b>0.163</b>	<b>0.173</b>	0.477	0.322	0.311	0.313	1.160	0.495	0.624	0.638	0.638	0.410
BVARIP-M	0.471	0.610	0.531	0.457	0.573	0.913	0.713	0.850	0.969	0.395	0.476	0.568	0.484	0.430	0.601
BVARIP-SV-M	0.403	0.374	0.302	0.268	0.597	0.438	0.377	0.447	0.520	0.476	0.570	0.592	0.527	0.484	0.625
BVARIP vs BVARIP-SV	1.004	0.952	0.935	0.933	1.232	0.773	0.814	0.821	0.823	0.271	0.540	0.487	0.478	0.475	0.900
Model vs bench	<b>Connecticut</b>														
BVARIP	0.289	0.325	0.324	0.323	<b>0.213</b>	1.243	1.18	1.182	1.186	<b>2.772</b>	0.278	0.304	0.302	0.298	<b>0.183</b>
BVARIP-SV	0.281	0.311	0.319	0.319	0.246	0.516	0.460	0.448	0.448	2.360	0.484	0.522	0.533	0.531	0.239
BVARSSVS	0.349	0.272	0.276	0.276	0.226	0.770	0.747	0.736	0.736	1.898	0.407	0.398	0.405	0.404	0.334
BVARSSVS-SV	0.341	0.349	0.351	0.35	0.292	0.048	-0.006	-0.019	-0.019	1.971	0.756	0.808	0.826	0.825	0.331
BVARHS	0.538	0.574	0.586	0.605	0.386	0.233	0.182	0.171	0.169	1.432	0.678	0.722	0.734	0.737	0.538
BVARHS-SV	0.285	0.340	0.358	0.362	0.281	0.343	0.285	0.271	0.270	1.993	0.565	0.613	0.629	0.629	0.322
BVARIP-M	0.312	0.304	0.224	0.199	0.481	1.237	<b>1.202</b>	<b>1.386</b>	<b>1.475</b>	0.743	0.304	<b>0.297</b>	<b>0.222</b>	<b>0.201</b>	0.466
BVARIP-SV-M	0.262	<b>0.247</b>	<b>0.195</b>	<b>0.185</b>	0.424	0.440	0.366	0.542	0.624	0.888	0.517	0.538	0.409	0.373	0.439
BVARIP vs BVARIP-SV	1.027	1.043	1.016	1.014	0.868	0.727	0.720	0.734	0.738	0.412	0.574	0.582	0.567	0.562	0.763

TABLE 2.2: Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model.

Model vs bench	Tilting using new information														
	Point Forecasting - MRSFE					Density Forecasting - logS					Density Forecasting - CRPS				
	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4
	<b>Delaware</b>														
BVARIP	0.368	<b>0.125</b>	<b>0.122</b>	<b>0.122</b>	0.351	1.062	1.419	1.445	1.446	1.265	0.349	<b>0.193</b>	<b>0.188</b>	<b>0.189</b>	<b>0.351</b>
BVARIP-SV	0.487	0.192	0.182	0.185	0.333	0.350	0.693	0.721	0.722	1.087	0.665	0.382	0.371	0.372	0.374
BVARSSVS	0.506	0.175	0.165	0.163	<b>0.326</b>	0.865	1.222	1.252	1.255	0.824	0.438	0.243	0.234	0.233	0.446
BVARSSVS-SV	0.491	0.190	0.182	0.182	0.383	0.067	0.412	0.439	0.441	0.786	0.867	0.496	0.482	0.482	0.476
BVARHS	0.618	0.205	0.192	0.192	0.416	0.330	0.683	0.712	0.714	0.385	0.696	0.388	0.375	0.376	0.663
BVARHS-SV	0.554	0.217	0.209	0.209	0.411	0.254	0.597	0.625	0.627	0.838	0.735	0.422	0.410	0.410	0.470
BVARIP-M	0.418	0.212	0.254	0.228	0.565	1.171	<b>1.791</b>	<b>1.520</b>	<b>1.696</b>	0.579	0.341	0.203	0.256	0.224	0.588
BVARIP-SV-M	0.373	0.209	0.256	0.235	0.379	0.288	1.214	0.977	1.140	1.060	0.684	0.273	0.335	0.297	0.379
BVARIP vs BVARIP-SV	0.756	0.654	0.671	0.660	1.053	0.712	0.725	0.724	0.724	0.178	0.524	0.506	0.508	0.508	0.939
	<b>District of Columbia</b>														
BVARIP	0.125	0.172	0.175	0.172	0.194	1.692	<b>1.479</b>	<b>1.476</b>	<b>1.490</b>	<b>2.431</b>	0.165	<b>0.207</b>	<b>0.203</b>	<b>0.200</b>	<b>0.201</b>
BVARIP-SV	0.083	0.113	0.117	0.116	0.186	0.799	0.604	0.604	0.619	1.902	0.355	0.424	0.415	0.406	0.291
BVARSSVS	0.355	0.418	0.421	0.41	0.565	0.916	0.733	0.731	0.747	1.292	0.398	0.462	0.453	0.443	0.612
BVARSSVS-SV	0.069	0.116	0.118	0.113	<b>0.156</b>	0.193	-0.002	-0.002	0.014	1.379	0.645	0.771	0.753	0.736	0.461
BVARHS	0.183	0.197	0.191	0.185	0.374	0.704	0.515	0.516	0.532	1.264	0.403	0.472	0.460	0.449	0.554
BVARHS-SV	0.077	<b>0.111</b>	<b>0.115</b>	<b>0.111</b>	0.159	0.341	0.147	0.147	0.163	1.508	0.555	0.662	0.647	0.633	0.408
BVARIP-M	0.305	0.331	0.347	0.347	0.764	1.156	1.181	1.123	1.127	-0.140	0.305	0.333	0.345	0.346	0.779
BVARIP-SV-M	0.275	0.261	0.292	0.290	0.646	0.282	0.525	0.415	0.422	0.378	0.585	0.463	0.511	0.511	0.645
BVARIP vs BVARIP-SV	1.503	1.528	1.492	1.477	1.042	0.893	0.875	0.872	0.871	0.529	0.466	0.487	0.490	0.492	0.692
	<b>Florida</b>														
BVARIP	0.140	0.237	0.223	0.230	<b>0.400</b>	1.509	1.152	1.169	1.162	<b>0.816</b>	0.182	<b>0.273</b>	<b>0.260</b>	<b>0.262</b>	<b>0.404</b>
BVARIP-SV	0.274	0.345	0.336	0.335	0.619	0.599	0.270	0.281	0.278	0.103	0.431	0.594	0.575	0.573	0.733
BVARSSVS	0.340	0.455	0.432	0.444	0.879	0.596	0.257	0.272	0.266	-0.668	0.449	0.629	0.605	0.607	1.422
BVARSSVS-SV	0.235	0.289	0.281	0.278	0.620	0.022	-0.313	-0.303	-0.307	-0.425	0.717	1.005	0.976	0.975	1.097
BVARHS	0.590	0.925	0.893	0.927	2.165	0.109	-0.247	-0.234	-0.243	-1.115	0.742	1.099	1.059	1.069	2.503
BVARHS-SV	0.078	<b>0.161</b>	<b>0.168</b>	<b>0.163</b>	0.422	0.163	-0.175	-0.165	-0.170	-0.311	0.609	0.865	0.841	0.840	0.950
BVARIP-M	0.425	0.266	0.277	0.248	0.843	1.029	<b>1.212</b>	<b>1.180</b>	<b>1.212</b>	-0.012	0.410	0.285	0.291	0.277	0.895
BVARIP-SV-M	0.430	0.447	0.486	0.413	0.844	0.432	0.276	0.232	0.269	0.104	0.565	0.664	0.698	0.656	0.935
BVARIP vs BVARIP-SV	0.510	0.688	0.664	0.685	0.646	0.910	0.882	0.888	0.885	0.714	0.422	0.459	0.453	0.457	0.551
	<b>Georgia</b>														
BVARIP	0.238	0.125	0.146	0.135	0.338	1.338	<b>1.306</b>	<b>1.301</b>	<b>1.310</b>	<b>1.041</b>	0.251	<b>0.225</b>	<b>0.226</b>	<b>0.223</b>	<b>0.337</b>
BVARIP-SV	0.288	0.331	0.332	0.316	0.487	0.536	0.370	0.375	0.381	0.428	0.464	0.580	0.564	0.561	0.560
BVARSSVS	0.375	0.253	0.241	0.225	0.399	0.621	0.509	0.516	0.523	-0.197	0.462	0.497	0.48	0.477	0.882
BVARSSVS-SV	0.240	0.273	0.265	0.250	0.397	-0.016	-0.190	-0.184	-0.179	-0.059	0.754	0.963	0.935	0.934	0.776
BVARHS	0.315	0.266	0.241	0.237	0.390	0.316	0.157	0.164	0.169	-0.415	0.565	0.688	0.664	0.664	1.076
BVARHS-SV	0.088	<b>0.115</b>	<b>0.098</b>	<b>0.085</b>	<b>0.259</b>	0.143	-0.033	-0.027	-0.022	0.059	0.631	0.812	0.788	0.787	0.673
BVARIP-M	0.171	0.333	0.323	0.332	0.440	1.503	1.145	1.168	1.153	0.849	0.198	0.334	0.318	0.327	0.426
BVARIP-SV-M	0.220	0.351	0.355	0.375	0.354	0.418	0.222	0.228	0.212	0.438	0.495	0.630	0.615	0.634	0.475
BVARIP vs BVARIP-SV	0.828	0.379	0.441	0.429	0.693	0.802	0.937	0.927	0.929	0.614	0.541	0.389	0.401	0.398	0.602
	<b>Hawaii</b>														
BVARIP	0.274	0.291	0.262	0.26	0.255	0.978	1.007	1.021	1.020	0.908	0.354	0.337	0.331	0.328	0.352
BVARIP-SV	0.282	0.244	0.232	0.246	0.316	0.166	0.225	0.231	0.227	0.353	0.759	0.664	0.664	0.665	0.568
BVARSSVS	0.734	0.43	0.413	0.419	0.584	0.357	0.441	0.448	0.445	-0.254	0.707	0.573	0.574	0.572	1.043
BVARSSVS-SV	0.319	0.252	0.233	0.234	0.350	-0.440	-0.380	-0.373	-0.377	-0.231	1.381	1.198	1.198	1.197	0.964
BVARHS	0.557	0.400	0.373	0.391	0.730	-0.081	-0.019	-0.012	-0.016	-0.614	0.988	0.855	0.853	0.855	1.461
BVARHS-SV	0.294	0.206	0.200	0.205	0.306	-0.123	-0.062	-0.055	-0.059	-0.027	1.005	0.870	0.871	0.871	0.789
BVARIP-M	0.183	0.173	0.156	0.143	0.244	1.527	<b>1.901</b>	<b>2.021</b>	<b>2.137</b>	<b>2.309</b>	0.207	<b>0.147</b>	<b>0.135</b>	<b>0.124</b>	<b>0.209</b>
BVARIP-SV-M	0.230	<b>0.130</b>	<b>0.119</b>	<b>0.112</b>	0.243	0.492	1.186	1.241	1.341	1.923	0.514	0.238	0.229	0.213	0.256
BVARIP vs BVARIP-SV	0.970	1.189	1.127	1.054	0.807	0.812	0.783	0.789	0.793	0.555	0.466	0.507	0.498	0.492	0.621
	<b>Idaho</b>														
BVARIP	0.269	0.551	0.538	0.537	0.196	1.163	0.944	0.965	0.970	<b>1.597</b>	0.297	0.454	0.444	0.437	<b>0.233</b>
BVARIP-SV	0.290	0.323	0.306	0.303	0.215	0.401	0.358	0.354	0.352	1.138	0.569	0.605	0.614	0.613	0.327
BVARSSVS	0.495	0.489	0.504	0.508	0.475	0.763	0.754	0.749	0.747	0.591	0.481	0.473	0.484	0.481	0.610
BVARSSVS-SV	0.306	0.331	0.328	0.326	0.297	0.042	-0.002	-0.007	-0.01	0.641	0.799	0.851	0.868	0.867	0.521
BVARHS	0.389	0.357	0.328	0.326	0.435	0.283	0.248	0.246	0.244	0.193	0.649	0.673	0.680	0.679	0.805
BVARHS-SV	0.165	<b>0.196</b>	<b>0.196</b>	<b>0.195</b>	<b>0.164</b>	0.286	0.241	0.235	0.232	0.823	0.616	0.659	0.674	0.673	0.414
BVARIP-M	0.460	0.308	0.301	0.265	0.412	0.934	<b>1.122</b>	<b>1.145</b>	<b>1.266</b>	1.294	0.432	<b>0.313</b>	<b>0.308</b>	<b>0.266</b>	0.407
BVARIP-SV-M	0.450	0.282	0.275	0.239	0.362	0.242	0.257	0.256	0.378	1.491	0.641	0.546	0.555	0.477	0.364
BVARIP vs BVARIP-SV	0.927	1.707	1.757	1.772	0.915	0.762	0.586	0.611	0.618	0.459	0.522	0.750	0.722	0.714	0.712
	<b>Illinois</b>														
BVARIP	0.481	0.333	0.301	0.315	0.487	0.935	1.119	1.139	1.132	<b>0.683</b>	0.445	0.326	0.311	0.313	0.508
BVARIP-SV	0.283	0.247	<b>0.152</b>	<b>0.157</b>	<b>0.412</b>	0.203	0.141	0.161	0.157	-0.010	0.647	0.754	0.718	0.718	0.786
BVARSSVS	0.564	0.357	0.304	0.305	0.661	0.529	0.573	0.593	0.590	-0.399	0.573	0.513	0.488	0.486	1.177
BVARSSVS-SV	0.194	0.301	0.251	0.268	0.499	-0.429	-0.500	-0.483	-0.487	-0.638	1.183	1.421	1.365	1.365	1.422
BVARHS	0.508	0.469	0.425	0.428	0.780	0.127	0.080	0.098	0.095	-0.656	0.737	0.822	0.789	0.787	1.501
BVARHS-SV	0.245	0.296	0.225	0.236	0.499	-0.186	-0.256	-0.238	-0.242	-0.437	0.932	1.114	1.067	1.067	1.177
BVARIP-M	0.996	0.278	0.295	0.283	0.807	0.934	<b>1.303</b>	<b>1.239</b>	<b>1.275</b>	-0.168	0.528	<b>0.281</b>	<b>0.300</b>	<b>0.293</b>	0.799
BVARIP-SV-M	0.564	<b>0.181</b>	0.204	0.202	0.445	0.142	0.546	0.531	0.555	0.489	0.841	0.428	0.435	0.431	<b>0.506</b>
BVARIP vs BVARIP-SV	1.703	1.351	1.976	2.008	1.184	0.732	0.978	0.978	0.975	0.693	0.687	0.432	0.433	0.435	0.646

TABLE 2.3: Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model.

	Tilting using new information														
	Point Forecasting - MRSFE					Density Forecasting - logS					Density Forecasting - CRPS				
	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4
Model vs bench	<b>Indiana</b>														
BVARIP	0.265	0.228	0.255	0.260	0.345	1.207	1.158	1.154	1.149	0.893	0.278	0.284	0.295	0.293	0.404
BVARIP-SV	0.182	0.209	0.190	0.201	0.364	0.5	0.306	0.325	0.318	0.359	0.438	0.612	0.597	0.597	0.571
BVARSSVS	0.344	0.454	0.407	0.413	0.657	0.952	0.834	0.858	0.855	0.005	0.361	0.435	0.419	0.415	0.872
BVARSSVS-SV	0.167	0.276	0.253	0.254	0.409	0.053	-0.148	-0.129	-0.136	-0.148	0.671	0.961	0.937	0.936	0.888
BVARHS	0.378	0.534	0.485	0.495	0.805	0.508	0.333	0.353	0.347	-0.392	0.488	0.651	0.633	0.631	1.225
BVARHS-SV	0.095	<b>0.171</b>	<b>0.169</b>	<b>0.174</b>	<b>0.316</b>	0.332	0.130	0.148	0.141	0.064	0.503	0.723	0.706	0.706	0.715
BVARIP-M	0.319	0.272	0.261	0.243	0.479	1.237	<b>1.302</b>	<b>1.318</b>	<b>1.386</b>	1.645	0.314	<b>0.265</b>	<b>0.254</b>	<b>0.238</b>	0.439
BVARIP-SV-M	0.336	0.266	0.252	0.244	0.462	0.594	0.518	0.479	0.511	<b>2.187</b>	0.497	0.435	0.436	0.427	<b>0.410</b>
BVARIP vs BVARIP-SV	1.458	1.088	1.343	1.294	0.948	0.707	0.852	0.828	0.83	0.534	0.636	0.463	0.494	0.491	0.706
Model vs bench	<b>Iowa</b>														
BVARIP	0.373	0.179	0.189	0.179	<b>0.477</b>	1.024	<b>1.186</b>	<b>1.180</b>	<b>1.193</b>	<b>0.726</b>	0.357	<b>0.253</b>	<b>0.256</b>	<b>0.249</b>	<b>0.468</b>
BVARIP-SV	0.237	<b>0.161</b>	<b>0.161</b>	<b>0.155</b>	0.488	0.306	0.386	0.385	0.394	0.399	0.602	0.517	0.517	0.507	0.560
BVARSSVS	0.499	0.377	0.381	0.369	0.636	0.821	0.914	0.913	0.925	0.120	0.462	0.388	0.388	0.379	0.741
BVARSSVS-SV	0.369	0.304	0.310	0.297	0.617	-0.037	0.039	0.037	0.047	-0.033	0.851	0.741	0.741	0.726	0.808
BVARHS	0.552	0.398	0.403	0.388	0.782	0.335	0.425	0.424	0.434	-0.317	0.650	0.546	0.546	0.533	1.060
BVARHS-SV	0.317	0.249	0.253	0.248	0.505	0.299	0.376	0.375	0.384	0.195	0.614	0.532	0.533	0.523	0.649
BVARIP-M	0.447	0.507	0.411	0.391	0.746	0.839	0.504	0.858	0.952	-1.453	0.448	0.519	0.418	0.393	0.878
BVARIP-SV-M	0.471	0.393	0.326	0.311	0.752	0.091	0.304	0.526	0.574	0.391	0.702	0.552	0.472	0.552	0.735
BVARIP vs BVARIP-SV	1.577	1.110	1.175	1.152	0.977	0.718	0.800	0.795	0.798	0.327	0.594	0.489	0.496	0.491	0.835
Model vs bench	<b>Kansas</b>														
BVARIP	0.269	0.585	0.441	0.418	0.369	1.241	<b>1.080</b>	<b>1.129</b>	<b>1.133</b>	<b>0.961</b>	0.276	<b>0.391</b>	<b>0.353</b>	<b>0.341</b>	<b>0.386</b>
BVARIP-SV	0.275	0.404	<b>0.311</b>	<b>0.290</b>	<b>0.342</b>	0.333	0.247	0.265	0.269	0.351	0.597	0.711	0.684	0.661	0.549
BVARSSVS	0.530	0.726	0.604	0.569	0.690	0.659	0.609	0.627	0.631	-0.108	0.515	0.566	0.544	0.525	0.923
BVARSSVS-SV	0.230	0.408	0.328	0.298	<b>0.318</b>	-0.246	-0.335	-0.319	-0.315	-0.221	1.038	1.249	1.207	1.168	0.905
BVARHS	0.362	0.553	0.460	0.436	0.477	0.287	0.201	0.218	0.221	-0.36	0.633	0.754	0.729	0.706	1.059
BVARHS-SV	0.229	0.425	0.351	0.335	0.371	0.062	-0.028	-0.012	-0.009	-0.005	0.766	0.924	0.893	0.865	0.752
BVARIP-M	0.568	0.593	0.568	0.552	0.785	0.772	0.302	0.398	0.454	-1.164	0.517	0.579	0.534	0.530	0.767
BVARIP-SV-M	0.434	<b>0.386</b>	0.365	0.356	0.656	0.202	0.267	0.278	0.270	0.333	0.686	0.587	0.567	0.576	0.663
BVARIP vs BVARIP-SV	0.978	1.447	1.418	1.442	1.078	0.907	0.833	0.863	0.864	0.610	0.463	0.550	0.517	0.515	0.703
Model vs bench	<b>Kentucky</b>														
BVARIP	0.417	0.772	0.713	0.719	0.537	0.944	0.813	0.792	0.797	0.622	0.424	0.587	0.585	0.595	0.566
BVARIP-SV	0.148	0.241	0.185	<b>0.171</b>	0.474	0.622	0.390	0.414	0.410	0.239	0.388	0.600	0.563	0.577	0.658
BVARSSVS	0.290	0.465	0.403	0.413	0.721	0.973	0.783	0.808	0.802	-0.161	0.329	0.460	0.432	0.445	0.989
BVARSSVS-SV	0.179	0.390	0.332	0.327	0.561	0.088	-0.151	-0.128	-0.133	-0.285	0.652	1.027	0.970	0.995	1.032
BVARHS	0.179	0.216	0.175	0.186	0.646	0.601	0.375	0.398	0.392	-0.428	0.399	0.604	0.570	0.586	1.189
BVARHS-SV	0.083	<b>0.203</b>	<b>0.175</b>	0.179	<b>0.303</b>	0.353	0.114	0.136	0.131	-0.068	0.494	0.779	0.736	0.757	0.799
BVARIP-M	0.689	0.397	0.38	0.368	0.429	0.869	<b>0.983</b>	<b>1.053</b>	<b>1.094</b>	<b>0.897</b>	0.508	<b>0.416</b>	<b>0.395</b>	<b>0.386</b>	<b>0.488</b>
BVARIP-SV-M	0.485	0.283	0.282	0.267	0.399	-0.019	0.431	0.426	0.464	0.598	0.917	0.530	0.539	0.524	0.558
BVARIP vs BVARIP-SV	2.807	3.202	3.852	4.205	1.133	0.322	0.422	0.377	0.387	0.383	1.093	0.979	1.038	1.031	0.860
Model vs bench	<b>Louisiana</b>														
BVARIP	0.193	0.196	0.196	0.193	<b>0.286</b>	1.456	1.346	1.35	1.352	<b>1.318</b>	0.213	0.224	0.224	0.223	<b>0.277</b>
BVARIP-SV	0.164	0.174	<b>0.166</b>	0.163	0.347	0.724	0.582	0.590	0.590	0.964	0.371	0.422	0.417	0.416	0.364
BVARSSVS	0.234	0.235	0.226	0.223	0.438	1.365	1.274	1.284	1.286	0.724	0.245	0.252	0.248	0.246	0.466
BVARSSVS-SV	0.203	0.239	0.233	0.231	0.424	0.611	0.465	0.473	0.472	0.656	0.419	0.481	0.476	0.475	0.478
BVARHS	0.257	0.302	0.283	0.283	0.548	0.770	0.627	0.638	0.637	0.166	0.376	0.430	0.422	0.422	0.731
BVARHS-SV	0.139	0.176	0.174	0.172	0.305	0.960	0.811	0.819	0.818	0.897	0.294	0.341	0.338	0.337	0.365
BVARIP-M	1.007	0.181	0.186	0.174	0.691	0.504	<b>3.042</b>	<b>2.782</b>	<b>3.213</b>	-1.391	0.747	<b>0.149</b>	<b>0.157</b>	<b>0.143</b>	0.702
BVARIP-SV-M	0.854	<b>0.166</b>	0.168	<b>0.153</b>	0.616	0.016	2.650	2.347	2.796	0.736	0.938	0.184	0.199	0.178	0.580
BVARIP vs BVARIP-SV	1.178	1.124	1.179	1.178	0.825	0.731	0.764	0.759	0.762	0.354	0.574	0.532	0.539	0.535	0.762
Model vs bench	<b>Maine</b>														
BVARIP	0.227	0.240	0.232	0.24	0.376	1.417	<b>1.114</b>	<b>1.121</b>	<b>1.113</b>	<b>1.153</b>	0.237	<b>0.289</b>	<b>0.287</b>	<b>0.287</b>	<b>0.374</b>
BVARIP-SV	0.116	<b>0.125</b>	<b>0.117</b>	<b>0.128</b>	<b>0.152</b>	0.484	0.073	0.078	0.068	0.713	0.456	0.710	0.710	0.710	0.398
BVARSSVS	0.277	0.387	0.372	0.392	0.476	0.779	0.384	0.390	0.381	0.205	0.380	0.566	0.564	0.565	0.711
BVARSSVS-SV	0.134	0.233	0.235	0.258	0.255	-0.201	-0.615	-0.61	-0.619	0.035	0.897	1.410	1.411	1.412	0.778
BVARHS	0.248	0.39	0.366	0.384	0.552	0.329	-0.083	-0.076	-0.086	-0.118	0.546	0.853	0.849	0.85	0.953
BVARHS-SV	0.135	0.231	0.23	0.239	0.264	0.147	-0.267	-0.262	-0.271	0.324	0.633	0.996	0.997	0.996	0.594
BVARIP-M	0.500	0.376	0.375	0.374	0.466	0.863	1.008	0.993	0.977	0.319	0.483	0.358	0.356	0.362	0.504
BVARIP-SV-M	0.357	0.195	0.181	0.179	0.426	-0.007	0.459	0.485	0.486	0.454	0.794	0.451	0.432	0.432	0.532
BVARIP vs BVARIP-SV	1.948	1.924	1.984	1.876	2.469	0.933	1.041	1.043	1.045	0.439	0.520	0.407	0.405	0.404	0.938
Model vs bench	<b>Maryland</b>														
BVARIP	0.154	0.202	0.207	0.213	0.450	1.709	1.247	<b>1.242</b>	<b>1.241</b>	<b>0.848</b>	0.166	<b>0.255</b>	<b>0.258</b>	<b>0.256</b>	0.454
BVARIP-SV	0.110	<b>0.172</b>	<b>0.179</b>	<b>0.189</b>	<b>0.288</b>	0.627	0.099	0.097	0.093	0.412	0.392	0.705	0.706	0.706	0.531
BVARSSVS	0.319	0.417	0.414	0.422	0.541	0.795	0.307	0.306	0.305	-0.106	0.386	0.623	0.623	0.618	0.905
BVARSSVS-SV	0.140	0.198	0.197	0.203	0.298	-0.136	-0.665	-0.666	-0.67	-0.298	0.832	1.499	1.499	1.499	1.027
BVARHS	0.137	0.242	0.228	0.24	0.350	0.471	-0.059	-0.059	-0.063	-0.339	0.457	0.826	0.824	0.824	1.074
BVARHS-SV	0.113	0.203	0.202	0.211	0.306	0.254	-0.276	-0.277	-0.281	0.073	0.563	1.018	1.017	1.017	0.724
BVARIP-M	0.460	0.269	0.338	0.335	0.476	0.875	<b>1.272</b>	1.160	1.163	0.487	0.460	0.271	0.329	0.325	0.450
BVARIP-SV-M	0.189	0.258	0.283	0.272	0.401	0.202	0.247	0.164	0.175	0.752	0.601	0.561	0.645	0.630	<b>0.435</b>
BVARIP vs BVARIP-SV	1.407	1.172	1.158	1.126	1.561	1.081	1.148	1.145	1.148	0.435	0.423	0.362	0.366	0.363	0.8



Model vs bench	Tilting using new information														
	Point Forecasting - MRSFE					Density Forecasting - logS					Density Forecasting - CRPS				
	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4
	<b>Massachusetts</b>														
BVARIP	0.232	0.242	0.238	0.241	0.236	1.355	<b>1.259</b>	<b>1.259</b>	<b>1.257</b>	<b>1.910</b>	0.243	<b>0.269</b>	<b>0.268</b>	<b>0.268</b>	<b>0.217</b>
BVARIP-SV	0.192	0.206	0.206	0.209	0.165	0.428	0.275	0.262	0.261	1.252	0.500	0.627	0.645	0.645	0.321
BVARSSVS	0.523	0.492	0.519	0.511	0.485	0.678	0.632	0.620	0.622	0.822	0.515	0.518	0.532	0.528	0.557
BVARSSVS-SV	0.156	<b>0.181</b>	0.171	0.19	<b>0.140</b>	-0.253	-0.410	-0.423	-0.424	0.583	0.964	1.224	1.261	1.262	0.604
BVARHS	0.322	0.317	0.329	0.327	0.256	0.338	0.194	0.181	0.180	0.571	0.564	0.689	0.709	0.708	0.623
BVARHS-SV	0.125	0.182	<b>0.163</b>	<b>0.170</b>	0.141	0.166	0.008	-0.004	-0.006	0.941	0.634	0.808	0.830	0.831	0.426
BVARIP-M	0.344	0.425	0.382	0.421	0.442	1.106	0.877	0.955	0.856	0.837	0.335	0.420	0.378	0.422	0.458
BVARIP-SV-M	0.217	0.341	0.287	0.308	0.443	0.075	0.243	0.316	0.232	0.282	0.745	0.586	0.524	0.576	0.600
BVARIP vs BVARIP-SV	1.208	1.176	1.159	1.152	1.435	0.927	0.984	0.996	0.996	0.658	0.487	0.429	0.416	0.416	0.674
	<b>Michigan</b>														
BVARIP	0.234	0.134	0.183	0.199	0.339	1.329	<b>1.267</b>	<b>1.257</b>	<b>1.241</b>	<b>1.004</b>	0.247	<b>0.236</b>	<b>0.247</b>	<b>0.254</b>	<b>0.347</b>
BVARIP-SV	0.130	0.136	0.135	0.136	0.411	0.629	0.410	0.427	0.416	0.462	0.383	0.534	0.517	0.525	0.518
BVARSSVS	0.229	0.255	0.226	0.232	0.407	1.292	1.155	1.180	1.170	0.299	0.250	0.292	0.276	0.279	0.580
BVARSSVS-SV	0.133	0.207	0.193	0.198	0.444	0.248	0.023	0.041	0.029	-0.076	0.554	0.786	0.759	0.772	0.799
BVARHS	0.232	0.235	0.214	0.219	0.397	0.700	0.499	0.518	0.507	-0.297	0.378	0.499	0.481	0.488	0.962
BVARHS-SV	0.085	<b>0.122</b>	<b>0.121</b>	<b>0.126</b>	<b>0.303</b>	0.682	0.458	0.475	0.463	0.317	0.358	0.506	0.490	0.498	0.538
BVARIP-M	0.605	0.372	0.421	0.348	0.757	0.638	0.853	0.783	1.093	-2.409	0.546	0.345	0.397	0.326	0.766
BVARIP-SV-M	0.629	0.300	0.351	0.287	0.648	0.122	0.782	0.485	0.779	0.718	0.758	0.412	0.514	0.426	0.584
BVARIP vs BVARIP-SV	1.804	0.99	1.356	1.461	0.824	0.700	0.857	0.83	0.826	0.542	0.646	0.442	0.478	0.484	0.670
	<b>Minnesota</b>														
BVARIP	0.318	0.197	<b>0.190</b>	<b>0.164</b>	0.481	1.186	<b>1.223</b>	<b>1.218</b>	<b>1.226</b>	<b>0.736</b>	0.312	<b>0.266</b>	<b>0.269</b>	<b>0.263</b>	<b>0.472</b>
BVARIP-SV	0.295	0.354	0.374	0.364	0.547	0.414	0.289	0.278	0.280	0.377	0.531	0.653	0.672	0.670	0.601
BVARSSVS	0.462	0.448	0.465	0.454	0.632	0.692	0.638	0.630	0.634	0.015	0.483	0.502	0.513	0.509	0.795
BVARSSVS-SV	0.361	0.501	0.525	0.519	0.625	-0.198	-0.330	-0.341	-0.340	-0.166	0.945	1.191	1.226	1.225	0.903
BVARHS	0.298	0.292	0.286	0.276	0.316	0.315	0.193	0.184	0.185	-0.256	0.578	0.704	0.721	0.72	0.905
BVARHS-SV	0.112	<b>0.196</b>	0.206	0.202	<b>0.250</b>	0.138	0.003	-0.008	-0.007	0.186	0.66	0.837	0.863	0.862	0.589
BVARIP-M	0.288	0.42	0.386	0.373	0.498	1.267	1.083	1.173	1.187	0.577	0.286	0.386	0.346	0.332	0.518
BVARIP-SV-M	0.289	0.236	0.256	0.260	0.462	0.176	0.258	0.233	0.252	0.280	0.672	0.630	0.647	0.618	0.583
BVARIP vs BVARIP-SV	1.078	0.555	0.508	0.451	0.879	0.773	0.934	0.939	0.946	0.359	0.587	0.407	0.400	0.392	0.786
	<b>Mississippi</b>														
BVARIP	0.175	0.219	0.198	0.198	0.536	1.240	1.004	1.012	1.008	0.661	0.234	0.321	0.318	0.317	0.555
BVARIP-SV	0.146	0.235	0.231	0.226	0.358	0.414	0.136	0.140	0.136	0.159	0.474	0.728	0.728	0.727	0.664
BVARSSVS	0.271	0.445	0.418	0.421	0.664	0.864	0.604	0.612	0.608	-0.204	0.347	0.509	0.505	0.503	0.997
BVARSSVS-SV	0.146	0.316	0.309	0.307	0.452	-0.17	-0.451	-0.446	-0.451	-0.480	0.838	1.300	1.300	1.299	1.210
BVARHS	0.167	0.204	0.204	0.204	0.478	0.381	0.107	0.112	0.107	-0.582	0.491	0.743	0.744	0.744	1.334
BVARHS-SV	0.128	0.274	0.269	0.269	0.386	0.239	-0.042	-0.037	-0.042	-0.061	0.558	0.865	0.866	0.865	0.812
BVARIP-M	0.199	0.428	0.534	0.531	<b>0.195</b>	1.257	0.728	-0.021	0.004	<b>1.385</b>	0.247	0.455	0.594	0.589	0.213
BVARIP-SV-M	0.186	<b>0.142</b>	<b>0.174</b>	<b>0.175</b>	0.210	0.137	<b>1.920</b>	<b>1.055</b>	<b>1.078</b>	0.284	0.645	<b>0.273</b>	<b>0.367</b>	<b>0.368</b>	<b>0.519</b>
BVARIP vs BVARIP-SV	1.200	0.929	0.856	0.874	1.497	0.826	0.868	0.871	0.872	0.502	0.494	0.442	0.437	0.436	0.836
	<b>Missouri</b>														
BVARIP	0.317	0.333	0.345	0.342	0.467	1.201	<b>1.190</b>	<b>1.185</b>	<b>1.193</b>	0.798	0.321	<b>0.318</b>	<b>0.316</b>	<b>0.319</b>	0.485
BVARIP-SV	0.184	0.165	<b>0.170</b>	<b>0.171</b>	0.228	0.415	0.321	0.314	0.314	0.715	0.516	0.593	0.593	0.610	0.400
BVARSSVS	0.378	0.331	0.34	0.342	0.413	0.82	0.79	0.784	0.786	0.414	0.419	0.412	0.411	0.421	0.568
BVARSSVS-SV	0.150	0.204	0.207	0.213	<b>0.192</b>	-0.321	-0.421	-0.428	-0.428	-0.001	1.055	1.236	1.236	1.272	0.774
BVARHS	0.339	0.386	0.391	0.393	0.500	0.428	0.337	0.332	0.332	0.110	0.54	0.615	0.612	0.629	0.753
BVARHS-SV	0.125	<b>0.165</b>	<b>0.170</b>	0.175	<b>0.215</b>	0.147	0.047	0.040	0.040	0.428	0.661	0.774	0.774	0.797	0.514
BVARIP-M	0.335	0.638	0.634	0.636	0.335	1.176	0.215	0.314	0.313	<b>1.224</b>	0.336	0.681	0.659	0.657	<b>0.333</b>
BVARIP-SV-M	0.242	0.178	0.191	0.206	0.214	0.15	0.283	0.250	0.251	0.560	0.678	0.576	0.591	0.594	0.446
BVARIP vs BVARIP-SV	1.719	2.022	2.028	2.006	2.043	0.786	0.869	0.871	0.879	0.083	0.621	0.535	0.532	0.523	1.212
	<b>Montana</b>														
BVARIP	0.187	0.291	0.276	0.270	0.420	1.407	<b>1.214</b>	<b>1.241</b>	<b>1.247</b>	0.830	0.227	<b>0.297</b>	<b>0.285</b>	<b>0.282</b>	0.445
BVARIP-SV	0.147	0.195	0.191	0.189	0.370	0.536	0.432	0.448	0.448	0.418	0.476	0.510	0.501	0.501	0.513
BVARSSVS	0.327	0.29	0.285	0.281	0.662	1.045	0.979	0.995	0.997	0.031	0.350	0.340	0.335	0.333	0.803
BVARSSVS-SV	0.259	0.257	0.251	0.247	0.505	0.060	-0.039	-0.023	-0.022	-0.186	0.768	0.808	0.793	0.793	0.883
BVARHS	0.311	0.308	0.302	0.304	0.697	0.561	0.466	0.482	0.482	-0.357	0.492	0.513	0.504	0.505	1.075
BVARHS-SV	0.153	<b>0.156</b>	<b>0.156</b>	<b>0.158</b>	0.368	0.510	0.411	0.427	0.427	0.191	0.487	0.513	0.504	0.504	0.611
BVARIP-M	0.179	0.373	0.413	0.402	<b>0.222</b>	1.426	1.027	0.983	1.013	<b>1.400</b>	0.218	0.368	0.393	0.388	<b>0.227</b>
BVARIP-SV-M	0.214	0.171	0.192	0.197	0.233	0.263	0.478	0.373	0.394	0.478	0.612	0.463	0.515	0.509	0.441
BVARIP vs BVARIP-SV	1.269	1.493	1.442	1.430	1.135	0.871	0.782	0.793	0.799	0.412	0.477	0.583	0.569	0.563	0.869
	<b>Nebraska</b>														
BVARIP	0.369	<b>0.221</b>	<b>0.199</b>	<b>0.176</b>	0.229	1.149	<b>1.237</b>	<b>1.253</b>	<b>1.267</b>	1.933	0.327	<b>0.264</b>	<b>0.253</b>	<b>0.245</b>	<b>0.219</b>
BVARIP-SV	0.307	0.320	0.270	0.248	0.198	0.363	0.368	0.387	0.397	1.431	0.600	0.599	0.57	0.557	0.280
BVARSSVS	0.456	0.439	0.389	0.358	0.299	0.766	0.784	0.801	0.812	1.042	0.452	0.436	0.417	0.407	0.414
BVARSSVS-SV	0.251	0.233	0.205	0.189	0.200	-0.202	-0.196	-0.179	-0.169	0.844	1.026	1.023	0.978	0.957	0.469
BVARHS	0.359	0.368	0.318	0.292	0.249	0.457	0.463	0.481	0.491	0.823	0.556	0.553	0.527	0.515	0.486
BVARHS-SV	0.260	0.289	0.255	0.236	0.225	0.285	0.289	0.306	0.316	1.197	0.637	0.639	0.611	0.597	0.346
BVARIP-M	0.394	0.732	0.825	1.021	<b>0.145</b>	1.133	0.821	0.821	0.803	<b>2.595</b>	0.340	0.551	0.563	0.595	0.140
BVARIP-SV-M	0.286	0.409	0.451	0.570	0.160	0.105	0.083	0.069	0.046	1.772	0.752	0.808	0.829	0.880	0.255
BVARIP vs BVARIP-SV	1.201	0.691	0.739	0.710	1.160	0.786	0.869	0.867	0.870	0.501	0.544	0.441	0.444	0.440	0.784



Model vs bench	Tilting using new information														
	Point Forecasting - MRSFE					Density Forecasting - logS					Density Forecasting - CRPS				
	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4
	<b>Nevada</b>														
BVARIP	0.200	0.155	0.156	0.155	0.551	1.548	1.498	1.517	1.510	0.599	0.21	0.192	0.19	0.188	0.548
BVARIP-SV	0.154	0.156	0.134	0.136	0.371	0.654	0.522	0.546	0.536	0.101	0.409	0.447	0.436	0.436	0.653
BVARSSVS	0.287	0.26	0.247	0.249	0.774	0.854	0.742	0.768	0.757	-0.588	0.372	0.385	0.375	0.375	1.311
BVARSSVS-SV	0.134	0.166	0.161	0.16	0.436	0.101	-0.034	-0.012	-0.022	-0.570	0.696	0.767	0.754	0.754	1.226
BVARHS	0.466	0.435	0.423	0.432	1.487	0.454	0.340	0.363	0.352	-0.918	0.565	0.589	0.577	0.579	1.945
BVARHS-SV	0.125	0.126	0.124	0.126	0.503	0.581	0.448	0.470	0.460	-0.143	0.433	0.474	0.466	0.467	0.842
BVARIP-M	0.183	0.138	0.127	0.122	<b>0.267</b>	1.514	<b>3.183</b>	<b>3.545</b>	<b>3.598</b>	<b>1.231</b>	0.198	<b>0.117</b>	<b>0.110</b>	<b>0.107</b>	<b>0.275</b>
BVARIP-SV-M	0.241	<b>0.110</b>	<b>0.097</b>	<b>0.097</b>	0.352	0.329	2.125	2.471	2.514	0.069	0.528	0.251	0.238	0.236	0.676
BVARIP vs BVARIP-SV	1.301	0.993	1.164	1.142	1.486	0.894	0.976	0.971	0.974	0.497	0.512	0.428	0.435	0.431	0.840
	<b>New Hampshire</b>														
BVARIP	0.271	0.209	0.262	0.268	0.259	1.498	1.480	1.443	1.440	<b>2.128</b>	0.246	0.218	0.238	0.239	0.240
BVARIP-SV	0.177	0.164	0.199	0.216	<b>0.174</b>	0.679	0.538	0.502	0.499	1.663	0.420	0.480	0.521	0.523	0.280
BVARSSVS	0.244	0.250	0.302	0.308	0.311	0.984	0.849	0.814	0.812	1.146	0.331	0.372	0.402	0.402	0.474
BVARSSVS-SV	0.297	0.318	0.385	0.396	0.311	-0.048	-0.193	-0.228	-0.231	0.912	0.861	0.995	1.080	1.081	0.581
BVARHS	0.275	0.299	0.358	0.370	0.296	0.515	0.373	0.338	0.335	0.825	0.504	0.578	0.626	0.627	0.626
BVARHS-SV	0.146	0.164	0.205	0.213	0.192	0.52	0.374	0.338	0.336	1.455	0.485	0.561	0.610	0.611	0.340
BVARIP-M	0.330	0.152	0.175	0.173	0.289	1.375	<b>2.037</b>	<b>1.786</b>	<b>1.816</b>	1.783	0.292	<b>0.144</b>	<b>0.172</b>	<b>0.167</b>	<b>0.281</b>
BVARIP-SV-M	0.211	<b>0.100</b>	<b>0.125</b>	<b>0.127</b>	0.175	0.465	1.147	0.864	0.896	1.308	0.504	0.239	0.297	0.289	0.325
BVARIP vs BVARIP-SV	1.527	1.273	1.316	1.24	1.492	0.818	0.942	0.941	0.94	0.464	0.586	0.454	0.456	0.457	0.859
	<b>New Jersey</b>														
BVARIP	0.335	0.631	0.548	0.515	0.875	1.093	<b>0.963</b>	<b>1.004</b>	<b>1.019</b>	0.120	0.347	0.453	0.417	0.404	0.904
BVARIP-SV	0.226	0.622	0.606	0.630	0.771	0.093	-0.175	-0.175	-0.329	-0.329	0.679	1.126	1.125	1.128	1.106
BVARSSVS	0.409	0.542	0.485	0.470	0.639	0.517	0.316	0.321	0.323	-0.607	0.515	0.709	0.698	0.695	1.366
BVARSSVS-SV	0.209	0.733	0.720	0.745	1.008	-0.679	-0.947	-0.948	-0.947	-1.087	1.442	<b>2.392</b>	<b>2.391</b>	<b>2.393</b>	<b>2.204</b>
BVARHS	0.326	0.480	0.464	0.454	0.674	0.161	-0.09	-0.090	-0.089	-0.830	0.654	1.025	1.023	1.022	1.684
BVARHS-SV	0.110	0.364	0.315	0.291	0.373	-0.124	-0.391	-0.391	-0.390	-0.550	0.825	1.366	1.363	1.361	1.255
BVARIP-M	0.379	0.514	0.412	0.427	0.520	1.062	0.792	0.947	0.936	<b>0.524</b>	0.380	0.507	0.407	0.419	0.571
BVARIP-SV-M	0.274	<b>0.257</b>	<b>0.203</b>	<b>0.218</b>	<b>0.307</b>	-0.080	-0.115	-0.054	-0.069	-0.094	0.830	0.854	0.744	0.679	0.777
BVARIP vs BVARIP-SV	1.483	1.016	0.904	0.817	1.135	1.000	1.137	1.179	1.194	0.449	0.512	0.402	0.371	0.358	0.818
	<b>New Mexico</b>														
BVARIP	0.230	0.206	0.210	0.207	0.303	1.353	<b>1.419</b>	<b>1.399</b>	<b>1.400</b>	1.326	0.235	<b>0.211</b>	<b>0.216</b>	<b>0.215</b>	0.298
BVARIP-SV	0.183	0.142	0.125	0.123	0.283	0.665	0.709	0.692	0.685	0.842	0.364	0.337	0.340	0.343	0.369
BVARSSVS	0.212	0.177	0.178	0.18	0.446	1.292	1.365	1.348	1.34	0.55	0.234	0.205	0.208	0.210	0.520
BVARSSVS-SV	0.144	0.15	0.152	0.151	0.335	0.282	0.316	0.295	0.289	0.274	0.514	0.490	0.500	0.505	0.605
BVARHS	0.201	0.149	0.144	0.145	0.465	0.760	0.809	0.790	0.783	0.045	0.338	0.308	0.312	0.315	0.77
BVARHS-SV	0.103	<b>0.107</b>	<b>0.108</b>	<b>0.108</b>	0.252	0.818	0.852	0.831	0.824	0.76	0.302	0.289	0.294	0.297	0.383
BVARIP-M	0.292	0.481	0.426	0.434	<b>0.172</b>	1.232	-0.489	-0.228	-0.347	<b>1.902</b>	0.288	0.574	0.496	0.508	<b>0.183</b>
BVARIP-SV-M	0.174	0.200	0.170	0.162	0.139	0.424	0.847	1.105	1.119	0.981	0.451	0.327	0.278	0.274	0.337
BVARIP vs BVARIP-SV	1.255	1.457	1.676	1.674	1.070	0.688	0.710	0.707	0.715	0.484	0.645	0.626	0.636	0.627	0.809
	<b>New York</b>														
BVARIP	0.544	0.446	0.404	0.451	1.018	0.909	<b>1.080</b>	<b>1.137</b>	<b>1.110</b>	-0.805	0.463	<b>0.375</b>	<b>0.346</b>	<b>0.361</b>	1.018
BVARIP-SV	0.343	0.365	0.325	0.366	0.866	0.157	0.129	0.127	0.105	-0.165	0.712	0.75	0.762	0.789	0.987
BVARSSVS	0.624	<b>0.292</b>	<b>0.249</b>	<b>0.261</b>	0.542	0.584	0.705	0.710	0.691	-0.272	0.593	0.434	0.433	0.445	0.943
BVARSSVS-SV	0.446	0.413	0.373	0.433	0.853	-0.526	-0.554	-0.559	-0.58	-0.782	1.379	1.45	1.489	1.539	1.562
BVARHS	0.498	0.378	0.373	0.374	0.65	0.286	0.283	0.281	0.263	-0.505	0.666	0.65	0.663	0.678	1.179
BVARHS-SV	0.492	0.514	0.451	0.497	0.967	-0.007	-0.033	-0.032	-0.053	-0.335	0.852	0.891	0.899	0.928	1.131
BVARIP-M	0.325	0.399	0.371	0.370	<b>0.219</b>	1.187	0.797	0.858	0.883	<b>1.364</b>	0.318	0.408	0.366	0.365	<b>0.230</b>
BVARIP-SV-M	0.366	0.294	0.299	0.291	0.676	-0.043	0.397	0.506	0.487	-0.034	0.849	0.507	0.457	0.459	0.826
BVARIP vs BVARIP-SV	1.587	1.221	1.241	1.23	1.175	0.752	0.951	1.010	1.004	-0.641	0.651	0.500	0.455	0.458	1.032
	<b>North Carolina</b>														
BVARIP	0.246	0.405	0.438	0.42	0.683	1.282	1.031	1.015	1.035	0.268	0.254	0.385	0.404	0.390	0.762
BVARIP-SV	0.164	0.236	0.250	0.230	<b>0.299</b>	0.427	0.189	0.169	0.171	0.048	0.444	0.655	0.691	0.688	0.714
BVARSSVS	0.380	0.394	0.413	0.401	0.733	0.812	0.687	0.673	0.678	-0.248	0.398	0.458	0.475	0.470	1.049
BVARSSVS-SV	0.189	0.305	0.305	0.276	0.393	-0.294	-0.535	-0.555	-0.553	-0.701	0.895	1.331	1.404	1.401	1.474
BVARHS	0.747	1.149	1.256	1.249	2.771	0.283	0.070	0.049	0.054	-0.989	0.758	1.087	1.150	1.143	3.020
BVARHS-SV	0.117	0.220	<b>0.211</b>	<b>0.189</b>	0.393	0.250	0.007	-0.012	-0.010	-0.113	0.519	0.776	0.816	0.814	0.846
BVARIP-M	0.293	0.235	0.237	0.235	0.367	1.207	<b>1.308</b>	<b>1.298</b>	<b>1.301</b>	<b>1.024</b>	0.293	<b>0.250</b>	<b>0.253</b>	<b>0.254</b>	<b>0.369</b>
BVARIP-SV-M	0.190	<b>0.218</b>	0.260	0.277	0.329	0.16	0.135	0.115	0.108	0.014	0.584	0.626	0.648	0.667	0.732
BVARIP vs BVARIP-SV	1.502	1.716	1.751	1.823	2.286	0.855	0.842	0.846	0.864	0.221	0.572	0.588	0.584	0.567	1.067
	<b>North Dakota</b>														
BVARIP	0.328	0.300	0.279	0.267	0.182	1.289	<b>1.332</b>	<b>1.361</b>	<b>1.370</b>	<b>4.515</b>	0.276	0.263	0.251	0.246	0.158
BVARIP-SV	0.326	0.235	0.223	0.216	0.154	0.637	0.736	0.758	0.763	4.348	0.465	0.393	0.38	0.375	0.162
BVARSSVS	0.548	0.382	0.363	0.352	0.324	1.185	1.329	1.355	1.358	4.027	0.39	0.313	0.301	0.298	0.286
BVARSSVS-SV	0.367	0.276	0.263	0.253	0.213	0.696	0.793	0.814	0.819	4.177	0.448	0.381	0.369	0.364	0.204
BVARHS	0.645	0.437	0.419	0.403	0.366	0.747	0.866	0.887	0.893	3.600	0.503	0.407	0.394	0.388	0.361
BVARHS-SV	0.189	<b>0.133</b>	<b>0.126</b>	<b>0.122</b>	<b>0.120</b>	1.105	1.206	1.227	1.232	4.445	0.288	<b>0.243</b>	<b>0.235</b>	<b>0.232</b>	<b>0.140</b>
BVARIP-M	0.354	0.333	0.306	0.273	0.183	1.205	1.244	1.343	1.514	4.174	0.312	0.306	0.279	0.251	0.150
BVARIP-SV-M	0.350	0.277	0.259	0.234	0.166	0.333	0.707	0.777	0.933	3.911	0.614	0.400	0.371	0.335	0.166
BVARIP vs BVARIP-SV	1.005	1.280	1.255	1.235	1.182	0.652	0.595	0.603	0.607	0.167	0.594	0.671	0.662	0.656	0.976

Model vs bench	Tilting using new information														
	Point Forecasting - MRSFE					Density Forecasting - logS					Density Forecasting - CRPS				
	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4
	<b>Ohio</b>														
BVARIP	0.456	0.571	0.558	0.546	0.455	0.975	0.936	0.93	0.958	<b>0.778</b>	0.443	0.503	0.497	0.486	<b>0.475</b>
BVARIP-SV	0.171	<b>0.168</b>	<b>0.158</b>	<b>0.140</b>	<b>0.226</b>	0.356	0.271	0.281	0.285	0.445	0.547	0.64	0.623	0.621	0.477
BVARSSVS	0.371	0.288	0.263	0.255	0.369	0.902	0.901	0.916	0.922	0.273	0.392	0.372	0.357	0.355	0.589
BVARSSVS-SV	0.159	0.221	0.199	0.19	0.272	-0.375	-0.464	-0.454	-0.451	-0.315	1.117	1.324	1.289	1.289	0.989
BVARHS	0.359	0.344	0.324	0.316	0.486	0.486	0.419	0.43	0.434	-0.059	0.518	0.576	0.559	0.557	0.815
BVARHS-SV	0.145	0.226	0.234	0.228	0.259	0.233	0.142	0.151	0.154	0.263	0.611	0.728	0.711	0.71	0.569
BVARIP-M	0.592	0.251	0.255	0.251	0.491	0.704	<b>1.432</b>	<b>1.429</b>	<b>1.434</b>	0.494	0.57	<b>0.245</b>	<b>0.255</b>	<b>0.250</b>	0.538
BVARIP-SV-M	0.382	0.295	0.254	0.243	0.245	0.065	0.148	0.153	0.148	0.164	0.786	0.658	0.663	0.661	0.594
BVARIP vs BVARIP-SV	2.664	3.402	3.543	3.9	2.011	0.62	0.665	0.649	0.673	0.333	0.811	0.786	0.798	0.782	0.997
	<b>Oklahoma</b>														
BVARIP	0.209	0.172	0.168	0.160	0.232	1.537	1.594	1.587	1.622	3.559	0.216	0.19	0.188	0.183	0.229
BVARIP-SV	0.199	0.185	0.189	0.182	0.26	0.733	0.776	0.76	0.79	3.175	0.399	0.365	0.369	0.361	0.289
BVARSSVS	0.299	0.222	0.224	0.215	0.35	1.291	1.381	1.368	1.4	2.954	0.293	0.242	0.243	0.237	0.376
BVARSSVS-SV	0.202	0.179	0.183	0.175	0.235	0.343	0.390	0.374	0.404	2.709	0.575	0.520	0.526	0.515	0.397
BVARHS	0.255	0.208	0.21	0.202	0.311	0.896	0.947	0.932	0.962	2.604	0.357	0.319	0.321	0.315	0.454
BVARHS-SV	0.173	0.154	0.16	0.153	0.224	0.866	0.912	0.895	0.925	3.177	0.348	0.316	0.32	0.313	0.271
BVARIP-M	0.292	<b>0.151</b>	<b>0.148</b>	<b>0.127</b>	<b>0.197</b>	1.427	<b>2.164</b>	<b>2.039</b>	<b>2.313</b>	<b>3.804</b>	0.276	<b>0.147</b>	<b>0.149</b>	<b>0.128</b>	<b>0.191</b>
BVARIP-SV-M	0.319	0.168	0.172	0.151	0.273	0.494	1.171	1.000	1.259	3.228	0.525	0.306	0.334	0.294	0.298
BVARIP vs BVARIP-SV	1.051	0.928	0.890	0.877	0.892	0.804	0.818	0.828	0.832	0.384	0.541	0.521	0.510	0.505	0.793
	<b>Oregon</b>														
BVARIP	0.335	0.618	0.571	0.55	0.491	1.171	0.977	1.024	1.031	0.647	0.321	0.447	0.414	0.41	0.495
BVARIP-SV	0.295	0.31	0.305	0.287	0.33	0.201	0.124	0.122	0.122	0.292	0.68	0.776	0.775	0.774	0.559
BVARSSVS	0.637	0.709	0.716	0.692	0.791	0.632	0.618	0.62	0.623	-0.047	0.586	0.593	0.589	0.584	0.917
BVARSSVS-SV	0.226	0.338	0.324	0.302	0.292	-0.465	-0.548	-0.549	-0.55	-0.449	1.288	1.499	1.497	1.496	1.098
BVARHS	0.496	0.585	0.587	0.574	0.716	0.291	0.219	0.218	0.217	-0.359	0.664	0.746	0.744	0.743	1.098
BVARHS-SV	0.207	0.308	0.309	0.294	0.294	0.216	0.132	0.13	0.13	0.258	0.657	0.767	0.766	0.765	0.566
BVARIP-M	0.325	0.201	0.16	0.154	0.268	1.203	<b>1.669</b>	<b>1.977</b>	<b>2.037</b>	<b>1.222</b>	0.31	<b>0.193</b>	<b>0.151</b>	<b>0.149</b>	<b>0.266</b>
BVARIP-SV-M	0.33	<b>0.096</b>	<b>0.064</b>	<b>0.060</b>	<b>0.258</b>	-0.035	0.647	0.915	0.977	0.171	0.866	0.395	0.322	0.305	0.563
BVARIP vs BVARIP-SV	1.137	1.992	1.87	1.918	1.489	0.97	0.853	0.902	0.91	0.355	0.472	0.576	0.534	0.53	0.885
	<b>Pennsylvania</b>														
BVARIP	0.424	0.533	0.508	0.494	0.777	0.982	0.926	0.955	0.966	0.069	0.405	0.454	0.438	0.428	0.794
BVARIP-SV	0.243	0.268	0.264	0.269	0.56	-0.181	-0.251	-0.245	-0.246	-0.345	0.935	1.064	1.064	1.064	1.002
BVARSSVS	0.731	0.493	0.478	0.463	0.866	0.256	0.29	0.297	0.298	-0.547	0.762	0.668	0.664	0.659	1.271
BVARSSVS-SV	0.223	0.39	0.431	0.442	0.566	-1.039	-1.112	-1.106	-1.107	-1.16	2.184	2.501	2.504	2.504	2.156
BVARHS	0.413	<b>0.254</b>	<b>0.230</b>	<b>0.248</b>	0.657	-0.031	-0.086	-0.079	-0.081	-0.674	0.832	0.902	0.899	0.901	1.368
BVARHS-SV	0.221	0.313	0.313	0.319	<b>0.423</b>	-0.364	-0.437	-0.431	-0.432	-0.463	1.114	1.276	1.276	1.277	1.090
BVARIP-M	0.527	0.532	0.401	0.403	0.605	0.933	<b>0.944</b>	<b>1.083</b>	<b>1.082</b>	<b>0.274</b>	0.471	<b>0.448</b>	<b>0.364</b>	<b>0.376</b>	<b>0.666</b>
BVARIP-SV-M	0.454	0.342	0.292	0.318	0.501	-0.399	-0.407	-0.352	-0.365	-0.265	1.224	1.206	1.1	1.153	0.948
BVARIP vs BVARIP-SV	1.745	1.991	1.923	1.835	1.389	1.163	1.177	1.2	1.212	0.413	0.433	0.427	0.412	0.403	0.792
	<b>Rhode Island</b>														
BVARIP	0.311	0.288	0.282	0.286	0.523	1.256	<b>1.233</b>	<b>1.238</b>	<b>1.236</b>	0.478	0.324	<b>0.292</b>	<b>0.286</b>	<b>0.293</b>	0.545
BVARIP-SV	0.163	0.207	0.201	0.212	0.257	0.439	0.186	0.18	0.177	0.472	0.49	0.609	0.608	0.625	0.496
BVARSSVS	0.221	0.258	0.25	0.251	0.475	0.967	0.724	0.72	0.719	0.231	0.318	0.385	0.381	0.39	0.667
BVARSSVS-SV	0.114	0.202	0.197	0.198	<b>0.180</b>	-0.357	-0.61	-0.617	-0.619	-0.328	1.063	1.327	1.326	1.359	1.055
BVARHS	0.203	0.247	0.241	0.246	0.503	0.536	0.284	0.278	0.276	-0.061	0.453	0.56	0.559	0.573	0.859
BVARHS-SV	0.128	<b>0.151</b>	<b>0.146</b>	<b>0.147</b>	0.302	0.334	0.082	0.076	0.074	0.326	0.536	0.665	0.665	0.681	0.575
BVARIP-M	0.339	0.426	0.403	0.414	0.273	1.156	0.554	0.601	0.574	<b>1.555</b>	0.346	0.48	0.457	0.467	<b>0.253</b>
BVARIP-SV-M	0.232	0.366	0.359	0.371	0.56	0.146	0.168	0.231	0.21	0.497	0.652	0.599	0.572	0.585	0.628
BVARIP vs BVARIP-SV	1.901	1.393	1.401	1.351	2.038	0.816	1.047	1.058	1.059	0.006	0.661	0.479	0.47	0.469	1.099
	<b>South Carolina</b>														
BVARIP	0.576	0.481	0.471	0.449	0.456	1.079	0.962	0.958	0.969	0.59	0.442	0.463	0.454	0.447	0.490
BVARIP-SV	0.26	<b>0.115</b>	<b>0.106</b>	<b>0.104</b>	0.179	0.3	0.407	0.428	0.446	0.86	0.652	0.528	0.505	0.495	0.373
BVARSSVS	0.649	0.34	0.323	0.299	0.472	0.654	0.806	0.828	0.851	<b>0.480</b>	0.559	0.407	0.388	0.376	0.599
BVARSSVS-SV	0.519	0.38	0.361	0.355	0.549	-0.609	-0.506	-0.486	-0.468	-0.039	1.606	1.324	1.267	1.243	0.937
BVARHS	0.701	0.458	0.439	0.414	0.729	0.336	0.444	0.464	0.484	0.246	0.708	0.574	0.55	0.537	0.808
BVARHS-SV	0.182	0.13	0.116	0.112	0.193	0.183	0.285	0.306	0.325	0.721	0.723	0.595	0.569	0.557	0.426
BVARIP-M	0.713	0.327	0.324	0.312	0.417	0.957	<b>1.160</b>	<b>1.186</b>	<b>1.237</b>	0.568	0.513	<b>0.333</b>	<b>0.325</b>	<b>0.309</b>	0.442
BVARIP-SV-M	0.38	0.156	0.157	0.159	<b>0.137</b>	0.041	0.291	0.275	0.278	0.866	0.849	0.526	0.536	0.536	<b>0.365</b>
BVARIP vs BVARIP-SV	2.215	4.188	4.45	4.306	2.553	0.779	0.555	0.53	0.522	-0.269	0.678	0.877	0.899	0.903	1.313
	<b>South Dakota</b>														
BVARIP	0.175	0.193	0.183	0.162	0.128	1.624	<b>1.642</b>	<b>1.651</b>	<b>1.666</b>	<b>4.162</b>	0.178	<b>0.182</b>	<b>0.179</b>	<b>0.170</b>	<b>0.113</b>
BVARIP-SV	0.158	0.154	0.153	0.138	0.144	0.705	0.778	0.774	0.785	3.618	0.413	0.368	0.371	0.357	0.169
BVARSSVS	0.323	0.234	0.236	0.213	0.209	1.371	1.448	1.445	1.457	3.388	0.252	0.221	0.223	0.214	0.221
BVARSSVS-SV	0.187	0.143	0.138	0.124	0.166	0.179	0.256	0.252	0.263	2.954	0.693	0.610	0.615	0.591	0.301
BVARHS	0.184	<b>0.110</b>	<b>0.111</b>	<b>0.103</b>	<b>0.107</b>	0.893	0.973	0.969	0.98	2.983	0.345	0.300	0.303	0.291	0.284
BVARHS-SV	0.205	0.165	0.165	0.149	0.143	0.885	0.959	0.955	0.966	3.604	0.349	0.31	0.313	0.300	0.169
BVARIP-M	0.205	0.332	0.393	0.372	0.152	1.585	1.157	1.078	1.133	3.883	0.195	0.324	0.362	0.348	0.122
BVARIP-SV-M	0.170	0.189	0.217	0.208	0.129	0.448	0.733	0.592	0.635	3.358	0.531	0.37	0.430	0.417	0.166
BVARIP vs BVARIP-SV	1.108	1.254	1.197	1.172	0.889	0.918	0.864	0.877	0.881	0.545	0.432	0.496	0.481	0.476	0.667

TABLE 2.7: Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by

Model vs bench	Tilting using new information														
	Point Forecasting - MRSFE					Density Forecasting - logS					Density Forecasting - CRPS				
	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4
<b>Tennessee</b>															
BVARIP	0.412	0.466	0.477	0.459	0.819	0.938	<b>0.999</b>	<b>0.987</b>	<b>1.000</b>	-0.652	0.418	0.442	0.442	0.439	0.927
BVARIP-SV	0.175	0.204	0.195	0.19	0.245	0.393	0.197	0.206	0.213	-0.02	0.481	0.654	0.637	0.637	0.68
BVARSSVS	0.328	0.239	0.224	0.214	0.641	0.79	0.666	0.678	0.685	-0.38	0.38	<b>0.425</b>	<b>0.411</b>	<b>0.410</b>	1.034
BVARSSVS-SV	0.189	0.204	0.184	0.17	0.391	-0.402	-0.601	-0.592	-0.584	-0.854	1.045	1.431	1.395	1.395	1.546
BVARHS	0.304	0.309	0.302	0.288	0.747	0.523	0.345	0.355	0.362	-0.526	0.452	0.582	0.566	0.565	1.197
BVARHS-SV	0.117	<b>0.183</b>	<b>0.160</b>	<b>0.152</b>	0.255	0.311	0.11	0.12	0.127	-0.068	0.513	0.707	0.687	0.687	0.713
BVARIP-M	0.512	0.521	0.553	0.525	0.623	0.637	0.274	0.277	0.41	-0.352	0.523	0.584	0.629	0.586	0.742
BVARIP-SV-M	0.229	0.199	0.22	0.22	<b>0.237</b>	0.167	0.175	0.103	0.104	<b>0.090</b>	0.628	0.589	0.655	0.655	<b>0.621</b>
BVARIP vs BVARIP-SV	2.361	2.278	2.44	2.422	3.348	0.545	0.802	0.781	0.786	-0.632	0.869	0.676	0.694	0.689	1.362
<b>Texas</b>															
BVARIP	0.362	0.369	0.308	0.284	0.266	1.14	1.196	1.235	1.253	<b>1.856</b>	0.357	0.323	<b>0.294</b>	<b>0.282</b>	<b>0.255</b>
BVARIP-SV	0.264	0.358	0.292	0.263	0.273	0.425	0.214	0.209	0.22	1.254	0.518	0.709	0.712	0.693	0.369
BVARSSVS	0.427	0.571	0.607	0.624	0.576	0.762	0.598	0.586	0.585	0.774	0.454	0.555	0.573	0.576	0.653
BVARSSVS-SV	0.244	0.342	0.292	0.262	0.31	-0.243	-0.461	-0.471	-0.462	0.494	0.968	1.355	1.381	1.348	0.723
BVARHS	0.352	0.343	0.35	0.343	0.548	0.499	0.315	0.304	0.313	0.56	0.506	0.64	0.654	0.64	0.742
BVARHS-SV	0.088	<b>0.228</b>	<b>0.196</b>	<b>0.177</b>	<b>0.143</b>	0.491	0.268	0.258	0.268	1.186	0.46	0.656	0.667	0.651	0.360
BVARIP-M	0.403	0.358	0.373	0.325	0.305	1.076	<b>1.229</b>	<b>1.188</b>	<b>1.281</b>	1.435	0.395	<b>0.310</b>	0.334	0.295	0.309
BVARIP-SV-M	0.372	0.402	0.389	0.332	0.348	0.201	0.129	0.153	0.199	0.948	0.658	0.695	0.684	0.628	0.448
BVARIP vs BVARIP-SV	1.368	1.03	1.054	1.082	0.971	0.715	0.982	1.026	1.033	0.602	0.689	0.456	0.413	0.407	0.691
<b>Utah</b>															
BVARIP	0.491	0.453	0.448	0.448	0.705	0.978	1.159	1.166	1.17	0.085	0.447	0.356	0.354	0.348	0.759
BVARIP-SV	0.2	0.229	0.233	0.217	0.324	0.127	0.05	0.052	0.049	-0.005	0.703	0.828	0.828	0.827	0.703
BVARSSVS	0.661	0.873	0.862	0.884	1.005	0.437	0.383	0.388	0.384	-0.434	0.655	0.734	0.729	0.729	1.229
BVARSSVS-SV	0.344	0.465	0.466	0.47	0.651	-0.664	-0.743	-0.741	-0.744	-0.846	1.54	1.826	1.826	1.826	1.614
BVARHS	0.807	1.049	1.06	1.079	1.54	0.166	0.111	0.111	0.108	-0.678	0.829	0.932	0.937	0.934	1.688
BVARHS-SV	0.135	<b>0.162</b>	<b>0.167</b>	<b>0.174</b>	0.35	0.069	-0.009	-0.007	-0.012	-0.089	0.735	0.871	0.871	0.871	0.764
BVARIP-M	0.572	0.3	0.314	0.304	0.525	0.821	<b>1.374</b>	<b>1.299</b>	<b>1.330</b>	<b>0.313</b>	0.524	<b>0.293</b>	<b>0.314</b>	<b>0.298</b>	<b>0.591</b>
BVARIP-SV-M	0.3	0.211	0.216	0.213	<b>0.298</b>	-0.054	0.533	0.457	0.467	0.104	0.855	0.5	0.533	0.515	0.651
BVARIP vs BVARIP-SV	2.46	1.984	1.923	2.062	2.172	0.851	1.109	1.114	1.121	0.09	0.637	0.43	0.427	0.421	1.081
<b>Vermont</b>															
BVARIP	0.237	0.211	0.226	0.224	0.251	1.323	<b>1.302</b>	<b>1.289</b>	<b>1.289</b>	<b>1.537</b>	0.251	<b>0.252</b>	<b>0.257</b>	<b>0.257</b>	0.231
BVARIP-SV	0.137	0.196	0.207	0.21	0.199	0.449	0.363	0.357	0.357	0.781	0.486	0.571	0.573	0.573	0.377
BVARSSVS	0.184	0.198	0.188	0.18	0.273	0.887	0.81	0.808	0.809	0.447	0.328	0.376	0.372	0.371	0.526
BVARSSVS-SV	0.158	0.246	0.246	0.252	0.249	-0.366	-0.451	-0.456	-0.456	-0.067	1.085	1.271	1.27	1.271	0.845
BVARHS	0.215	0.239	0.241	0.246	0.412	0.431	0.35	0.345	0.345	0.098	0.502	0.582	0.582	0.583	0.748
BVARHS-SV	0.141	<b>0.172</b>	<b>0.176</b>	<b>0.176</b>	0.222	0.415	0.332	0.327	0.327	0.71	0.5	0.584	0.584	0.584	0.405
BVARIP-M	0.358	0.447	0.312	0.215	0.208	1.153	1.148	1.208	1.345	1.572	0.347	0.355	0.303	0.234	<b>0.216</b>
BVARIP-SV-M	0.12	0.443	0.354	0.25	<b>0.165</b>	0.144	-0.001	0.055	0.192	0.527	0.669	0.905	0.781	0.606	0.453
BVARIP vs BVARIP-SV	1.729	1.074	1.091	1.071	1.258	0.875	0.939	0.931	0.932	0.756	0.516	0.441	0.449	0.448	0.612
<b>Virginia</b>															
BVARIP	0.15	0.255	0.218	0.233	0.525	1.542	<b>1.144</b>	<b>1.179</b>	<b>1.163</b>	0.425	0.186	<b>0.289</b>	<b>0.270</b>	<b>0.275</b>	0.542
BVARIP-SV	0.139	0.305	0.312	0.331	0.385	0.135	-0.264	-0.258	-0.267	0.015	0.646	1.007	1.008	1.01	0.699
BVARSSVS	0.316	0.345	0.353	0.36	0.472	0.503	0.136	0.142	0.133	-0.306	0.489	0.694	0.696	0.695	0.951
BVARSSVS-SV	0.128	0.378	0.354	0.375	0.47	-0.833	-1.229	-1.222	-1.231	-0.906	1.684	2.604	2.603	2.604	1.675
BVARHS	0.342	0.48	0.48	0.497	0.785	0.262	-0.123	-0.116	-0.126	-0.443	0.604	0.905	0.904	0.905	1.148
BVARHS-SV	0.09	<b>0.117</b>	<b>0.098</b>	<b>0.107</b>	0.17	0.019	-0.375	-0.368	-0.377	-0.056	0.718	1.104	1.103	1.104	0.712
BVARIP-M	0.199	0.509	0.529	0.49	0.324	1.393	0.814	0.783	0.824	<b>1.223</b>	0.23	0.41	0.45	0.419	<b>0.319</b>
BVARIP-SV-M	0.199	0.505	0.617	0.57	0.417	-0.196	-0.36	-0.456	-0.417	0.074	0.892	1.077	1.235	1.135	0.726
BVARIP vs BVARIP-SV	1.082	0.838	0.696	0.704	1.364	1.407	1.408	1.436	1.43	0.41	0.289	0.288	0.268	0.272	0.774
<b>Washington</b>															
BVARIP	0.341	0.243	0.232	0.224	0.327	1.16	1.385	1.417	1.426	1.295	0.329	0.244	0.235	0.23	0.305
BVARIP-SV	0.228	0.211	0.201	0.206	0.35	0.335	0.496	0.515	0.51	0.605	0.554	0.458	0.45	0.451	0.46
BVARSSVS	0.506	0.282	0.274	0.27	0.552	0.728	1.02	1.042	1.041	0.353	0.505	0.32	0.313	0.311	0.629
BVARSSVS-SV	0.239	0.294	0.295	0.311	0.442	-0.664	-0.501	-0.483	-0.487	-0.4	1.464	1.204	1.187	1.189	1.144
BVARHS	0.294	0.145	0.136	0.129	0.27	0.51	0.698	0.718	0.715	0.157	0.485	0.369	0.362	0.361	0.657
BVARHS-SV	0.105	<b>0.095</b>	<b>0.089</b>	<b>0.091</b>	0.186	0.392	0.556	0.575	0.571	0.611	0.51	0.417	0.41	0.41	0.42
BVARIP-M	0.39	0.171	0.153	0.145	<b>0.175</b>	1.09	<b>1.713</b>	<b>1.730</b>	<b>1.730</b>	<b>1.889</b>	0.367	<b>0.172</b>	<b>0.164</b>	<b>0.158</b>	<b>0.170</b>
BVARIP-SV-M	0.233	0.23	0.231	0.225	0.243	0.141	0.52	0.486	0.469	0.747	0.66	0.446	0.462	0.46	0.421
BVARIP vs BVARIP-SV	1.491	1.153	1.15	1.086	0.932	0.825	0.89	0.902	0.916	0.689	0.593	0.532	0.523	0.509	0.664
<b>West Virginia</b>															
BVARIP	0.673	0.907	0.913	0.953	0.748	0.562	0.454	0.471	0.47	-0.387	0.65	0.762	0.772	0.77	0.843
BVARIP-SV	0.456	0.563	0.563	0.588	0.478	0.004	-0.076	-0.08	-0.085	0.426	0.82	0.935	0.961	0.96	0.594
BVARSSVS	0.724	0.897	0.919	0.964	0.887	0.498	0.458	0.451	0.444	0.206	0.686	0.744	0.768	0.77	0.894
BVARSSVS-SV	0.371	<b>0.509</b>	<b>0.516</b>	<b>0.543</b>	0.461	-0.785	-0.868	-0.873	-0.878	-0.372	1.719	1.982	2.043	2.043	1.163
BVARHS	0.52	0.61	0.616	0.644	0.725	0.233	0.165	0.16	0.155	0.004	0.691	0.768	0.79	0.789	0.906
BVARHS-SV	0.415	0.543	0.558	0.583	0.554	0.185	0.104	0.098	0.093	0.381	0.69	0.792	0.818	0.818	0.643
BVARIP-M	0.396	0.588	0.608	0.568	<b>0.422</b>	1.081	<b>0.739</b>	<b>0.727</b>	<b>0.769</b>	<b>0.890</b>	0.366	<b>0.534</b>	<b>0.550</b>	<b>0.510</b>	<b>0.437</b>
BVARIP-SV-M	0.368	0.577	0.603	0.564	0.519	-0.134	-0.125	-0.138	-0.118	0.299	0.915	0.935	0.964	0.91	0.683
BVARIP vs BVARIP-SV	1.476	1.61	1.621	1.62	1.563	0.559	0.529	0.551	0.555	-0.812	0.793	0.815	0.803	0.802	1.418

TABLE 2.8: Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model.

Model vs bench	Tilting using new information														
	Point Forecasting - MRSFE					Density Forecasting - logS					Density Forecasting - CRPS				
	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4	UF	Q1	Q2	Q3	Q4
	<b>Wisconsin</b>														
BVARIP	0.514	0.730	0.794	0.765	0.644	0.611	<b>0.850</b>	<b>0.850</b>	<b>0.860</b>	-0.724	0.538	0.569	0.588	0.582	0.781
BVARIP-SV	0.199	0.327	0.327	0.329	0.233	0.225	-0.068	-0.074	-0.073	0.281	0.608	0.927	0.959	0.96	0.567
BVARSSVS	0.390	0.516	0.558	0.535	0.484	0.826	0.599	0.593	0.596	0.172	0.411	<b>0.532</b>	<b>0.550</b>	<b>0.548</b>	0.683
BVARSSVS-SV	0.167	0.440	0.468	0.496	0.407	-0.715	-1.010	-1.017	-1.017	-0.674	1.528	2.352	2.438	2.441	1.447
BVARHS	0.491	<b>0.842</b>	0.914	0.881	0.953	0.495	0.220	0.214	0.216	-0.073	0.549	0.801	0.829	0.826	1.010
BVARHS-SV	0.155	<b>0.245</b>	<b>0.251</b>	<b>0.251</b>	0.241	0.161	-0.133	-0.139	-0.138	0.234	0.640	0.978	1.013	1.013	0.593
BVARIP-M	0.613	0.689	0.806	0.798	0.522	0.230	0.629	0.608	0.626	-0.429	0.655	0.615	0.660	0.652	0.611
BVARIP-SV-M	0.185	0.593	0.724	0.715	<b>0.163</b>	-0.02	-0.241	-0.272	-0.271	<b>0.474</b>	0.781	1.079	1.167	1.165	<b>0.499</b>
BVARIP vs BVARIP-SV	2.586	2.233	2.424	2.326	2.763	0.386	0.918	0.924	0.933	-1.005	0.885	0.614	0.613	0.607	1.378
	<b>Wyoming</b>														
BVARIP	0.242	0.185	0.182	0.180	0.328	1.348	1.522	1.514	1.531	1.965	0.253	0.205	0.204	0.202	0.338
BVARIP-SV	0.239	0.178	0.179	0.179	0.337	0.606	0.788	0.771	0.786	1.931	0.445	0.356	0.360	0.358	0.346
BVARSSVS	0.266	0.189	0.192	0.190	0.352	1.388	1.587	1.570	1.587	1.821	0.262	<b>0.203</b>	<b>0.205</b>	<b>0.203</b>	0.374
BVARSSVS-SV	0.281	0.207	0.208	0.207	0.389	0.318	0.501	0.484	0.499	1.445	0.587	0.468	0.474	0.470	0.487
BVARHS	0.258	0.190	0.190	0.189	0.359	1.112	1.296	1.281	1.296	1.561	0.301	0.239	0.241	0.239	0.436
BVARHS-SV	0.174	<b>0.133</b>	<b>0.134</b>	<b>0.133</b>	0.293	0.932	1.113	1.097	1.112	1.880	0.321	0.257	0.260	0.259	0.332
BVARIP-M	0.240	0.208	0.215	0.193	<b>0.165</b>	1.394	<b>1.704</b>	<b>1.634</b>	<b>1.764</b>	<b>2.918</b>	0.242	0.224	0.230	0.211	0.161
BVARIP-SV-M	0.291	0.197	0.201	0.197	0.233	0.389	0.809	0.739	0.843	2.236	0.551	0.412	0.422	0.405	<b>0.261</b>
BVARIP vs BVARIP-SV	1.014	1.041	1.016	1.005	0.975	0.741	0.734	0.743	0.744	0.034	0.568	0.577	0.566	0.565	0.977

TABLE 2.9: Nowcasting Performance (Results Relative to AR Benchmark). NOTE: the RMSFE and CRPS values from our nowcasting model are presented relative to (divided by) those from the benchmark AR model; the logS values from our nowcasting model are presented relative to (subtracted by) those from the benchmark AR model.

FIGURE 2.1: Comparison of point forecast accuracy. Each panel describes a US State with quarterly and annual data. The x axis reports the RMSFE for BVAR, the y axis reports the RMSFE for BVAR-SV. Each point corresponds to a different forecast horizon.

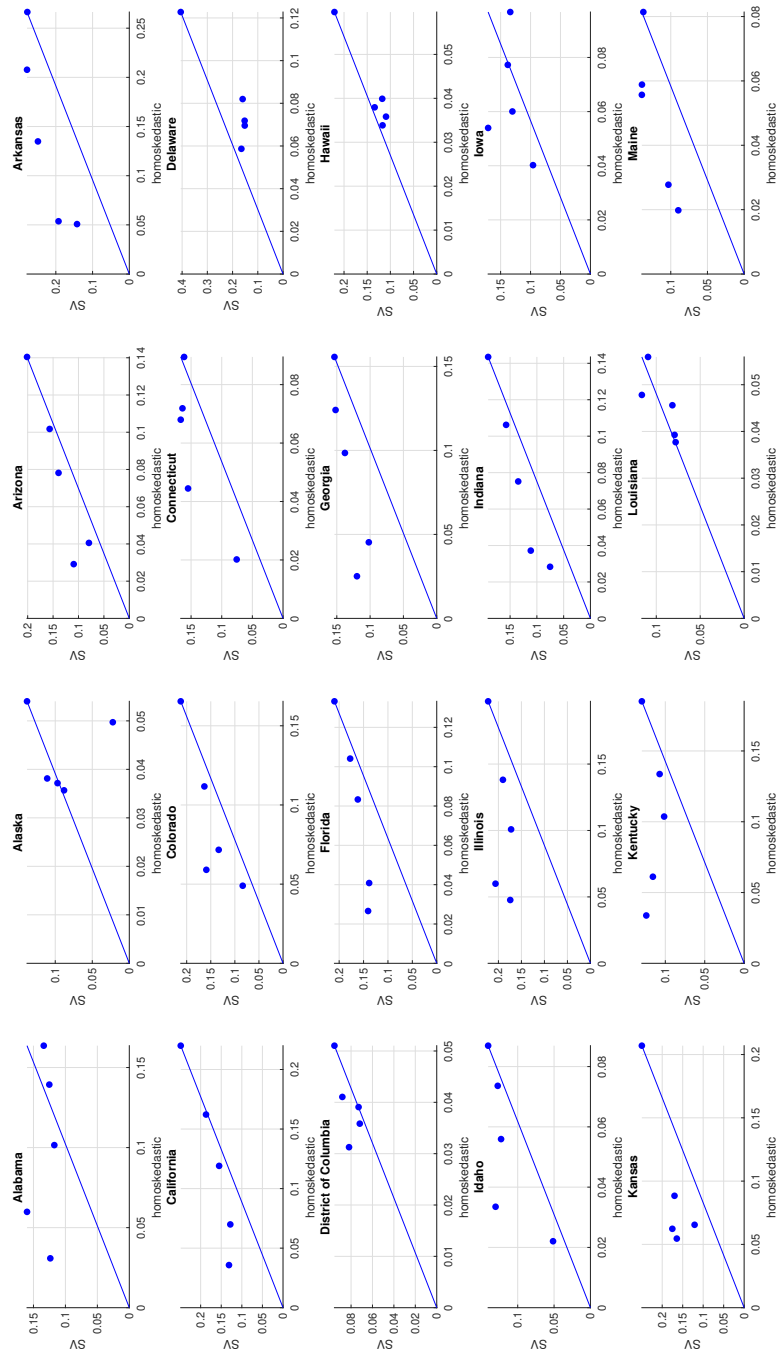


FIGURE 2.2: Comparison of point forecast accuracy. Each panel describes a US State with quarterly and annual data. The x axis reports the RMSFE for BVAR, the y axis reports the RMSFE for BVAR-SV. Each point corresponds to a different forecast horizon.

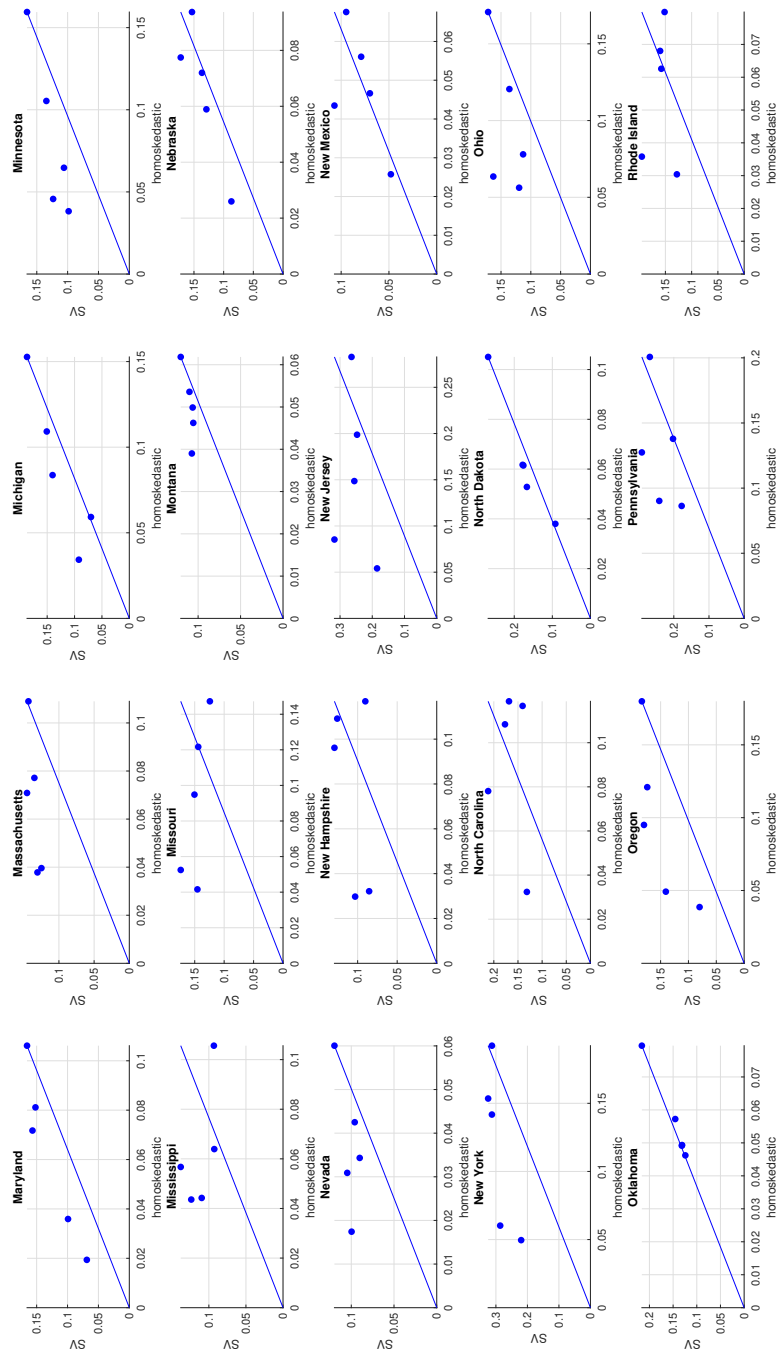
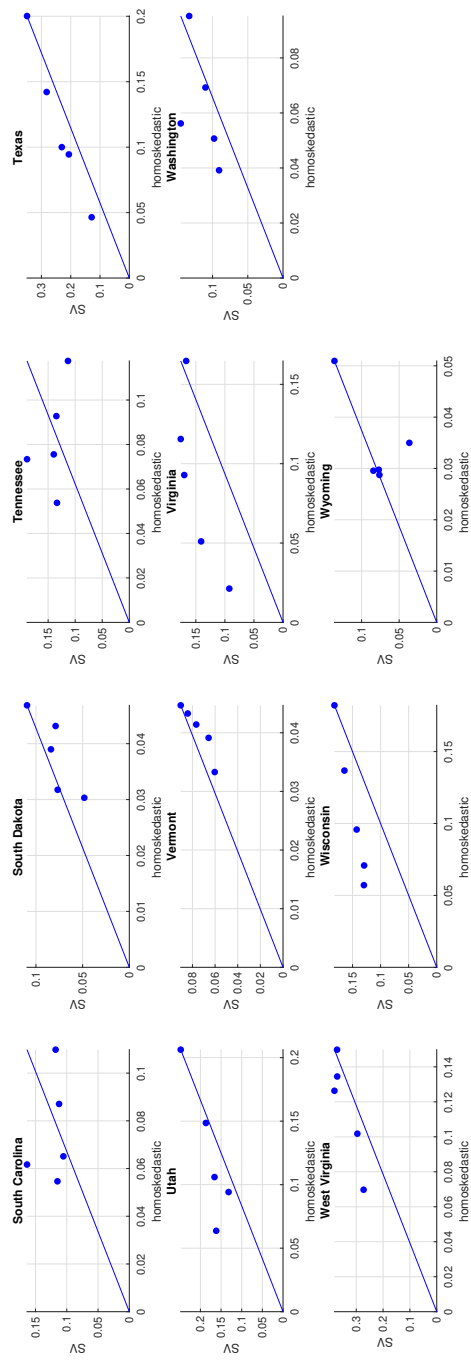


FIGURE 2.3: Comparison of point forecast accuracy. Each panel describes a US State with quarterly and annual data. The x axis reports the RMSFE for BVAR, the y axis reports the RMSFE for BVAR-SV. Each point corresponds to a different forecast horizon.



## Chapter 3

# Nowcasting Monthly GDP Growth for U.S. States Using a Mixed Frequency VAR

### 3.1 Introduction

The goal of this Chapter is to produce monthly nowcasts and historical estimates of GDP growth at the state-level using a mixed frequency vector autoregression (MF-VAR) in the U.S.. MF-VARs have enjoyed great popularity in policy circles since they can provide timely, high frequency nowcasts of low frequency variables such as GDP growth which is released with a delay. For instance, [Schorfheide and Song \(2015\)](#), [Brave, Butters, and Justiniano \(2019\)](#), and [McCracken, Owyang, and Sekhposyan \(2019\)](#) are influential papers associated with the Federal Reserve Banks of Minneapolis, Chicago, and St. Louis, respectively. A common set-up is to nowcast a quarterly variable (e.g. U.S. GDP growth) using several monthly variables. Nowcasting U.S. state-level GDP growth is a challenge, since there are 51 variables<sup>1</sup> to be nowcasted and the frequency mismatch is more complicated since we have a three-way frequency mismatch involving annual, quarterly and monthly variables which changes over time. Specifically, state-level GDP growth is available at an annual frequency only through 2004, before the U.S. Bureau of Economic Analysis (BEA)

---

<sup>1</sup>These are for the 50 states plus Washington DC. For simplicity, we will call the latter a state and refer to 51 states in this Chapter.



started producing quarterly estimates from 2005. U.S. GDP growth is available quarterly throughout the sample, and we include many monthly indicator variables in the model to better capture within-quarter business cycle dynamics. To complicate the data landscape further, the variables have a range of release delays and are available over different historical sample periods.

In light of these data features, in this Chapter we work with MF-VARs that are much larger than is conventional, and involve many more, and more complicated, latent states. This modelling approach raises challenges in terms of over-parameterization concerns and the computational burden. We develop a Bayesian modelling framework which overcomes these challenges, and use it to produce monthly estimates of state-level GDP. Notice that our estimates are “consistent” with the BEA’s estimates of both quarterly state and U.S.-level GDP. By imposing temporal and cross-sectional aggregation constraints within our MF-VAR that link published data with our model-based estimates and following [Koop, McIntyre, Mitchell, and Poon \(2022\)](#) consistency is achieved.

Our monthly state-level estimates of GDP can be contrasted with the state-level coincident indicators of [Crone and Clayton-Matthews \(2005\)](#), as published by the Federal Reserve Bank of Philadelphia, and the weekly state-level indicators of [Baumeister, Leiva-León, and Sims \(2021\)](#). While our estimates condition on the same state-level indicators as [Crone and Clayton-Matthews \(2005\)](#), as well as a range of additional indicators, their interpretation differs in that they are produced as direct estimates of state-level GDP. This also means that our estimates can be evaluated ex post, on receipt of the BEA’s quarterly estimates of state-level GDP. Econometrically, in contrast to both [Crone and Clayton-Matthews \(2005\)](#) and [Baumeister, Leiva-León, and Sims \(2021\)](#), in order to impose the cross-sectional constraint where state-level GDP sums to U.S. GDP, we estimate our model jointly across the 51 states. This implies that our model accommodates cross-state spillovers and dependencies, enabling it to provide a richer characterization of the higher-frequency effects of shocks on the U.S. states.

The remainder of the Chapter is organised as follows. Section [3.2](#) introduces the mixed frequency VAR and the proposed efficient approximate MCMC algorithm. Section [3.3](#) contains the empirical application results. Conclusions are set out in Section [3.4](#).

## 3.2 Econometric Methods

### 3.2.1 Notation and Data Observability

All our variables enter our model in growth rates using exact growth and are denoted by  $y$ . We use the following notational convention: subscripts  $a, q, m$  denote annual, quarterly and monthly frequencies, respectively. The first subscript on any variable denotes the frequency at which a variable is observed.<sup>2</sup> The second subscript denotes a different frequency than the observed one (e.g. prior to 2005 GDP growth in a state was observed at the annual frequency and so, to denote monthly state level GDP growth, we use subscripts  $a, m$ ). If the first two subscripts are the same then we suppress one of them (e.g. employment growth is observed at the monthly frequency and thus, we simply use subscript  $m$  for this concept instead of  $m, m$ ). The third subscript  $t = 1, \dots, T$  denotes time at the monthly frequency. Superscripts  $US$  and  $s = 1, \dots, S$  distinguish between variables for U.S. as a whole and the individual states. Our model involves the following variables:

- $y_{m,t}^{US}$  is a vector of macroeconomic variables for the U.S., which are always observed (e.g. employment growth).
- $y_{m,q,t}^{US}$  is a vector of quarterly observations on the U.S. macroeconomic variables. It is constructed from  $y_{m,t}^{US}$  and, thus, is always observed.
- $y_{q,m,t}^{US}$  is monthly GDP growth in the U.S. and is never observed.
- $y_{q,t}^{US}$  is quarterly GDP growth in the U.S.. It is observed for months 3, 6, 9 and 12, but not in other months.
- $y_{a,m,t}^s$  is monthly GDP growth in state  $s$ . It is never observed.
- $y_{a,q,t}^s$  is quarterly GDP growth in state  $s$ . Prior to 2005 it is never observed. From 2005 onwards it is observed for months 3, 6, 9 and 12, but not in other months.
- $y_{a,t}^s$  is annual GDP growth in state  $s$ . Prior to 2005 it is observed for month 12 of every year. From 2005 onwards it is observed for months 3, 6, 9 and 12, but not in other months.

---

<sup>2</sup>We denote state level variables with an  $a$  even though they are observed quarterly from 2005 onwards.

If the  $t$  subscript is suppressed, it denotes the vector of all observations on a variable. Superscript  $S$  denotes the vector containing quantities for all 51 states.

The monthly macroeconomic variables at the U.S. level include industrial production, personal consumption expenditure, hours worked, the unemployment rate, inflation, the S&P500, the effective Federal Funds rate, ten-year treasury rate, real personal income, and the employment rate. These are variables considered by [Schorfheide and Song \(2015\)](#). We also consider quarterly fixed investment and government expenditure. In addition to these U.S.-level variables, we have several monthly and quarterly predictors available at the state level. These are real personal income, wage and salaries, initial claims, the unemployment rate, the employment rate, and average weekly hours (for all sectors). In order to keep our large MF-VAR from becoming too large, we include these as exogenous predictors in the MF-VAR, with no cross-state spillovers. That is, a monthly exogenous variable for state  $i$  will appear as a right-hand side variable in the equation for state  $i$ , but not in the equation for state  $j$  for  $i \neq j$ . The exact definition, data sources, transformation and release schedule for each variable is given in the Data Appendix [A.1](#).

### 3.2.2 The MF-VAR

Similar to Chapter [2](#), we write the MF-VAR as

$$Ay_t = B_0 + B_1y_{t-1} + \dots + B_p y_{t-p} + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma), \quad (3.1)$$

for  $t = 1, \dots, T$  where  $y_t$  is a vector of  $N \times 1$  dependent variables,  $p$  is the order of the VAR,  $A$  is an  $N \times N$  lower triangular matrix with ones on the diagonal,  $B_q$ ,  $q = 1, \dots, p$  is the  $N \times N$  coefficient matrix and  $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$ <sup>34</sup>.

We re-write our model as:

$$y_{i,t} = w_{i,t}\alpha_i + x_{i,t}\beta_i + \epsilon_{i,t}, \quad (3.2)$$

where  $\epsilon_{i,t} \sim N(0, \sigma_i^2)$ ,  $w_{i,t} = (-y_{i,t}, \dots, -y_{i-1,t})'$ ,  $x_{i,t} = (1, y_{t-1}, \dots, y_{t-p})'$ ,  $\beta_i = (B_{0,i}, B_{1,i}, \dots, B_{p,i})'$  and  $a_i$  is the  $i^{\text{th}}$  row of  $A$  and is equal with  $\alpha_i =$

<sup>3</sup>Writing the MF-VAR with  $\Sigma$  being diagonal greatly reduces the computational burden since it allows for equation-by-equation estimation of the model (see, e.g., in [Carriero, Clark, and Marcellino \(2019\)](#))

<sup>4</sup>For simplicity, this specification does not include intercepts, more lags nor the state level exogenous variables. In our empirical work we include all of these and set the lag length equal to 2.

$(A_{i,1}, A_{i,2}, \dots, A_{i,i-1})'$ ,  $B_{0,i}$  denotes the  $i^{\text{th}}$  element of  $B_0$ ,  $B_{j,i}$  denotes the  $i^{\text{th}}$  row of  $B_j$ .

The model can be re-written as:

$$y_{i,t} = X_{i,t}\theta_i + \varepsilon_{i,t} \quad (3.3)$$

where  $\varepsilon_{i,t} \sim N(0, \sigma_i^2)$ ,  $X_{i,t} = (w_{i,t}, x_{i,t})$  and  $\theta_i = (\alpha'_i, \beta'_i)'$  and the dimension of  $\theta_i$  is  $k_i = Np + i$  for  $i = 1, \dots, N$ .

Since most of the values of  $y_t$  are not observed and treated as latent states, the MF-VAR is a state space model, where eq. (3.1) provides us with the state equations. The measurement equations in the state space model specify the observability conditions for every variable and link them to the unobserved latent states via inter-temporal restrictions. In our model we have a three-way frequency mismatch involving variables which are observed at monthly, quarterly and annual frequencies. Different inter-temporal restrictions apply for the various frequency mismatches.

Remember that quarterly state level GDP growth is available after 2005. To avoid a non-linear measurement equation, we proxy the weights in the inter-temporal restriction by 1/3. The inter-temporal restriction linking this to its monthly state counterpart post-2005 can be shown to be:

$$y_{a,q,t}^s = \frac{1}{3}y_{a,m,t}^s + \frac{2}{3}y_{a,m,t-1}^s + y_{a,m,t-2}^s + \frac{2}{3}y_{a,m,t-3}^s + \frac{1}{3}y_{a,m,t-4}^s \quad (3.4)$$

for  $s = 1, \dots, S$ . An inter-temporal restriction of the same form links monthly U.S. GDP growth to its observed quarterly value:

$$y_{q,t}^{US} = \frac{1}{3}y_{q,m,t}^{US} + \frac{2}{3}y_{q,m,t-1}^{US} + y_{q,m,t-2}^{US} + \frac{2}{3}y_{q,m,t-3}^{US} + \frac{1}{3}y_{q,m,t-4}^{US}. \quad (3.5)$$

Prior to 2005, state level GDP growth was only observed annually and the inter-temporal restriction linking the observed quantity to the desired monthly quantity

is:

$$\begin{aligned}
y_{a,t}^s = & \frac{1}{12}y_{a,m,t}^s + \frac{2}{12}y_{a,m,t-1}^s + \frac{3}{12}y_{a,m,t-2}^s + \frac{4}{12}y_{a,m,t-3}^s + \frac{5}{12}y_{a,m,t-4}^s + \frac{6}{12}y_{a,m,t-5}^s \\
& + \frac{7}{12}y_{a,m,t-6}^s + \frac{8}{12}y_{a,m,t-7}^s + \frac{9}{12}y_{a,m,t-8}^s + \frac{10}{12}y_{a,m,t-9}^s + \frac{11}{12}y_{a,m,t-10}^s + \frac{12}{12}y_{a,m,t-11}^s \\
& + \frac{11}{12}y_{a,m,t-12}^s + \frac{10}{12}y_{a,m,t-13}^s + \frac{9}{12}y_{a,m,t-14}^s + \frac{8}{12}y_{a,m,t-15}^s + \frac{7}{12}y_{a,m,t-16}^s + \frac{6}{12}y_{a,m,t-17}^s \\
& + \frac{5}{12}y_{a,m,t-18}^s + \frac{4}{12}y_{a,m,t-19}^s + \frac{3}{12}y_{a,m,t-20}^s + \frac{2}{12}y_{a,m,t-21}^s + \frac{1}{12}y_{a,m,t-22}^s, \tag{3.6}
\end{aligned}$$

for  $s = 1, \dots, S$ . Note that [Koop, McIntyre, Mitchell, and Poon \(2022\)](#) are adding errors to all of the inter-temporal restrictions, since they use the exact growth rates that involves an approximation. In this case, we proxy the weights in the inter-temporal restriction by  $1/12$ .

A final measurement equation is obtained through a cross-sectional restriction which arises from the fact that U.S. GDP is the sum of the GDP of all states. For exact growth rate data, this can be show to be (see [Koop, McIntyre, Mitchell, and Poon \(2020b\)](#)):

$$y_{q,t}^{US} = \sum_{s=1}^S w_t^s y_{a,q,t}^s + \eta_t, \quad w_t^s = \frac{y_{a,q,t-1}^s}{\sum_{s=1}^S y_{a,q,t-1}^s}, \tag{3.7}$$

where  $w_t^s$  is the share of regional output in aggregate output at quarter  $t$  and  $\eta_t \sim \mathcal{N}(0, cs^2)$ . We proxy  $w_t^s$  by the observed annual shares, noting that we expect to see little within-year variation in these weights. In the measurement equation we add an error since the cross-sectional restriction involves an approximation, since state-level GDP need not sum to U.S. GDP. We implement this using a method developed in [Doran \(1992\)](#) that preserves the linear structure of the state space. This measurement equation only applies prior to 2005. We find that the inclusion of this method improves estimation precision; and of course the restriction ensures that our new quarterly regional data are consistent with the observed quarterly U.S. totals.

In summary, our MF-VAR is a state space model with state equations given by eq. (3.1) and measurement equations given by eq. (3.4) – (3.7). The first of the measurement equations applies from 2005 onwards and the last two apply up to 2005. Given a prior (see below), standard Bayesian MCMC methods for state space models exist for carrying out posterior and predictive analysis with this MF-VAR. See, for instance, [Schorfheide and Song \(2015\)](#) for a conventional MF-VAR, [Koop, McIntyre, Mitchell,](#)

and Poon (2020b) for a MF-VAR with a regional structure similar to that in this Chapter and Koop, McIntyre, Mitchell, and Poon (2020a) for a case where the frequency mismatch changes over time.

### 3.2.3 The Prior

The MF-VARs estimated in this Chapter are relative high-dimensional, including more than 50 covariates/states. When working with VARs of this dimension, there is a strong need for prior shrinkage and many alternatives have been proposed in the literature including forms inspired by the classic Minnesota prior such as Banbura, Giannone, and Reichlin (2010), variable selection priors proposed by Korobilis (2013) and global-local shrinkage priors developed in Kastner and Huber (2020). With MF-VARs the need for prior shrinkage becomes even more important due to the need to additionally estimate the high frequency values of the variables which are only observed at a low frequency. In this Chapter, we adopt the global-local hierarchical shrinkage priors. Specifically, we adopt the horseshoe prior proposed by Carvalho, Polson, and Scott (2010) as a method for shrinkage in the presence of sparsity.

The horseshoe prior belongs to the family of global-local shrinkage priors. We implement the horseshoe prior one equation at a time. That is, each equation in the MF-VAR has its own prior, which allows for a different degree of shrinkage in each equation. For equation  $i$ , we have a global shrinkage prior parameter,  $\tau_i$ , and a local shrinkage parameter,  $\lambda_{ij}$ , which is specific to the  $j^{\text{th}}$  coefficient. The horseshoe prior has properties which are often found advantageous in sparse models. Specifically, it aggressively penalises small coefficients, but applies minimal shrinkage to large coefficients. Thus, the noise provided by large numbers of irrelevant coefficients in the MF-VAR is largely removed and the signal provided by the few non-zero coefficients is more precisely estimated in a data-based fashion.

A conventional horseshoe prior, for the  $i^{\text{th}}$  VAR equation,  $\theta_i$   $i = 1, \dots, N$ , is the following:

$$\theta_i | \Lambda_i^2, \tau_i^2, \sigma_i^2 \sim N(0, \Lambda_i^2 \tau_i^2 \sigma_i^2), \quad \Lambda_i^2 = \text{diag}(\lambda_{i,1}^2, \dots, \lambda_{i,k_i}^2) \quad (3.8)$$

$$\sigma_i^2 \sim \text{IG}(\underline{v}_\sigma, \underline{S}_\sigma), \quad (3.9)$$

$$\tau_i^2 \sim C^+(0, 1), \quad (3.10)$$

$$\lambda_{i,j}^2 \sim C^+(0, 1), \quad j = 1, \dots, k_i. \quad (3.11)$$

We assume that  $\nu_\sigma = 10$  and  $\bar{S}_\sigma = 0.04$ . These values imply that prior means of  $\sigma^2$  is 0.001 which is roughly consistent with the magnitude of our growth rates data, but relatively non-informative.

Despite its advantages, the horseshoe specification results in a non-standard conditional distribution, thus making standard Gibbs sampler intractable. There are various specifications of the horseshoe prior and associated MCMC sampling schemes. [Neal \(2003\)](#) and [Polson, Scott, and Windle \(2014\)](#) proposed the use of a slice sampler that allows  $\tau_i^2$  and  $\lambda_{i,j}^2$  to update efficiently in high-dimensions which we follow.

The above hierarchy makes Gibbs sampling from the posterior distribution straightforward:

1. The conditional posterior distribution of the regression coefficients  $\theta_i$  is drawn as follows:

$$\theta_i | \Lambda_i^2, \tau_i^2, \sigma_i^2 \sim N \left( A_i^{-1} X_i' y_i, \sigma_i^2 A_i^{-1} \right), \quad (3.12)$$

where  $A_i = (X_i' X_i + \Lambda_\star^{-1})$ ,  $\Lambda_\star = \tau_i^2 \Lambda_i^2$ . We use the efficient algorithm of [Bhattacharya, Chakraborty, and Mallick \(2016\)](#) to sample them.

2. The conditional posterior distribution of  $\sigma_i^2$  is drawn as follows:

$$\sigma_i^2 \sim IG(\bar{\nu}_\sigma, \bar{S}_\sigma), \quad (3.13)$$

where  $\bar{\nu}_\sigma = (T + K_i)/2$  and  $\bar{S}_\sigma = (y_i - X_i \theta_i)' (y_i - X_i \theta_i) / 2 + \theta_i' \Lambda_\star^{-1} \theta_i / 2$ .

3. The conditional posterior distribution of  $\tau_i^2$  is drawn as follows:

- set  $\eta_i = 1/\tau_i^2$
- sample  $u$  from

$$u | \eta_i \sim \mathcal{U} \left( 0, \frac{1}{1 + \eta_i} \right), \quad (3.14)$$

- sample  $\eta_i$  from

$$\eta_i \sim \gamma \left( (k_i + T)/2, u \frac{2\sigma_i^2}{\sum (\frac{\theta_i}{\lambda_{i,j}})^2} \right), \quad (3.15)$$

where  $\gamma$  is the lower incomplete gamma function

- set  $\tau_i = \frac{1}{\sqrt{\eta_i}}$ .

4. The conditional posterior distribution of  $\lambda_{i,j}^2$  is drawn as follows:

- set  $\eta_{i,j} = 1/\lambda_{i,j}^2$
- sample  $u$  from

$$u|\eta_{i,j} \sim \mathcal{U}\left(0, \frac{1}{1 + \eta_{i,j}}\right), \quad (3.16)$$

- sample  $\eta_{i,j}$  from

$$\eta_{i,j} \sim e^{\frac{\theta_i^2}{2\sigma_i^2}\eta_{i,j}} I\left(\frac{u}{1-u} > \eta_{i,j}\right) \quad (3.17)$$

- set  $\lambda_{i,j} = \frac{1}{\sqrt{\eta_{i,j}}}$ .

The reader is referred to [Bhattacharya, Chakraborty, and Mallick \(2016\)](#) for precise details about the efficient sampler for the Normal distribution, which makes full use of the Woodbury identity in order to sample  $\theta$  efficiently.

### 3.2.4 State Space setup

Suppose  $y_t$  is a  $N$ -dimensional series, in which not all of its variables will be observed at every period under a mixed frequency setup. We consider  $y_t = (y_{m,t}^{US}, y_{q,t}^{US}, y_a^S)'$ , where  $N_m$  collects the monthly U.S. variables,  $y_{m,t}^{US}$ , such as inflation,  $N_q$  collects the quarterly U.S. and state variables post-2005,  $y_{q,t}^{US}, y_{q,t}^S$ , respectively, that are observed every quarter and finally  $N_a$  collects the annual GDP by state, which is observed every year pre-2005. It holds that  $N = N_m + N_q + N_a$ .

To describe the monthly and quarterly dynamics, we denote  $y_{q,m,t}, y_{a,m,t}, y_{a,q,t}$  as the monthly and quarterly latent variables underlying the quarterly series and annual series,  $y_{q,t}^{US}, y_{q,t}^S$  and  $y_a^S$ , respectively. As we describe in Subsection 3.2.5 we run our MCMC in two steps; the former considers the annual-quarterly mismatch, while the latter depicts the monthly – the quarterly frequency. We combine these latent variables with the indicators observed at the quarterly and monthly frequency in  $X_t^k = [X_t^{HF}, X_t^{LF}]$ ,  $k = 1, 2$ <sup>5</sup>. Clearly, when  $k = 1$ , HF is the quarterly data and LF is the annual variables, while when  $k = 2$ , HF is the monthly indicator and LF is the quarterly series.

To alleviate the computational burden of the state space MF-VAR we follow [Schorfheide and Song \(2015\)](#) and [Ankargren and Jonéus \(2019\)](#). This suggests that monthly variables can be omitted from the state equation and instead enter the system through the

<sup>5</sup>1 corresponds to quarterly indicators, while 2 depicts the monthly latent variables and HF stands for high frequency data and LF for low frequency



exogenous terms of the state-space model. However, this implies that the state and measurement errors will be correlated, which requires the use of an algorithm based on a state-space model with between-equation correlation. The companion form of the monthly VAR specified in eq. (3.3) together with a measurement equation for  $y_t$  delivers the common two-equation state-space system given by

$$y_t = C_t + Z_t S_t + G_t \epsilon_t \quad (3.18)$$

$$S_t = D_t + T_t S_{t-1} + H_t \epsilon_t \quad (3.19)$$

$$\epsilon_t \sim N(0, I_N), \quad (3.20)$$

where

$$Z_t = \begin{pmatrix} 0_{N_{HF} \times N_{LF}} & \Phi_{HF,LF} \\ \Lambda_{LF} & 0_{N_{LF} \times N_{LF}} \end{pmatrix}, \quad T_t = \begin{pmatrix} \Phi_{LF,LF} & 0_{N_{LF} \times N_{LF}} \\ I_{pN_{LF}} & 0_{pN_{LF} \times N_{LF}} \end{pmatrix}, \quad (3.21)$$

$$C_t = \begin{pmatrix} \Phi_{HF,HF} & \Phi_{HF,0} \\ 0_{N_{LF} \times N_{HF}} & 0_{N_{LF} \times 1} \end{pmatrix} \begin{pmatrix} y_{m,t-1:t-p} \\ 1 \end{pmatrix} + \begin{pmatrix} \beta X_{HF,t-1:t-p} \\ 0_{N_{LF} \times 1} \end{pmatrix}, \quad (3.22)$$

$$D_t = \begin{pmatrix} \Phi_{LF,HF} & \Phi_{LF,0} \\ 0_{pN_{LF} \times pN_{HF}} & 0_{pN_{LF} \times 1} \end{pmatrix} \begin{pmatrix} y_{m,t-1:t-p} \\ 1 \end{pmatrix} + \begin{pmatrix} \beta X_{HF,t-1:t-p} \\ 0_{N_{LF} \times 1} \end{pmatrix}, \quad (3.23)$$

$$G_t = \begin{pmatrix} \Sigma_{HF}^{1/2} \\ 0_{N_{LF} \times N} \end{pmatrix}, \quad H_t = \begin{pmatrix} \Sigma_{LF}^{1/2} \\ 0_{pN_{LF} \times N} \end{pmatrix}, \quad \Sigma_t^{1/2} = \begin{pmatrix} \Sigma_{HF}^{1/2} \\ \Sigma_{LF}^{1/2} \end{pmatrix}. \quad (3.24)$$

We present the companion form of the VAR at the high frequency as:

$$\begin{pmatrix} X_t^k \\ X_{t-1}^k \\ \vdots \\ X_{t-p+1}^k \end{pmatrix} = \begin{pmatrix} \Phi_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I_N & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_N & 0 \end{pmatrix} \begin{pmatrix} X_{t-1}^k \\ X_{t-2}^k \\ \vdots \\ X_{t-p}^k \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (3.25)$$

$$S_t = \begin{pmatrix} X_t^k \\ X_{t-1}^k \\ \vdots \\ X_{t-p+1}^k \end{pmatrix}, \quad \Phi_0 = \begin{pmatrix} \Phi_{HF,0} \\ \Phi_{LF,0} \end{pmatrix}, \quad \Phi_i = \begin{pmatrix} \Phi_{HF,HF} & \Phi_{HF,LF} \\ \Phi_{LF,HF} & \Phi_{LF,LF} \end{pmatrix}. \quad (3.26)$$

Finally, in eq. (3.21) we augment  $Z_t$  with the cross sectional restriction. The term  $\beta X_{HF,t-1:t-p}$  collects the monthly state level indicators.

We use algorithm 2 of [Durbin and Koopman \(2002\)](#) and generate draws for the posterior distribution  $P(S|\Theta, Y)$  as follows:

1. Generate an artificial sample from eq. (3.18), (3.19);  $\{s_t^+, y_t^+\}_{t=1}^T$
2. Set  $y_t^* = y_t - y_t^+$  and use Kalman smoothing to get  $\hat{s}_t^*$
3. Take a draw  $\{s_t^+, +\hat{s}_t^*\}_{t=1}^T$  from  $P(S|\Theta, Y)$ .

Finally when smoothing  $y_t^*$ , we set the constant terms equal to zero as in [Jarociński \(2015\)](#).

### 3.2.5 A Computationally Efficient Approximate MCMC Algorithm

Bayesian inference in MF-VARs is implemented using MCMC methods. For the conventional MF-VAR with a single frequency mismatch and no cross-sectional restrictions, the algorithm of [Schorfheide and Song \(2015\)](#) is commonly used. [Koop, McIntyre, Mitchell, and Poon \(2020b\)](#) extend this algorithm to a MF-VAR with a cross-sectional restriction. [Koop, McIntyre, Mitchell, and Poon \(2020a\)](#) further extend the algorithm for the case where the frequency mismatch changes over time. Small adaptations of these algorithms are required to handle the three-way frequency mismatch involved in our U.S. state-level application. We have done such adaptations, but have found that the MCMC algorithm is simply too slow<sup>6</sup> to carry out extensive empirical work or a pseudo-real time forecasting exercise. The fact that MCMC methods are not scalable to models with high-dimensional parameter spaces is well-known to Bayesian econometricians. This lack of scalability is a particular problem in large VARs and has led to the use of approximate methods. For instance, [Gefang, Koop, and Poon \(2020\)](#) develop Variational Bayesian (VB) methods for the MF-VAR. VB methods are computationally efficient but are approximations. It is well-known that, although they provide accurate approximations to posterior means, they tend to under-estimate posterior variances. In a forecasting exercise this leads to an under-estimation of predictive variances, see the discussion in [Gefang, Koop, and Poon \(2020\)](#). In this Chapter, we propose an approximate method specifically designed for the MF-VAR with a three-way frequency mismatch which does not rely on VB methods or similar. It is

<sup>6</sup>The empirical work in this Chapter would take months on a high quality personal computer.

MCMC-based and allows for full exploration of the high-dimensional posterior and accurate reflection of posterior and predictive uncertainty.

MCMC algorithms for drawing the parameters of an MF-VAR (conditional on draws of latent states for the low frequency variables) are standard. Conditional on the latent states, the model reduces to a VAR, for which Bayesian methods are available. For instance, Gibbs samplers for Bayesian VARs with several global-local shrinkage including the horseshoe are given in [Gefang, Koop, and Poon \(2020\)](#). Hence, we will not describe them here nor explicitly list the VAR parameters as conditioning arguments in this Subsection. Instead, we will develop methods for drawing monthly GDP growth (conditional on draws of the parameters). As noted above, the proposed model is a linear state space model and standard methods exist for drawing latent states in such models. However, the exact MCMC algorithm is very computationally burdensome in large dimensional MF-VARs (e.g.  $N > 50$ ) such as the ones used in this Chapter. In practice, we find that the main computational bottleneck lies in the parts of the model involving the annual-monthly frequency mismatch and, in particular, the fact that in the inter-temporal restriction, given in eq. (3.6), which involves over 20 lags. Accordingly, we develop an approximate MCMC algorithm which avoids the use of the annual-monthly inter-temporal restriction. The idea underlying our algorithm is that it is much faster to draw from two separate algorithms, one involving a quarterly-monthly frequency mismatch and the other a quarterly-annual mismatch. Below we describe the MCMC algorithm.

Consider an MCMC algorithm which produces draws of state level monthly and quarterly GDP growth rates given U.S. data and annual state level data. The posterior in this case is  $p(y_{a,m}^S, y_{a,q}^S | y_a^S, y_q^{US}, y_m^{US})$ . A simple rule of probability implies:

$$p(y_{a,m}^S, y_{a,q}^S | y_a^S, y_q^{US}, y_m^{US}) = p(y_{a,m}^S | y_{a,q}^S, y_a^S, y_q^{US}, y_m^{US}) p(y_{a,q}^S | y_a^S, y_q^{US}, y_m^{US}) \quad (3.27)$$

The first of the term on the right-hand side of eq. (3.27) can be simplified to  $p(y_{a,m}^S | y_{a,q}^S, y_q^{US}, y_m^{US})$  since, conditional on knowing the quarterly growth rates ( $y_{a,q}^S$ ), the annual growth rates  $y_a^S$  provides no additional information. This is the posterior that arises in a conventional MF-VAR involving a quarterly/monthly frequency mismatch, where annual quantities are not considered. The MCMC algorithm for such a posterior is standard (see, e.g., [Schorfheide and Song \(2015\)](#)).

The second term on the right-hand side of eq. (3.27) involves quarterly, monthly and annual variables and thus requires an annual-monthly inter-temporal restriction. Consider the posterior  $p(y_{a,q}^S | y_a^S, y_q^{US}, y_{m,q}^{US})$ , which is identical to  $p(y_{a,m}^S | y_{a,q}^S, y_q^{US}, y_m^{US})$  except for the fact that  $y_m^{US}$  is replaced by  $y_{m,q}^{US}$  (i.e. monthly employment growth rates have been replaced by quarterly growth rates). The MCMC algorithm for this posterior is also a conventional one for an MF-VAR involving an annual/quarterly frequency mismatch as in, e.g., [Koop, McIntyre, Mitchell, and Poon \(2020b\)](#).

The reasoning above suggests the following strategy: use two conventional MF-VARs, one with a quarterly/monthly frequency mismatch and another with an annual/quarterly frequency mismatch. Draws of  $y_{a,q}^S$  produced by the algorithm of the second MF-VAR are then conditioned on in the first MF-VAR model. We find that this strategy is much more computationally efficient than drawing directly from the algorithm with the three way frequency mismatch, since it avoids dealing with the computational bottleneck caused by having annual and monthly variables in the same model.

This strategy is an approximate one since the algorithm which produces quarterly draws of state level GDP growth is conditional on quarterly employment growth instead of monthly. But the loss of information is likely to be small (i.e. to produce quarterly estimates of state level GDP knowing the U.S. quarterly quantities will be useful, but knowing them at the monthly frequency is likely to provide only minimal improvements). We stress that this approximation is only used in one part of the algorithm. The important part of the algorithm is  $p(y_{a,m}^S | y_{a,q}^S, y_q^{US}, y_m^{US})$ , which produces the draws of monthly GDP growth for each state. This does condition on the monthly data and, thus, our monthly estimates of state level GDP growth do reflect the information contained in monthly predictors. If we further consider that this approximate algorithm is only used for producing pre-2005 draws of quarterly state level GDP growth (since subsequently quarterly state level data is available), the argument in favor of gaining large computational benefits by using this approximate algorithm is strengthened.

Evidence on the accuracy of this algorithm is provided in the Monte Carlo Section below.

### 3.2.5.1 Monte Carlo

In this Section we use Monte Carlo experiments in order to study the finite properties of our MF-BVAR using the traditional slow algorithm, henceforth Model 1 (i.e. monthly – quarterly – annual) and MF-BVAR using the efficient fast algorithm see eq. (3.27), henceforth Model 2 (i.e. first estimate the model at quarterly – annual frequency and get the historical estimates for the quarterly variable, then estimate the model at quarterly – monthly and produce the monthly variable). We generate artificial data  $\{y_t\}$  with three regressors measured at monthly, quarterly and annual frequency. For all experiments  $u_t \sim iidN(0, \sigma^2)$ , random variable independent of  $\{X_t\}$  and  $\{\sigma_t\}$ . We consider two Data Generating Processes (DGPs). The subscript denotes the frequency each variable is being generated, where  $a, q, m$  stand for annual, quarterly and monthly frequency, respectively.

We consider the following sample sizes,  $N_m \in (120, 360, 600, 840)$ ,  $N_q \in (40, 120, 200, 280)$  and  $N_a \in (10, 30, 50, 70)$  with 10 replications. Each Gibbs sampler runs for 20,000 draws from which we discard 18,000.

In DGP II, the sample is adjusted to the max number of observations we can use i.e.  $N_m \in (120, 360, 600)$ ,  $N_q \in (40, 120, 200)$  and  $N_a \in (10, 30, 50)$ .

To evaluate the performance of the computationally efficient MCMC, we present (i) the root mean squared error of Model 1, (ii) the root mean squared error of Model 2, (iii) the coverage ratio of lower bound 95% and (iv) the coverage ratio of upper bound 95%. The coverage depicts how many times the mean posterior estimates of the annual frequency variable of Model 2 lies inside the lower and upper 95<sup>th</sup> percentile of Model 1 posterior mean.

We set the lag order of VAR  $p = 2$  and assume the following:

$$\beta_{true} = \begin{bmatrix} 0.5 & 0.5 & 0.5 \\ 0.95 & 0 & 0 \\ 0 & 0.95 & 0 \\ 0 & 0 & 0.95 \\ -0.1 & 0 & 0 \\ 0 & -0.1 & 0 \\ 0 & 0 & -0.1 \end{bmatrix}, \Phi = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0 & 1 & 0.5 \\ 0 & 0 & 1 \end{bmatrix}, \sigma_{true} = (\Phi * \Phi')^{-1}$$

**DGP I:** We consider the case of  $N(0, 1)$  simulated predictors that are simulated as follows:

$$Y_t = \beta_{true} * Y_{t-p} + u_t, \quad u_t \sim iidN(0, \sigma_{true}^2),$$

where  $Y_t = [X_{N_m}, X_{N_q}, X_{N_a}]$  is a  $N_m \times 3$  matrix,  $X_{N_m} \sim N(0, 1)$ , is a  $N_m \times 1$  vector,  $X_{N_q} \sim N(0, 1)$ , is a  $N_q \times 1$  vector and  $X_{N_a} \sim N(0, 1)$ , is a  $N_a \times 1$  vector.

**DGP II:** We consider real data to simulate predictors as follows:

$$Y_t = \beta_{true} * Y_{t-p} + u_t, \quad u_t \sim iidN(0, \sigma_{true}^2),$$

where  $Y_t = [X_{N_m}, X_{N_q}, X_{N_a}]$  is a  $N_m \times 3$  matrix,  $X_{N_m}$  = U.S. employment, is a  $N_m \times 1$  vector,  $X_{N_q}$  = U.S. GDP, is a  $N_q \times 1$  vector and  $X_{N_a}$  = Alabama GDP, is a  $N_a \times 1$  vector.

TABLE 3.1: DGP I: X is i.i.d.

MF-BVAR						
$N_a$	$N_q$	$N_m$	$\widehat{MSE}$ Model 1	$\widehat{MSE}$ Model 2	Lower 95% coverage ratio	Upper 95% coverage ratio
10	40	120	0.0139	0.0136	99.7500	99.4900
30	120	360	0.0120	0.0120	100.0000	99.9700
50	200	600	0.0118	0.0117	100.0000	100.0000
70	280	840	0.0129	0.0129	100.0000	100.0000

TABLE 3.2: DGP II: X is Real data

MF-BVAR						
$N_a$	$N_q$	$N_m$	$\widehat{MSE}$ Model 1	$\widehat{MSE}$ Model 2	Lower 95% coverage ratio	Upper 95% coverage ratio
10	40	120	0.0094	0.0094	100.0000	100.0000
30	120	360	0.0094	0.0094	100.0000	99.9700
50	200	600	0.0086	0.0086	99.9800	100.0000

Tables 3.1 and 3.2 show that the computationally efficient approximate MCMC algorithm (Model 2) suggests results similar to the traditional slow MCMC algorithm (Model 1) in terms of  $\widehat{MSE}$ .

### 3.3 Empirical Application

#### 3.3.1 Data

Our underlying data are monthly, quarterly and annual U.S. macroeconomic variables, covering the period from January 1964 to December 2019, taken from Federal Reserve Economic Data (FRED), Bureau of Economic Analysis (BEA) and Bureau of Labor Statistics (BLS). Our dataset consists of the following monthly variables: civilian unemployment rate, CPI, industrial production index, personal consumption expenditure, effective Federal Funds Rate, ten-year treasury rate, S&P500, average weekly hours, real personal income, all employees (total non-farm), and oil prices. The following quarterly variables: real GDP, GDP by state (post 2005), fixed investment and government expenditures. The annual frequency variables are: the GDP by state (pre-2005). Further, we include the following state-level variables: real personal income, wage and salary, initial claims, all employees (total non-farm), average weekly hours and the unemployment rate. We relegate the rest of the information about the data to the Data Appendix [A.1](#).

#### 3.3.2 Historical Estimates of monthly regional growth

We estimate the proposed MF-VAR-SV model on the 2019 vintage data, without a ragged edge, to produce historical monthly estimates of real regional growth in the U.S.. The full set of historical estimates are relegated to the end of this Chapter. In the Appendix [B.1](#), Figures [3.1](#) – [3.10](#) present the real estimates alongside the U.S. growth rate from 1964 to 2019. In the next Section, we present how the historical estimates can enrich our understanding of the U.S. economy. Before doing this we discuss some features of the historical estimates. First, we inspect the 68% credible intervals around our regional real GDP growth estimates and find that our regional estimates are quite precise, as illustrated in Figure [3.1](#) – [3.10](#). Our credible intervals go to zero once per year pre-2005 and once every quarter post 2005. The reason why we observe this pattern is that we plot annualized quarterly and monthly estimates once a year and a quarter, which impose the inter-temporal restriction in eq. [\(3.4\)](#) - [\(3.6\)](#), which is equal to the actual observed annual regional growth rate.

Second, we repeat the analysis using the Minnesota prior as a further robustness check. We find that, the estimates of regional monthly GDP growth produced by this

prior are very similar to those produced using the Horseshoe prior. Results using the Minnesota prior are available upon request.

Third, we re-estimated the MF-VAR-SV using growth rates in log differences (instead of using the exact growth rate) and find results to show similar patterns. Results using the log differences growth rate are available upon request.

### 3.3.3 Measuring the connectedness of U.S. states

We use the connectedness measures developed in [Diebold and Yilmaz \(2014\)](#) to complement our historical estimates and examine the dynamic connections between the U.S. states and U.S. macroeconomic variables.

To compute the connectedness measures we consider the generalised variance decomposition following [Koop, Pesaran, and Potter \(1996\)](#) and [Pesaran and Shin \(1998\)](#) due to its invariance to the ordering of the variables in the VAR. Each draw from our approximate MCMC algorithm provides the parameters in eq. (3.1) and we use these to compute the variance decompositions. This provides us with draws of  $d_{nj}^h$ , that we average to produce estimates, for  $n, j = 1, \dots, N$ .

We now define the forecast error at the  $h^{\text{th}}$  horizon as

$$y_{N,t+h} - y_{N,t}(h) = \sum_{i=1}^{\infty} \Phi_i u_{t+h-i}, \quad (3.28)$$

where  $\Phi_i = [\Phi_{nj,i}]_{N \times N}$  is the response of variable  $n$  to a unit shock  $u_{it}$ ,  $j = 1 \dots, N$ ,  $i$  periods ago. The forecast error variance decomposition is computed at horizon  $h = H$  as

$$d_{nj}^H = \frac{\sigma^{-1} \sum_{h=0}^{H-1} (e_n' \Phi_h \Sigma_u e_j)^2}{\sum_{h=0}^{H-1} (e_n' \Phi_h \Sigma_u e_j)}, \quad (3.29)$$

where  $e_n$  is the column  $n^{\text{th}}$  of the  $I_n$  matrix. Due to the fact that forecast error variance decompositions do not sum up to 1, we use the following normalization

$$\widehat{d}_{nj}^H = \frac{d_{nj}^H}{\sum_{j=1}^N d_{nj}^H}. \quad (3.30)$$

It holds that  $\sum_{j=1}^N \widehat{d}_{nj}^H = 1$  and  $\sum_{n,j=1}^N \widehat{d}_{nj}^H = N$ . Using these variance decompositions we can define the total directional connectedness from other states to state  $n$  at horizon  $h$



as:

$$\text{Connectedness from: } \sum_{n \neq j} d_{nj}^h. \quad (3.31)$$

This is a measure of how information in other states impacts the forecast error variance of region  $n$  (i.e. the summation is over  $j$ ). Following [Diebold and Yilmaz \(2014\)](#) we call this measure connectedness from. Then we define the total directional connectedness to other regions from region  $j$  at horizon  $h$  is:

$$\text{Connectedness to: } \sum_{j \neq n} d_{nj}^h. \quad (3.32)$$

This is a measure of how information in state  $j$  impacts the forecast error variances of other states (i.e. the summation is over  $n$ ). This is called a connectedness to measure. We emphasize that our connectedness measures are based on a monthly frequency VAR. Thus, e.g., results for  $h = 1$  measure connectedness in terms of the one month ahead forecast error variances and results for  $h = 12$  measure connectedness in terms of the one year ahead forecast error variances. Since a key contribution of this Chapter is to produce monthly estimates of state level GDP growth, in this Section we limit our attention on the connectedness measures at  $h = 1$  and  $h = 12$ . For ease of exposition we split connectedness results into four blocks; the first three illustrate the connectedness from measures, while the last one depicts the connectedness to measures to other variables. Specifically, the first block consists of the connectedness from results for the U.S. macroeconomic variables, the second block depicts results for own state effects and the third one consists of the impact from other states. Finally, the fourth block shows the connectedness to measures.

Table 3.3 shows that in the short-run ( $h = 1$ ) the degree of connectedness is high and dominated by state-specific effects. This is seen by focusing on the own variable block that illustrates the importance of state-specific effects. Across states, we see that idiosyncratic or state-specific shocks explain up to 98% of short-run state growth dynamics. The connectedness from measure varies only slightly across states, but the connectedness to measures vary more. For these, Alabama, Illinois and Pennsylvania exhibit higher values than the other states indicating that these states have the strongest effects on the others.

Table 3.4 contains the connectedness measures for  $h = 12$  and shows a very different picture. It indicates that the inter-connections between states are much higher at this

longer forecast horizon. Across states, the idiosyncratic state-specific shocks typically<sup>7</sup> now explain less, and the highest effect appears in North Dakota at 83%. These findings show that state growth dynamics become much more important in explaining state growth dynamics in the longer run.

---

<sup>7</sup>Koop, McIntyre, Mitchell, and Poon (2020b) observe similar behavior across the UK regions

TABLE 3.3: Connectedness in percent (normalized to sum to 100) at  $h = 1$  months

State	Macroeconomic Variables	Own State	Other state	To other
Alabama	4.33	93.92	1.76	5.29
Alaska	1.38	97.84	0.79	1.71
Arizona	1.44	97.88	0.69	1.66
Arkansas	2.39	96.57	1.04	2.77
California	2.03	97.38	0.60	2.20
Colorado	2.72	95.94	1.35	3.34
Connecticut	3.09	95.33	1.59	3.86
Delaware	1.41	97.75	0.85	1.81
District of Columbia	2.56	96.09	1.36	3.19
Florida	1.99	96.98	1.03	2.39
Georgia	3.58	94.45	1.98	4.72
Hawaii	1.85	96.98	1.17	2.40
Idaho	2.66	95.73	1.61	3.38
Illinois	3.78	94.12	2.11	5.00
Indiana	3.69	94.37	1.96	4.72
Iowa	2.98	95.41	1.62	3.80
Kansas	2.44	96.14	1.43	3.17
Kentucky	1.95	96.97	1.08	2.41
Louisiana	2.41	96.34	1.26	3.05
Maine	2.41	96.43	1.17	2.88
Maryland	2.55	96.16	1.30	3.16
Massachusetts	1.23	98.45	0.32	1.30
Michigan	1.30	98.20	0.50	1.44
Minnesota	2.04	96.98	0.99	2.41
Mississippi	2.07	96.90	1.03	2.48
Missouri	2.05	97.00	0.95	2.43
Montana	1.75	97.51	0.75	2.00
Nebraska	0.87	98.83	0.30	0.94
Nevada	1.55	97.69	0.77	1.83
New Hampshire	1.43	97.85	0.73	1.69
New Jersey	2.41	96.45	1.15	2.91
New Mexico	2.16	96.79	1.06	2.59
New York	2.40	96.51	1.10	2.84
North Carolina	2.60	96.30	1.11	3.07
North Dakota	1.35	98.00	0.65	1.58
Ohio	3.04	95.65	1.32	3.67
Oklahoma	1.43	97.73	0.85	1.81
Oregon	1.33	97.85	0.82	1.68
Pennsylvania	3.78	94.41	1.82	4.97
Rhode Island	2.63	96.09	1.28	3.22
South Carolina	2.06	96.94	1.00	2.48
South Dakota	2.11	96.89	1.00	2.54
Tennessee	2.54	95.87	1.60	3.52
Texas	1.83	97.05	1.12	2.39
Utah	2.85	95.58	1.57	3.64
Vermont	1.14	98.52	0.34	1.22
Virginia	1.51	97.83	0.66	1.73
Washington	2.37	96.28	1.36	3.04
West Virginia	2.73	95.67	1.60	3.52
Wisconsin	2.62	95.91	1.47	3.30
Wyoming	1.44	97.77	0.79	1.72

TABLE 3.4: Connectedness in percent (normalized to sum to 100) at  $h = 12$  months

State	Macroeconomic Variables	Own State	Other states	To other
Alabama	23.49	58.11	18.56	16.03
Alaska	13.84	76.63	9.75	48.26
Arizona	14.91	71.80	13.47	24.78
Arkansas	19.57	65.57	15.05	15.69
California	16.73	68.93	14.86	12.28
Colorado	17.27	70.07	13.05	15.44
Connecticut	18.22	66.71	15.36	15.97
Delaware	14.57	75.05	10.58	32.31
District of Columbia	18.80	69.19	12.22	11.17
Florida	15.44	67.02	17.84	10.18
Georgia	21.63	56.35	22.22	14.89
Hawaii	15.04	76.49	8.63	13.19
Idaho	15.07	72.78	12.43	18.05
Illinois	19.58	57.56	23.01	13.60
Indiana	21.24	47.21	31.73	17.39
Iowa	16.97	60.11	23.12	17.01
Kansas	16.85	64.29	19.05	10.13
Kentucky	16.51	65.24	18.61	9.25
Louisiana	14.10	74.42	11.67	14.54
Maine	18.03	65.55	16.63	10.93
Maryland	18.43	65.62	16.14	10.04
Massachusetts	15.30	65.57	19.81	9.66
Michigan	15.68	73.26	11.26	147.47
Minnesota	17.90	57.00	25.37	8.11
Mississippi	19.82	62.35	18.04	13.57
Missouri	18.60	59.63	21.94	10.55
Montana	17.87	67.84	14.48	19.43
Nebraska	15.08	63.74	21.36	9.28
Nevada	13.49	75.76	10.95	15.30
New Hampshire	14.33	73.36	12.65	32.29
New Jersey	18.10	64.09	17.99	9.38
New Mexico	14.94	74.13	11.12	22.21
New York	17.94	69.22	13.29	10.05
North Carolina	19.33	61.11	19.73	9.97
North Dakota	10.03	82.21	7.99	86.88
Ohio	19.04	45.23	35.89	10.57
Oklahoma	13.45	73.72	13.09	14.68
Oregon	16.12	70.58	13.78	18.01
Pennsylvania	20.17	62.59	17.39	12.20
Rhode Island	17.69	66.94	15.68	14.93
South Carolina	18.56	63.24	18.37	9.16
South Dakota	15.24	64.73	20.25	31.30
Tennessee	23.42	56.87	19.90	14.34
Texas	15.19	66.07	18.93	8.56
Utah	16.54	68.82	14.93	15.48
Vermont	12.85	74.57	12.80	17.46
Virginia	18.79	65.76	15.70	6.71
Washington	15.71	73.57	10.89	13.87
West Virginia	15.50	72.49	12.20	13.13
Wisconsin	18.73	59.67	21.75	11.18
Wyoming	13.00	76.38	10.81	30.61

### **3.4 Conclusion**

Specifically, the proposed model imposes inter-temporal and cross-sectional constraints to ensure that the monthly estimates both “add up” to published quarterly/annual data and that the GDP estimates for the 50 states (plus DC) sum to published GDP data for the U.S. as a whole. We develop a computationally-fast approximate Bayesian Markov Chain Monte Carlo (MCMC) algorithm for estimating and nowcasting with this large scale MF-VAR. The model is used to produce historical estimates of monthly GDP for the 50 (plus DC) U.S. states back to the 1960s, and the utility of these estimates is illustrated by using them to better understand business cycle dynamics and cross-state dependencies. Further, we examine potential inter-connectedness among the U.S. states and U.S. macroeconomic variables.

## A.1 Data Appendix (Chapter 3)

This appendix provides a discussion regarding the construction of the real-time data set.

Our dataset consists of fourteen real-time macroeconomic U.S. nationwide data and eight real-time macroeconomic U.S. state-level variables, which are available from the ALFRED database maintained by the Federal Reserve Bank of St. Louis, from the Bureau of Economic Analysis (BEA), from the Bureau of Labor Statistics (BLS) and the Bloomberg database. Table 3.5 summarizes the series we use in this Chapter.

We use the same eleven real-time macroeconomic U.S. nationwide data as in the seminal work of [Schorfheide and Song \(2015\)](#) and add the following three: All Employees, Total non-farm, Personal Income and Crude Oil Prices. We added Personal income as [Arias, Gascon, and Rapach \(2016\)](#) and [Stock and Watson \(1989\)](#) are using it to estimate a monthly economic activity index for the United States.

Regarding the U.S. state level variables, normally, a range of economic indicators are used to capture the U.S. state economies. Among them, the most comprehensive measure of economic activity in a state, is the real GDP by state, but is available at annual frequency and with considerable publication lag, but the aim of this Chapter is to shed light on this measure by providing a more timely and frequent monthly GDP by state. As in [Crone and Clayton-Matthews \(2005\)](#) we add 3 monthly variables, namely Nonagricultural payroll employment, Unemployment rate, Average hours worked in manufacturing and one quarterly the real wage and salary disbursements. These variables have been commonly used in the literature of constructing state level coincident indexes, see [Stock and Watson \(1989\)](#) and [Crone and Clayton-Matthews \(2005\)](#). Also, the Federal Reserve Bank of Philadelphia is using these variables to produce a monthly coincident index for each state<sup>8</sup>.

Following [Carlino, DeFina, and Sill \(2001\)](#) and [Arias, Gascon, and Rapach \(2016\)](#), we also use industry employment data specifically average hours worked at the Metropolitan Statistical Areas (MSAs) level for seven industries: construction, manufacturing, transportation and public utilities (TPU), wholesale and retail trade (Trade), finance, insurance and real estate (FIRE), services, and government.

---

<sup>8</sup>State Coincident Indexes are available on <https://www.philadelphiafed.org/surveys-and-data/regional-economic-analysis/state-coincident-indexes>

As stated above, real GDP by state is available since 1977 only. For this reason we use the nominal GDP by state which is available since 1963 and the GDP deflator to produce real GDP by state from 1963 to 1976. [Del Negro \(2002\)](#) used U.S. GDP deflator and CPI by state to produce real GDP, he finds that both deflators produce similar results. For this reason we use the annual GDP by state (series ID: SAGDP2) and the U.S. GDP deflator (series ID: Gross Domestic Product: Implicit Price Deflator (A191RI1A225NBEA)). These two series IDs, namely SAGDP2 and A191RI1A225NBEA are not included in [Table 3.5](#). To produce real GDP by state we simply subtract the U.S. GDP from the growth rate of nominal GDP by state.

We seasonally adjust the non-seasonal adjusted series using the X-13 ARIMA-SEATS Seasonal Adjustment Program by the Census Bureau, in which we select the option for X11 and we do this recursively. For the data transformation we use the ones proposed by [McCracken and Ng \(2016\)](#), with the exception of U.S. GDP and GDP by state as we want to use exact growth rate.





## B.1 Appendix (Chapter 3) – Exact Growth Rates

This appendix illustrates the historical estimates for the U.S. growth and the U.S. growth by state.

FIGURE 3.1: Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC)

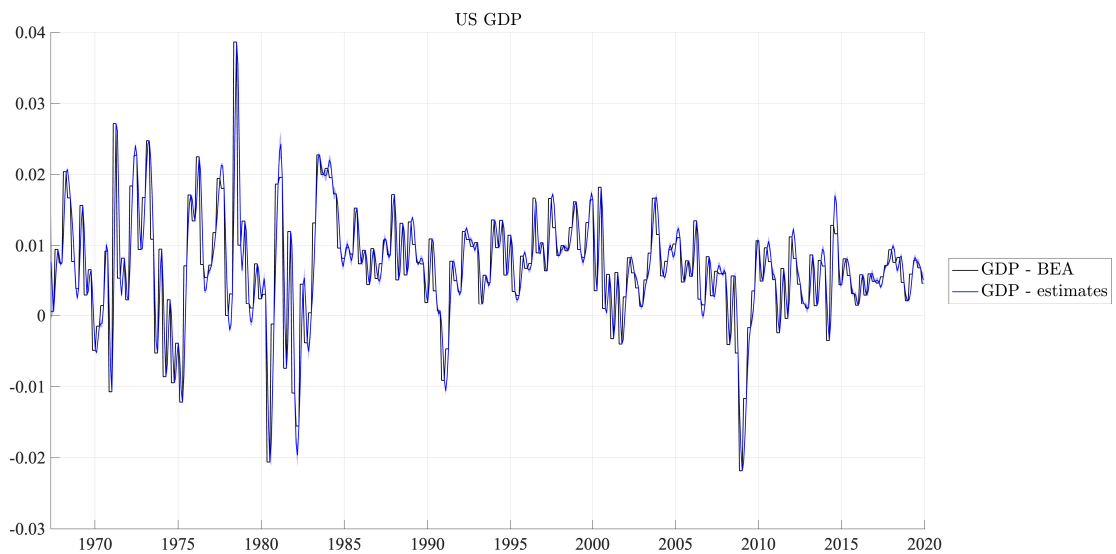


FIGURE 3.2: Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC)

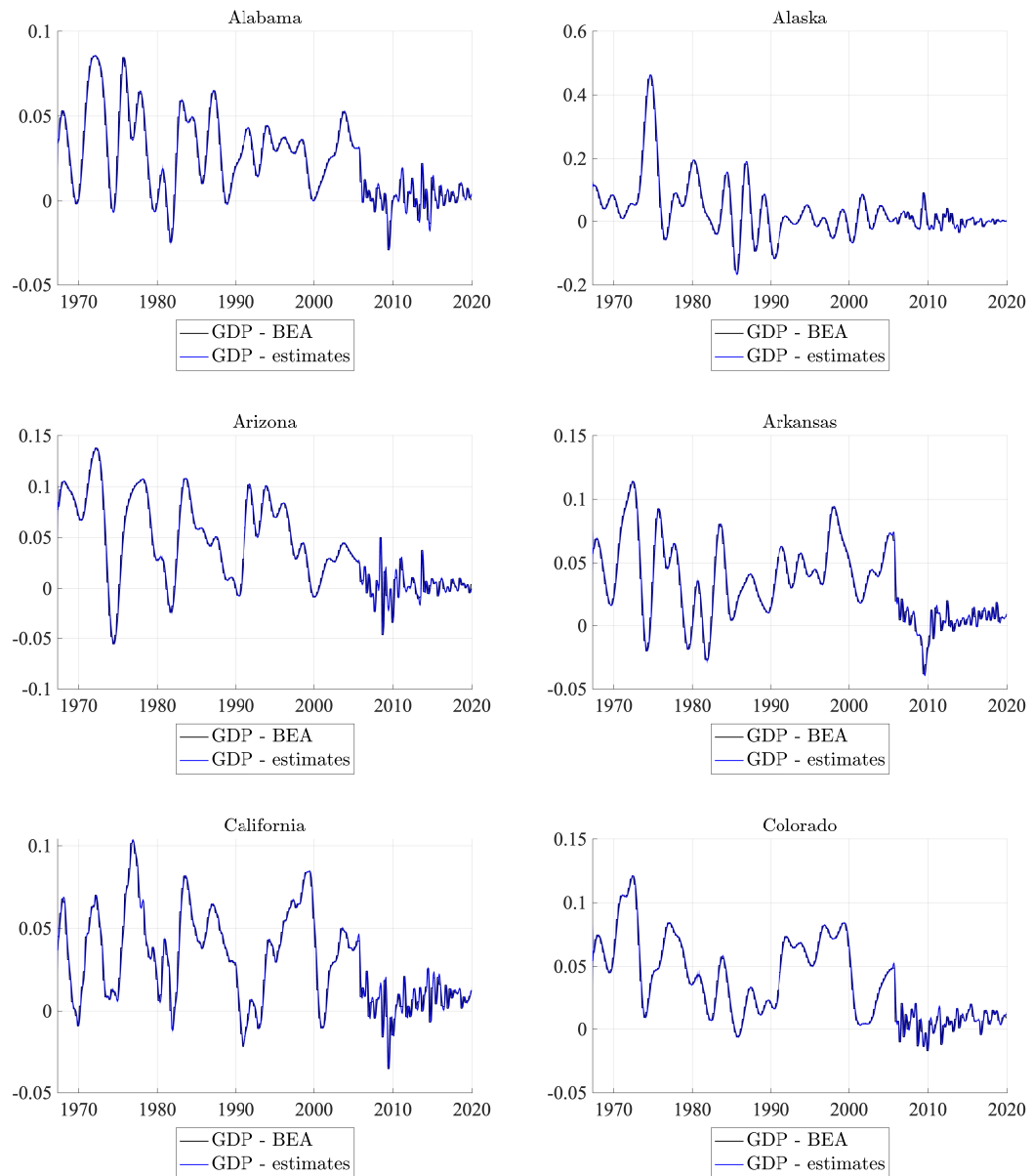


FIGURE 3.3: Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC)

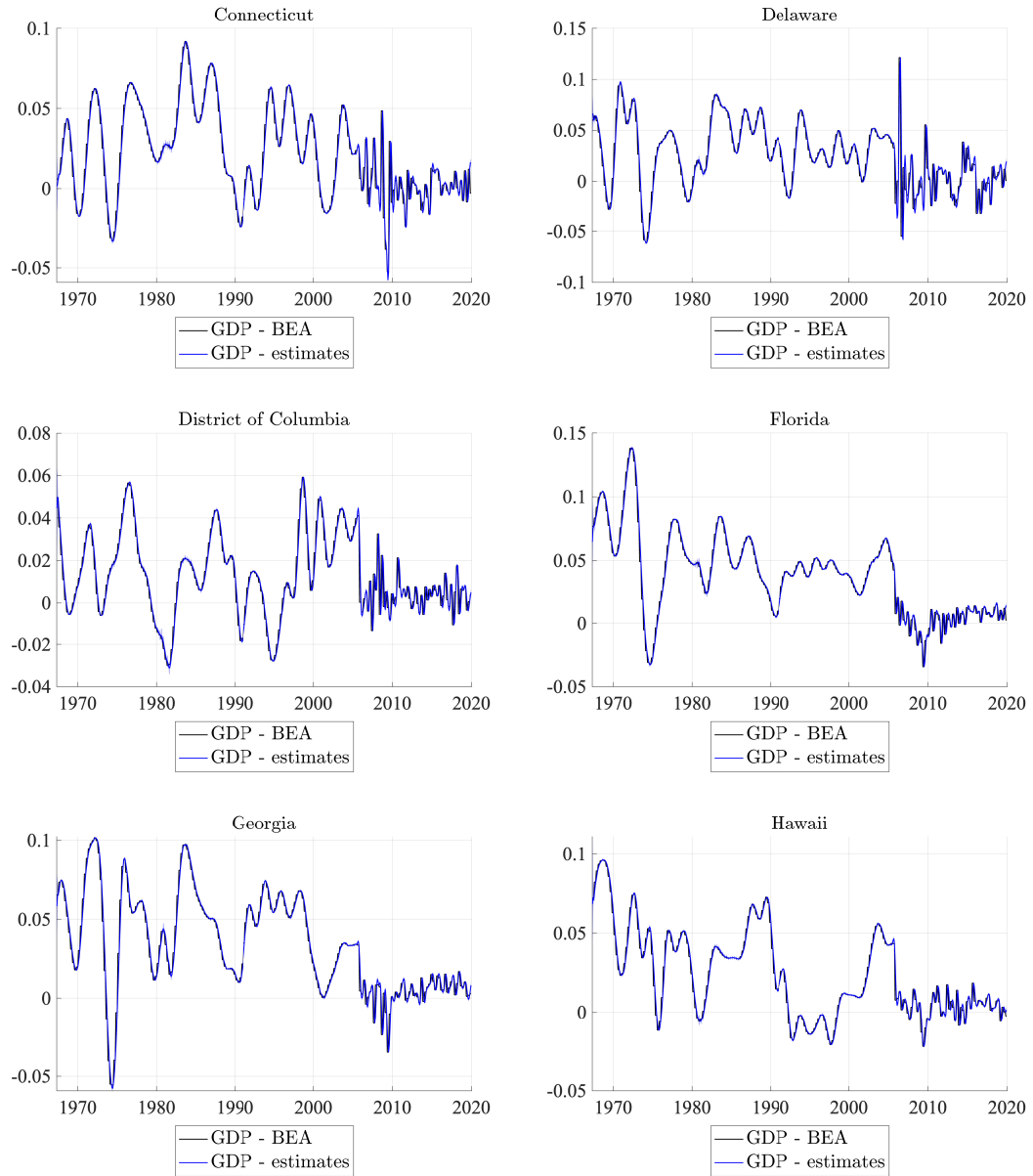


FIGURE 3.4: Historical Estimates

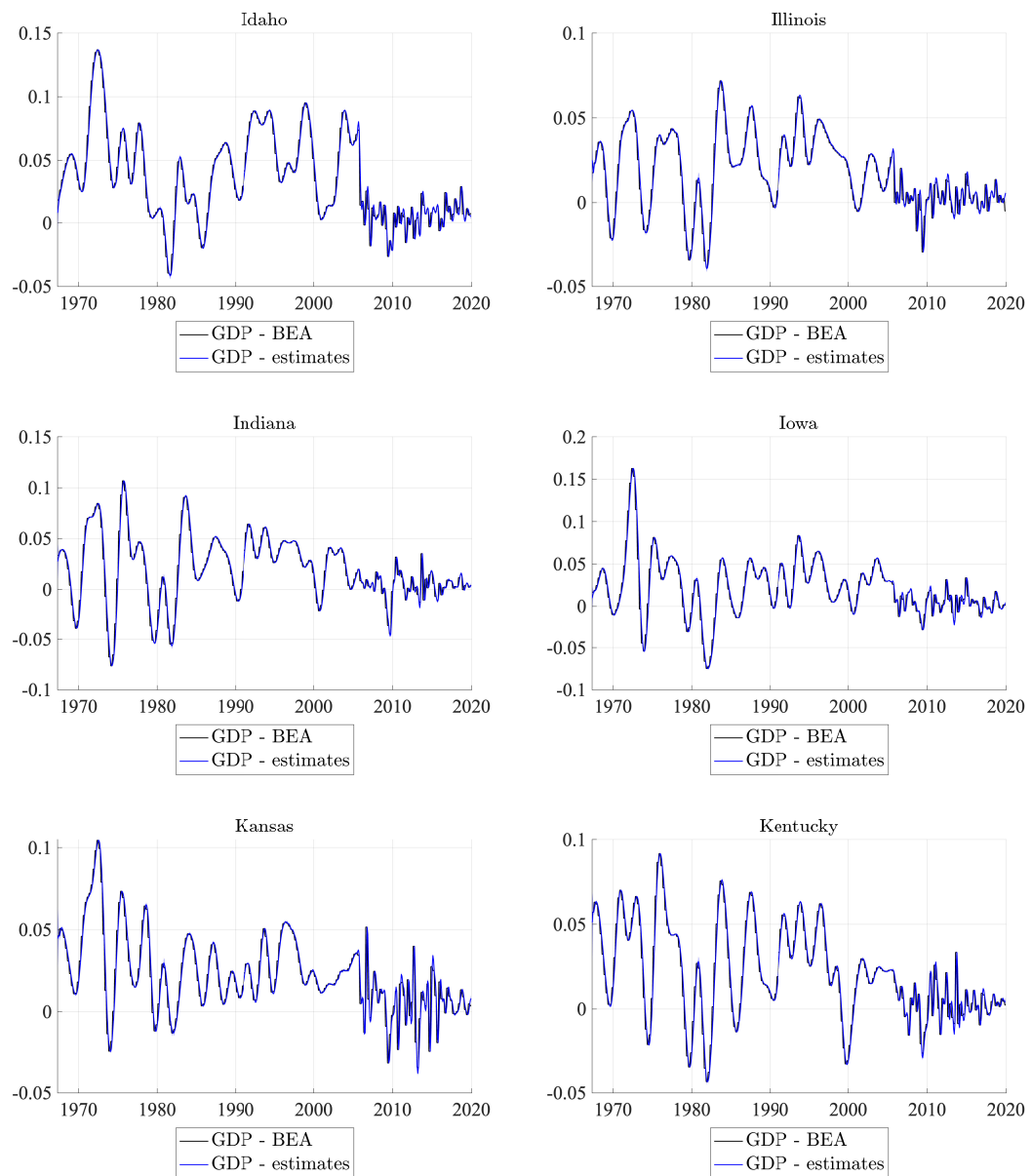


FIGURE 3.5: Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC)

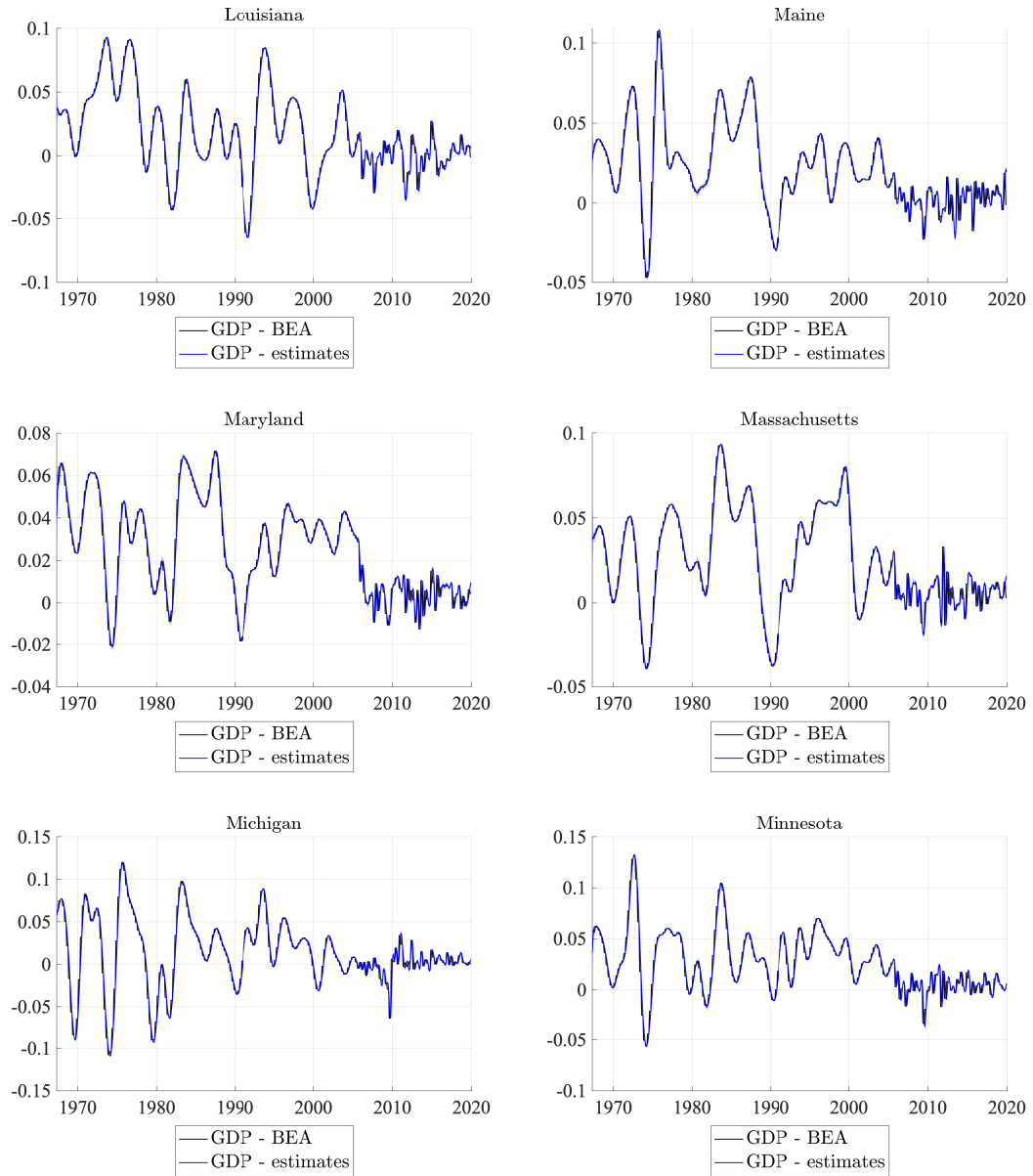


FIGURE 3.6: Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC)

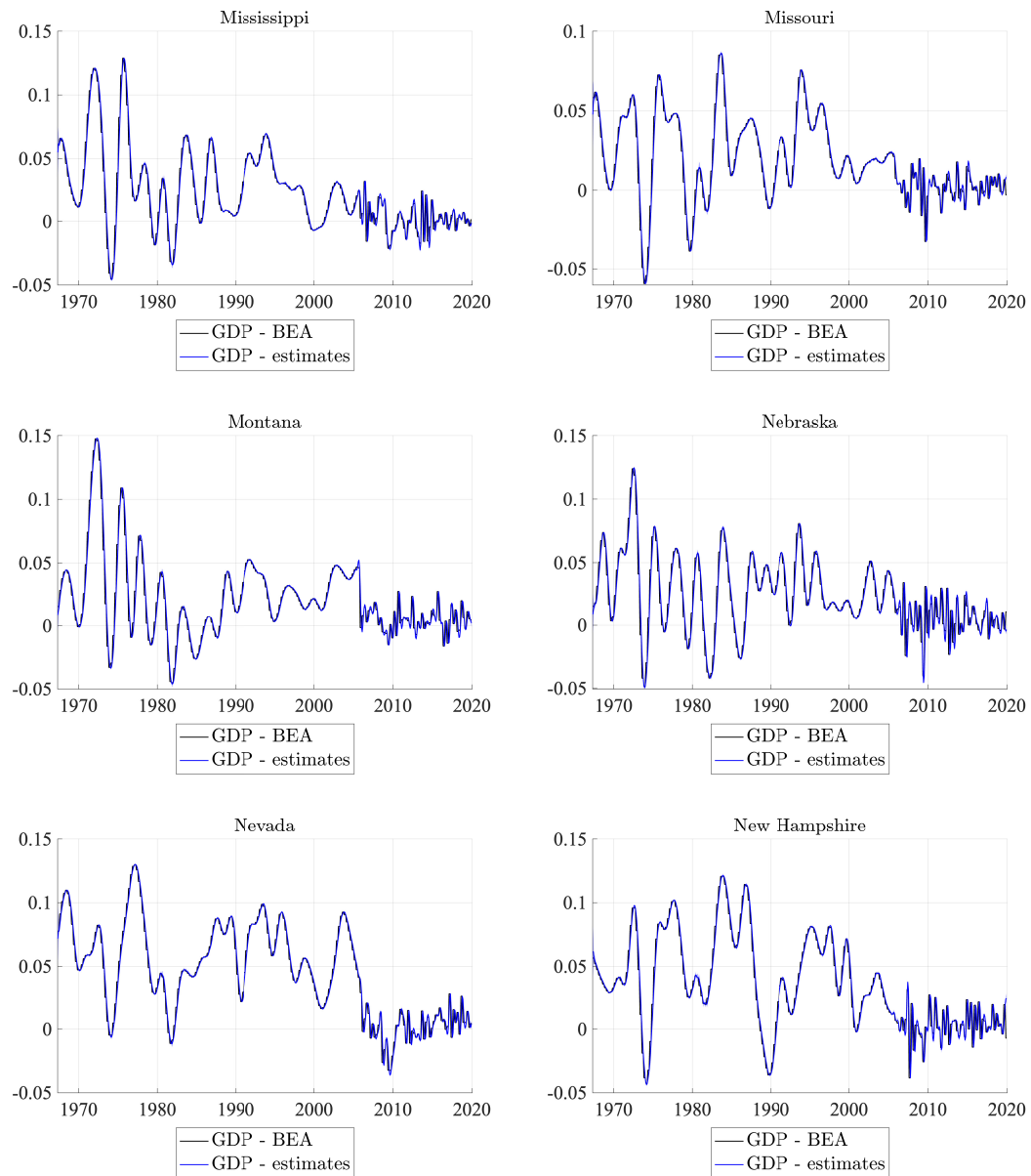


FIGURE 3.7: Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC)

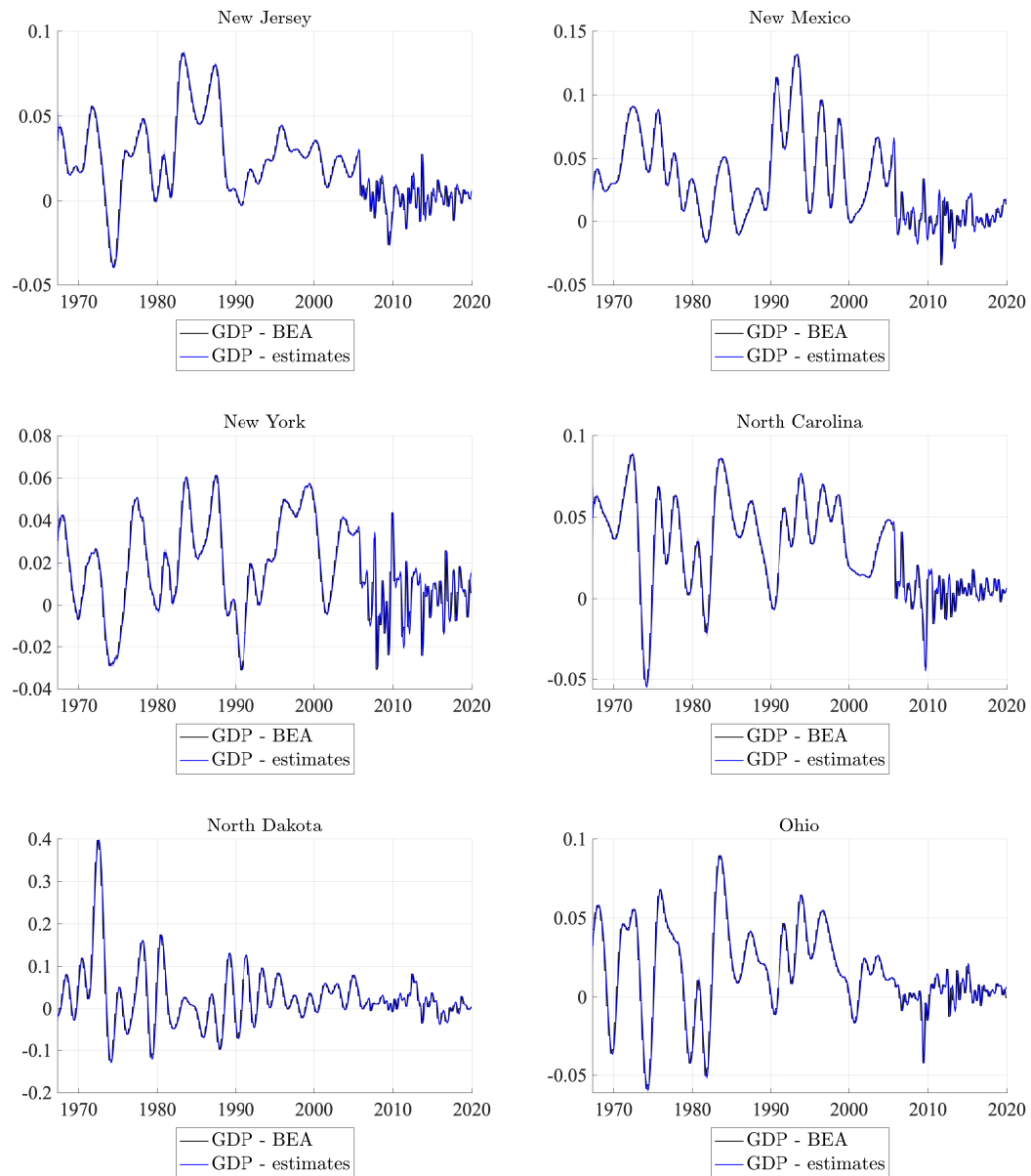


FIGURE 3.8: Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC)

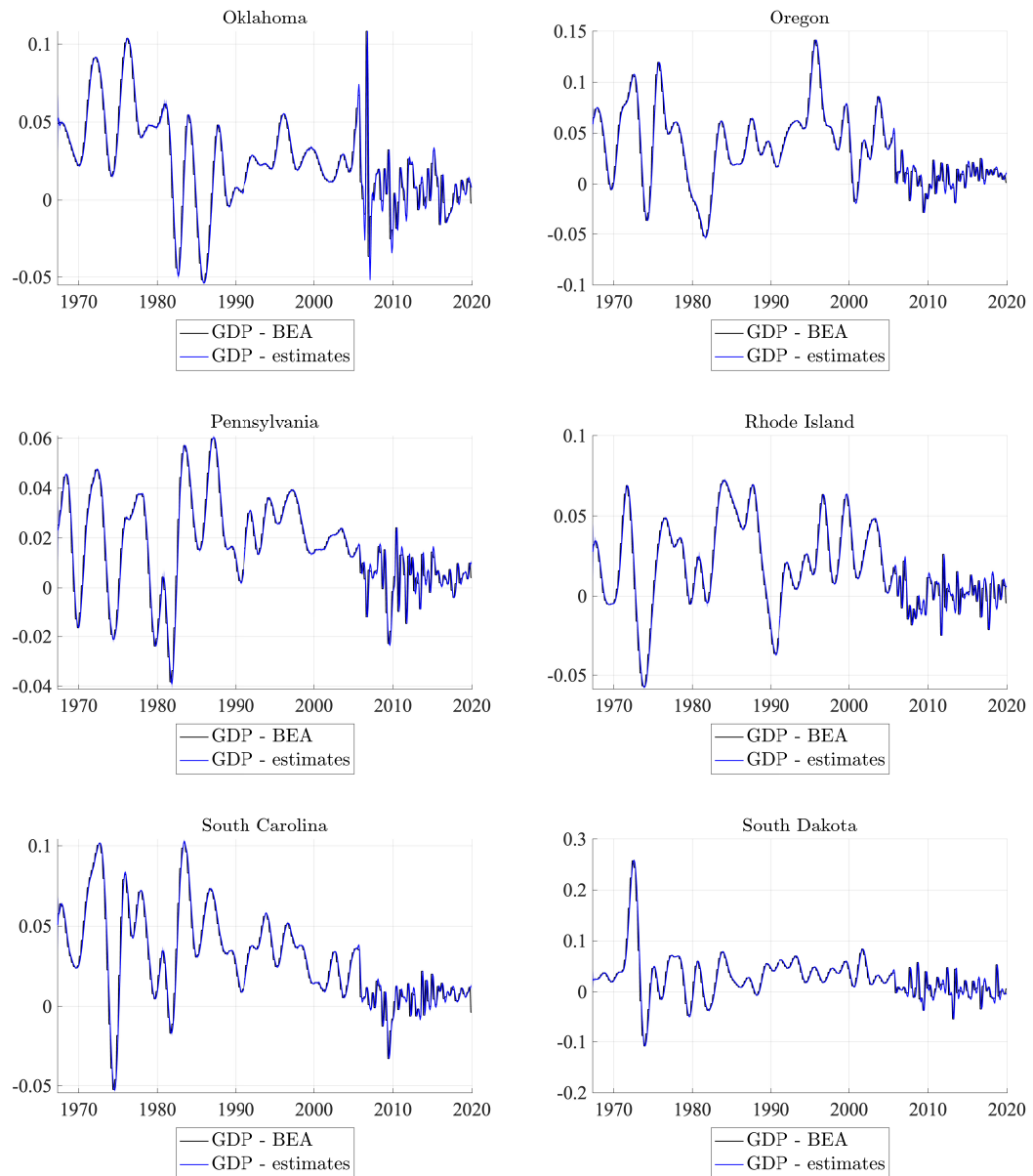




FIGURE 3.9: Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC)

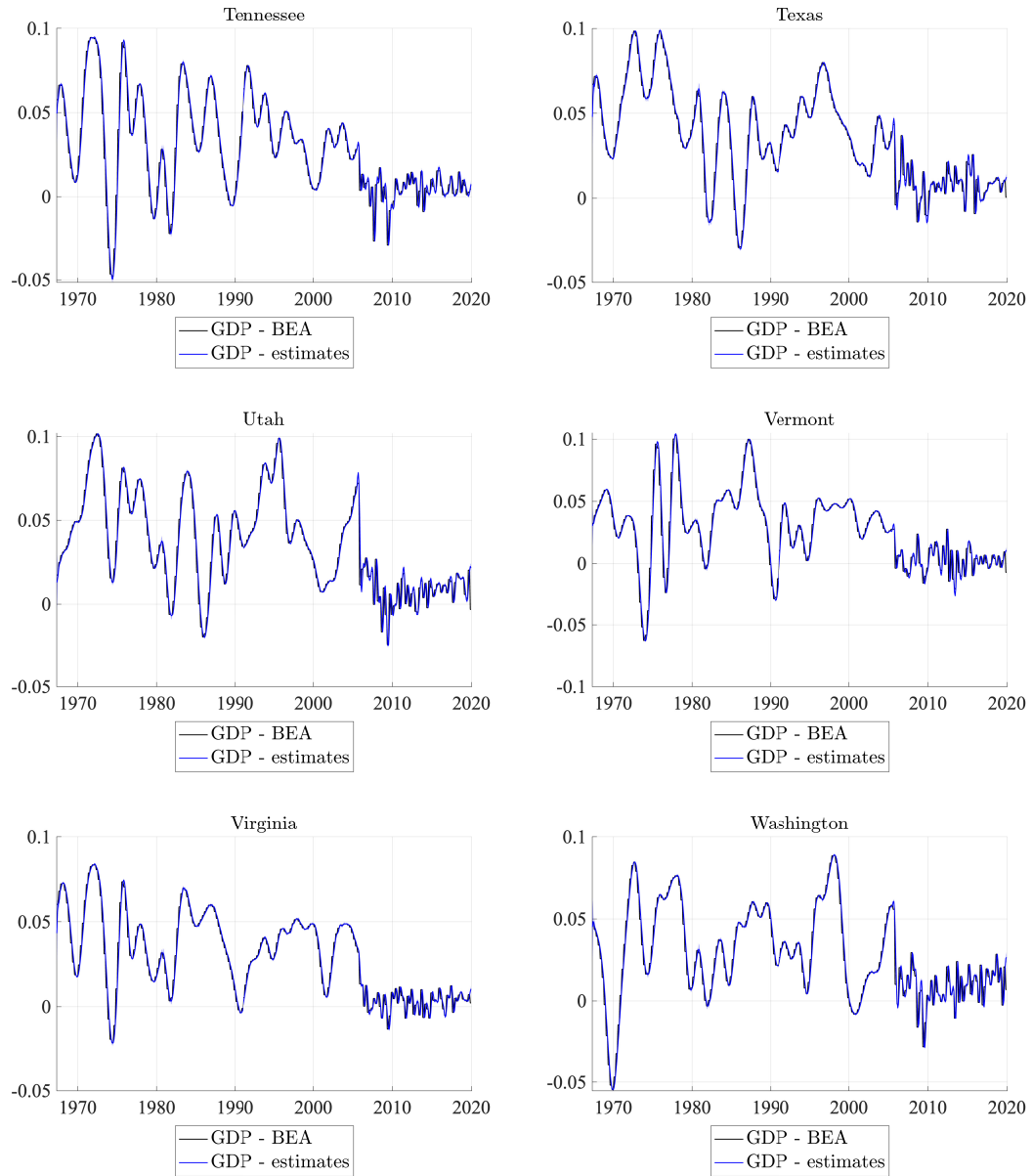
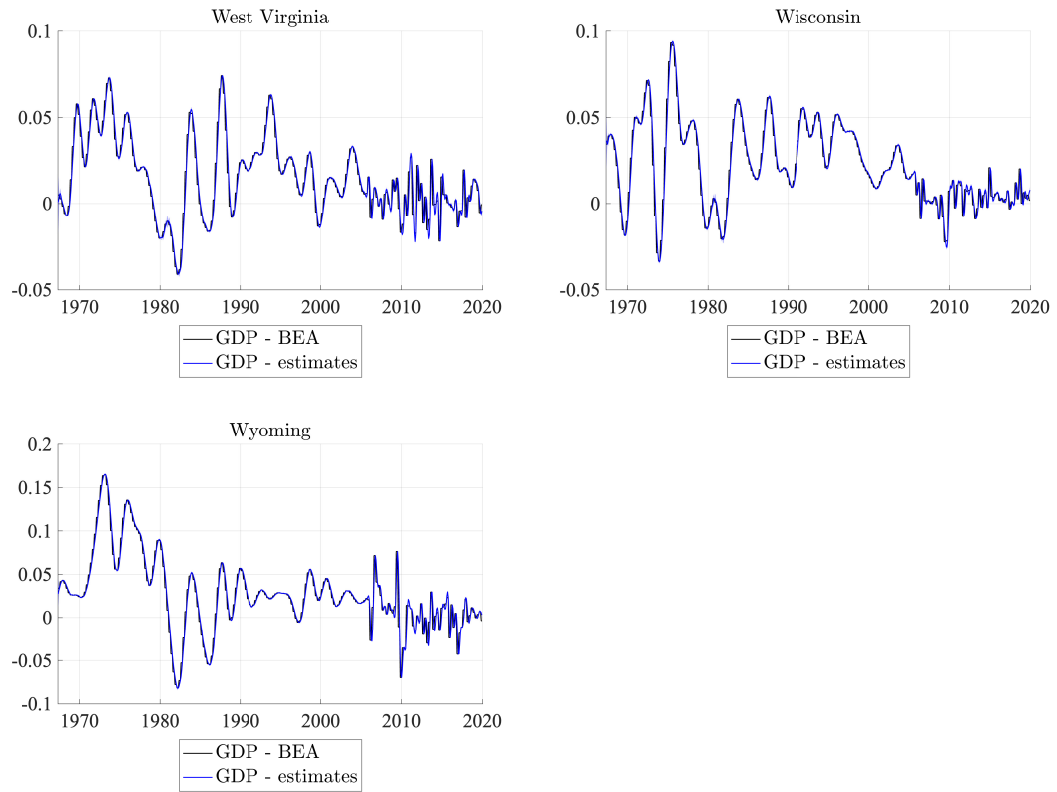


FIGURE 3.10: Historical Estimates of Monthly GDP in the U.S. and its 50 States (plus Washington, DC)



## Chapter 4

# Deep Quantile Regression

### 4.1 Introduction

Since the seminal work of [Koenker and Bassett Jr \(1978\)](#) and [Koenker and Hallock \(2001\)](#), quantile regression has grown in popularity and has found applications in several disciplines both in academia and industry, see e.g. [Chernozhukov and Umansev \(2001\)](#), [Adams, Adrian, Boyarchenko, and Giannone \(2021\)](#) and [Koenker, Chernozhukov, He, and Peng \(2017\)](#). They generalize ordinary sample quantiles to the regression setting, that give more extensive information on the conditional distribution of a dependent variable, given the covariates, relative to the classical regression setting; i.e. estimation of the conditional mean. This extension can be of great importance under extreme events, where the conditional distribution of variables such as asset returns tends to exhibit skewness, or under the presence of outliers and/or asymmetries, see e.g. [Baur and Schulze \(2005\)](#).

An assumption made in the early literature, was the linear association between the conditional quantile of the target variable and predictors. This was predominately an assumption that allowed for streamlined computation and theoretical inference, but was clearly restrictive. A more recent strand of the literature, relaxed the linearity assumption and considered non-parametric estimators for the conditional quantile, that is based on different methods, see e.g. [Belloni, Chernozhukov, Chetverikov, and Fernández-Val \(2019\)](#) and references therein. Recent advances in Machine Learning (ML) literature, which is the focus of this Chapter, show how modelling frameworks

such as neural networks can be used to estimate general, non-linear and potentially highly complicated associations.

Specifically, a large number of studies have shown that *feed-forward* neural networks can approximate arbitrarily well any continuous function of several real variables, see e.g. [Hornik \(1991\)](#), [Hornik, Stinchcombe, and White \(1989\)](#), [Galant and White \(1992\)](#) and [Park and Sandberg \(1991\)](#). Recent work by [Liang and Srikant \(2016\)](#) and [Yarotsky \(2017\)](#), extends this result for *feed-forward* neural networks with multiple layers, provided sufficiently many hidden neurons and layers are available. Notice that, besides neural networks, other non-parametric approaches, e.g. splines, wavelets, the Fourier basis, as well as simple polynomial approximations, do have the universal approximation property, based on the Stone-Weierstrass theorem.

There is considerable empirical work identifying non-linearities and asymmetries in financial variables, see e.g. [Gu, Kelly, and Xiu \(2020a\)](#), [Gu, Kelly, and Xiu \(2020b\)](#), [He and Krishnamurthy \(2013\)](#) and [Pohl, Schmedders, and Wilms \(2018\)](#), where they illustrate that ML offers richer functional form specifications that can capture potential non-linearities between dependent and independent variables. Some examples include [Gu, Kelly, and Xiu \(2020b\)](#) in which, they evaluate the forecast accuracy of machine learning methods in measuring equity risk premia, and find that neural networks give substantial forecasting gains in asset pricing compared to linear models, and [Bucci \(2020\)](#), where a recurrent neural network is proposed, that approximates realised volatility well and outperforms other classic non-linear estimators in forecasting. In a similar fashion, [Smalter Hall and Cook \(2017\)](#) use several neural network architectures to predict unemployment in the US and find that neural networks outperform forecasts from a linear benchmark model at short horizons. In addition, [Gu, Kelly, and Xiu \(2020a\)](#) propose the use of a conditional Autoencoder<sup>1</sup>, and illustrate its superior performance relative to linear unsupervised learning methods.

Before we discuss the contributions of this Chapter, we provide a succinct summary of the current machine learning literature on non-linear quantile and Value-at-Risk (*VaR*) estimation, but we note that the majority of this work, was not available during the writing of this Chapter. [Keilbar and Wang \(2021\)](#) use neural networks to estimate a non-linear conditional *VaR* model introduced by [Tobias and Brunnermeier \(2016\)](#) and find that, it gives significant gains in modelling systemic risk. In addition, [Tambwekar,](#)

---

<sup>1</sup>Autoencoders are artificial neural networks that can be used as a dimensionality reduction technique.

Maiya, Dhavala, and Saha (2021) estimate a non-linear binary quantile regression and develop confidence scores to assess the reliability of prediction. Padilla, Tansey, and Chen (2020) examine the performance of a quantile neural network using Rectified Linear Unit (ReLU) as activation function. They derive a theoretical upper bound for the mean squared error of a ReLU network and show that their non-linear quantile estimator has strong performance of ReLU neural networks for quantile regression across a broad range of function classes and error distributions. Chen, Liu, Ma, and Zhang (2020) propose a unified non-linear framework, based on *feed-forward* neural networks, that allows the estimation of treatment effects, for which they establish consistency and asymptotic normality. Their framework includes the quantile estimator and allows for high-dimensional covariates. ML based estimators for quantiles have been proposed in other fields, see e.g. Meinshausen (2006), where quantile random forests are introduced, and Zhang, Quan, and Srinivasan (2018) that propose a quantile neural network estimator.

In this Chapter, we contribute to the expanding literature on the use of ML in Finance and propose a novel *deep quantile* estimator that can capture non-linear associations between asset returns and predictors and that also allows for high dimensional data. We further consider an alternative architecture that allows the use of mixed frequency data. We also contribute towards the explainable machine learning literature, by proposing the use of partial derivatives as a means of “peeking” inside the black box.

We first explore the small sample properties of the proposed estimator via Monte Carlo experiments, which show that the estimator delivers good finite sample performance. Then we examine the performance of the proposed estimator, in the context of one of the most widely examined problems in finance: that of measuring and subsequently forecasting the risk of a portfolio adequately, via *VaR* modelling. *VaR* is a popular model that was first introduced in the late 80s and since then, has become a standard toolkit in measuring market risk. It measures how much value a portfolio can lose within a given time period with some small probability,  $\tau$ . *VaR* and quantiles are related in the following manner, let  $\mathbf{r} = (r_1, \dots, r_T)'$  denote the returns of a portfolio, then, the  $\tau^{th}$  *VaR* is equivalent of computing the negative value of the  $\tau^{th}$  quantile of  $\mathbf{r}$ ,  $-q_\tau(\mathbf{r})$ .

In this Chapter, we argue, following the non-parametric literature, that the linear relationship between *VaR* and predictors can be restrictive and propose a quantile neural

network estimator that allows a non-linear association between covariates and *VaR*. This method appears particularly suitable for developing sound predictions for the past stock return losses in the US over the sample period from September 1985 up to August 2020, the importance of which has been brought to the forefront by the recent COVID-19 pandemic. Specifically, our aim is to forecast ten-day ahead *VaR* produced from daily *VaR* forecasts. We use daily frequency returns in a fixed forecasting framework that is outlined below. Under this forecasting framework, mixed frequency models become relevant benchmarks to the non-linear quantile estimator, see e.g. [Ghysels, Plazzi, and Valkanov \(2016\)](#). Hence, we also include a linear MIXed DATA Sampling (MIDAS) model as a competitor and also a non-linear MIDAS model, which is an extension to the *deep quantile* estimator. Further, we consider ten-day compounded *VaR* forecasts that exhibit similar patterns, which we relegate to the Online Appendix.

[Shen, Jiao, Lin, Horowitz, and Huang \(2021\)](#) use deep neural networks to examine the convergence properties of a *deep quantile* Regression that can also mitigate the curse of dimensionality under some assumptions. While their work focuses on the statistical properties of *deep quantile* Regression, this Chapter uses ML based estimators to forecast ten-day ahead *VaR* and examines whether deep neural networks have significant gains over traditional forecasting methods. We are not the first to use ML methods for *VaR* forecasting, see e.g. [Du, Wang, and Xu \(2019\)](#), where they propose a recurrent neural network, as a novel forecasting methodology for the *VaR* model and exhibit an improved forecast performance relative to traditional methods. To the best of our knowledge though, there has been no application that uses a neural network quantile estimator in finance for forecasting *VaR*. Note that in this Chapter we also consider a large set of neural networks that also allow for mixed frequency estimation.

Our empirical analysis shows that the proposed *deep quantile* estimator outperforms the linear, MIDAS and other non-parametric quantile models, in forecasting *VaR*. We assess the forecasting accuracy between models based on two statistical tests. The first is the [Diebold and Mariano \(1995\)](#) test with the [Harvey, Leybourne, and Newbold \(1997\)](#) adjustment, and the second is the [Giacomini and White \(2006\)](#) test. Results from both tests suggest that our neural network estimator has higher accuracy in forecasting *VaR*. We use the linear quantile method as a benchmark to assess whether our proposed estimator has predictive gains or not. This measure illustrates gains up to 74% relative to the linear one, for the *deep quantile* estimator and up to 76% for the

non-linear MIDAS model. Further, we use the quantile score test that provides further evidence in favour of our neural network estimator.

We further examine whether our proposed estimator nests forecasts produced from the linear and other non-parametric models, using the encompassing test of [Giacomini and Komunjer \(2005\)](#). Overall, we find that forecasts from the *deep quantile* estimator encompass forecasts from competing models more times than vice versa. There are some cases where the test is inconclusive, suggesting that a forecast combination from a different pair of models would provide a better result, which is in line with the result of [Bates and Granger \(1969\)](#).

While ML methods show a great capacity at both approximating highly complicated non-linear functions and forecasting, they are routinely criticized as they lack interpretability and are considered a “black box”; in the sense that they do not offer simple summaries of relationships in the data. Recently though, there has been a number of studies that try to make ML output interpretable, see e.g. [Athey and Imbens \(2017\)](#), [Wager and Athey \(2018\)](#), [Belloni, Chernozhukov, and Hansen \(2014\)](#), [Joseph \(2019\)](#). In this Chapter we also try to understand in a semi-structural fashion, which variables impact the forecasting performance of the *deep quantile* estimator more. To this end, we first use Shapley Additive Explanation Values (SHAP) as proposed by [Lundberg and Lee \(2017\)](#) and further developed in [Joseph \(2019\)](#), that have started to become a standard tool for interpretability in ML methods. Further we use partial derivatives, as a means of investigating the marginal contribution/influence of each variable to the output. We compare the partial derivatives and SHAP values over time, and our results can be summarised as follows. First, partial derivatives overall are more stable than SHAP values, and are able to produce interpretable results, at a fraction of the computational time of SHAP. Second, the partial derivatives of the *deep quantile* estimator fluctuate around the estimate of the conditional linear quantile and i) exhibit time variation and ii) can capture stressful events in the U.S. economy for instance the COVID-19 pandemic and the 2008 financial crisis.

The remainder of the Chapter is organised as follows. Section [4.2](#) introduces the *deep quantile* estimator. Section [4.3](#) contains the Monte Carlo exercise. Section [4.4](#) presents our empirical application. Section [4.5](#) presents the semi-structural analysis. Conclusions are set out in Section [4.6](#). We relegate to the Online Appendix the specifications of the competing models, empirical results from one-step ahead *VaR* forecast, ten-day

compounded VaR forecasts and results from the quantile score test and predictive gains.

## 4.2 Theory

In this Section we start by summarising the underlying theory of a quantile regression as outlined by [Koenker and Bassett Jr \(1978\)](#) and [Koenker \(2005\)](#) and argue that the linear relationship of the conditional quantile between a dependent variable given the covariates, can be restrictive. We illustrate how some fundamental results on the universal approximation property of neural networks can be used to approximate a non-linear relationship instead, and propose a *deep quantile* estimator. We conclude with a discussion on how different penalisation schemes can be used and further how hyper-parameters can be selected via Cross Validation (CV).

### 4.2.1 Linear Quantile Regression

The standard goal in econometric analysis is to infer a relationship between a dependent variable and one or more covariates. Let  $\{y_t, \mathbf{x}_t\}_{t=1}^T$  be a random sample from the following linear regression model

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t, \quad (4.1)$$

where  $y_t$  is the dependent variable at time  $t$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  is a vector of unobserved slope parameters,  $\mathbf{x}_t = (x_{t1}, \dots, x_{tp})'$  is a vector of known covariates, and  $u_t$  is the random error of the regression which satisfies  $E(u_t | \mathbf{x}_t) = 0$ . Standard regression analysis tries to come up with an estimate of the conditional mean of  $y_t$  given  $\mathbf{x}_t$ , that minimises the expected squared error loss:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \frac{1}{T} \sum_{t=1}^T (y_t - \mathbf{x}_t' \boldsymbol{\beta})^2. \quad (4.2)$$

This can be restrictive though, when i) non-linearities and outliers exist and ii) since it provides just an aspect of the conditional distribution of  $y_t$ , given  $\mathbf{x}_t$  by construction. These potential limitations led to the development of quantile regression. In



their seminal work, [Koenker and Bassett Jr \(1978\)](#) generalise ordinary sample quantiles to the regression setting, that give more complete information on the conditional distribution of  $y_t$  given  $\mathbf{x}_t$ , for which we now provide a succinct description.

The quantile regression model can be defined as

$$Q_y(\tau|\mathbf{x}_t) = \mathbf{x}_t' \boldsymbol{\beta}(\tau), \quad \tau \in (0, 1), \quad (4.3)$$

such that  $y_t$  satisfies the quantile constraint  $Pr[y_t \leq \mathbf{x}_t' \boldsymbol{\beta}(\tau) | \mathbf{x}_t] = \tau$ , where  $\boldsymbol{\beta}(\tau)$  are regression coefficients that depend on  $\tau$ . Quantile regression tries to come up with an estimate for the  $\tau^{\text{th}}$  conditional quantile,  $\widehat{Q}_y(\tau, \mathbf{x}_t) := \widehat{\boldsymbol{\beta}}(\tau)$ , by minimizing the following function

$$\widehat{\boldsymbol{\beta}}(\tau) = \arg \min_{\boldsymbol{\beta}} \frac{1}{T} \sum_{t=1}^T \rho_{\tau}(y_t - \mathbf{x}_t' \boldsymbol{\beta}(\tau)), \quad (4.4)$$

where  $\rho_{\tau}(\cdot)$  is the quantile loss function defined as

$$\rho_{\tau}(u_t) = \begin{cases} \tau u_t(\tau), & \text{if } u_t(\tau) \geq 0 \\ (1 - \tau) u_t(\tau), & \text{if } u_t(\tau) < 0 \end{cases}$$

and  $u_t(\tau) = y_t - \mathbf{x}_t' \boldsymbol{\beta}(\tau)$ . The quantile estimator in eq. (4.4), provides i) much richer information on the whole conditional distribution of  $y_t$  as a function of  $\mathbf{x}_t$ , and ii) more robust estimates under the presence of outliers and non-linearities, when compared to the ordinary least squares estimator.

Notice that the linear association assumption,  $Q_y(\tau|\mathbf{x}_t) = \mathbf{x}_t' \boldsymbol{\beta}(\tau)$ , can be generally restrictive. Instead, we consider the case of the following non-linear association,

$$Q_y(\tau|\mathbf{x}_t) = h_{\tau}(\mathbf{x}_t),$$

where  $h_{\tau}(\cdot)$  is some unknown, (potentially highly) non-linear function. In this Chapter we propose an estimation strategy to approximate  $h_{\tau}(\mathbf{x}_t)$  with neural networks using their universal approximation property. Specifically, we assume that there exists a neural network with a function  $G_{\tau}(\mathbf{x}_t, \mathbf{w})$ , to be defined below, that can approximate  $h_{\tau}(\mathbf{x}_t)$  well. Before we illustrate how our methodology is implemented, we provide a discussion on how neural networks can approximate  $h_{\tau}(\mathbf{x}_t)$ .

## 4.2.2 Neural Networks

In this Chapter, we limit our attention to *feed-forward* neural networks, to approximate  $h_\tau(\mathbf{x}_t)$ . This architecture consists of an input layer of covariates, the hidden layer(s) where non-linear transformations of the covariates occur, and the output layer that gives the final prediction. Each hidden layer has several interconnected neurons relating it to both the previous and the next one. Specifically, information flows from one layer to the other, via neurons only in one direction, and the connections correspond to weights. Optimising a loss function *w.r.t* these weights makes neural networks capable of learning.

Throughout our exposition,  $L$  denotes the total number of hidden layers, a measure for the depth of a neural network, and  $J^{(l)}$  denotes the total number of neurons at layer  $l$ , a measure for its width. We start by presenting a general definition of a deep (multi-layer) *feed-forward* neural network. Let  $\sigma_l(\cdot)$ ,  $l = 0, \dots, L$  be the activation function used at the  $l^{\text{th}}$  layer, that is applied element-wise and induces non-linearity. We use the ReLU activation function,  $\sigma_l(\cdot) = \max(\cdot, 0)$ , for  $l = 1, \dots, L - 1$  and a linear one for the output layer,  $l = L$ . We denote by  $g^{(l)}$  the output of the  $l^{\text{th}}$  layer which is a vector of length equal to the number of the  $J^{(l)}$  neurons in that layer, such that  $g^{(0)} = \mathbf{x}_t$ . Then, the overall structure of the network is equal to:

$$G_\tau(\mathbf{x}_t, \mathbf{w}) = g^{(L)} \left( g^{(L-1)} \left( \dots \left( g^{(1)}(\cdot) \right) \right) \right), \quad (4.5)$$

where

$$g^{(l)}(\mathbf{x}_t) = \sigma_l \left( \mathbf{W}^{(l-1)} g^{(l-1)} + \mathbf{b}^{(l)} \right), \quad l = 1, \dots, L, \quad (4.6)$$

$\mathbf{W}^{(l)}$  is a  $J^{(l)} \times J^{(l-1)}$  matrix of weights,  $\mathbf{b}^{(l)}$  is a  $J^{(l)} \times 1$  vector of biases giving an overall vector  $\mathbf{w} = \left( \text{vec}(\mathbf{W}^{(0)})', \dots, \text{vec}(\mathbf{W}^{(L)})', \mathbf{b}^{(1)'}', \dots, \mathbf{b}^{(L)'}' \right)'$  of trainable parameters of dimensions  $J^{(l)}(1 + J^{(l-1)})$  total number of parameters in each hidden layer  $l$ ,  $J^{(0)} = p$  and  $J^{(L)} = 1$ .

According to various universal approximation theorems (see e.g. the theoretical results in [Hornik \(1991\)](#), [Hornik, Stinchcombe, and White \(1989\)](#), [Galant and White \(1992\)](#), [Kapetanios and Blake \(2010\)](#), [Liang and Srikant \(2016\)](#) and [Yarotsky \(2017\)](#)),  $G_\tau(\mathbf{x}_t, \mathbf{w})$  can approximate arbitrarily well  $h_\tau(\mathbf{x}_t)$ , such that, for any  $\epsilon > 0$ ,

$$\sup_t |G_\tau(\mathbf{x}_t, \mathbf{w}) - h_\tau(\mathbf{x}_t)| < \epsilon. \quad (4.7)$$

In this sense, the above  $(\epsilon)$ -approximation can be seen as a sieve type non-parametric estimation bound, where  $\epsilon$  can become arbitrarily small by increasing the complexity of  $G_\tau(\mathbf{x}_t, \mathbf{w})$ .

The increase in complexity can occur, either by letting  $L \rightarrow \infty$ , which stands for *deep learning*, or by letting  $J^{(l)} \rightarrow \infty$ . While asymptotically, both ways deliver the same results (see e.g. Farrell, Liang, and Misra (2021) and references therein), the approximation error has been shown to decline exponentially with  $L$ , see e.g. Babii, Chen, Ghysels, and Kumar (2020) but only polynomially with  $J^{(l)}$ , providing some evidence for the prevalent use of deep learning. Notice that there also exists an alternative approximation theory for sparse deep learning, see e.g. the work of Schmidt-Hieber (2020). As an illustration, in the Online Appendix we depict a simple *feed-forward* neural network with two inputs, two hidden layers, a total of five neurons and one output layer.

### 4.2.3 Non-linear Quantile Regression

We assume that the conditional quantile follows a non-linear relationship  $Q_y(\tau|\mathbf{x}_t) = h_\tau(\mathbf{x}_t)$  and there exists a function  $G_\tau(\mathbf{x}_t, \mathbf{w})$ , that can  $(\epsilon)$ -approximate  $h_\tau(\mathbf{x}_t)$ , see the bound in eq. (4.7). Using this assumption, we can formally define the conditional quantile function as the following approximation

$$Q_y(\tau|\mathbf{x}_t) = G_\tau(\mathbf{x}_t, \mathbf{w}) + O(\epsilon),$$

where  $G_\tau(\mathbf{x}_t, \mathbf{w})$  is the unknown non-linear function we want to estimate in order to approximate  $h_\tau(\mathbf{x}_t)$ . We obtain the deep neural network conditional quantile estimate from the solution of the following minimization problem:

$$Q_y(\tau|\mathbf{x}_t) = \arg \min_{\mathbf{w}} \frac{1}{T} \sum_{t=1}^T \rho_\tau(y_t - G_\tau(\mathbf{x}_t, \mathbf{w})), \quad (4.8)$$

where  $\mathbf{w} = (\text{vec}(\mathbf{W}^{(0)})', \dots, \text{vec}(\mathbf{W}^{(L)})', \mathbf{b}^{(1)'}, \dots, \mathbf{b}^{(L)'})'$  contains all model parameters, and  $G_\tau(\mathbf{x}_t, \mathbf{w})$  denotes the overall non-linear mapping, described in eq. (4.5) and (4.6). Notice that the choice of  $G_\tau(\mathbf{x}_t, \mathbf{w})$  will govern whether the model is parametric or non-parametric. If the number of neurons and layers is small, then the model is parametric, if the above number becomes large, then the model becomes

non-parametric, since the number of estimated parameters increases with the sample size, similar to sieve non-parametric approximations.

To allow the use of mixed frequency data, we can make the following changes to the structure of the network  $G_\tau(\mathbf{x}_t, \mathbf{w})$ :

In the input layer, we implement frequency alignment on each input variable  $x_t$  according to the corresponding maximum lag order  $K$ . Thus, each high frequency predictor  $x_t$  is transformed into a low frequency vector  $x_t^* = B(L_\varphi; \vartheta)x_t$ ,

$$B(L_\varphi; \vartheta) = \sum_{k=0}^K B(k; \vartheta)L_\varphi^k, \quad B(k; \vartheta) = \frac{\exp(\vartheta_1 k + \vartheta_2 k^2)}{\sum_{k=1}^K \exp(\vartheta_1 k + \vartheta_2 k^2)}, \quad (4.9)$$

where  $B(k; \vartheta)$  is the normalised Almon polynomial,  $L_\varphi^k$  is a lag operator such that  $L_\varphi^k x_t^\varphi = x_{t-k}^\varphi$ ; the lag coefficients in  $B(k; \vartheta)$  of the corresponding lag operator  $L^k$  are parameterised as a function of a small dimensional vector of parameters  $\vartheta$ . We use this weight function on the frequency alignment vector to reduce the number of parameters and ensure a parsimonious specification. As a consequence, the low frequency variable  $x_t^*$  which has the same frequency as the output  $y_t$  is obtained. The rest of the architecture of the *deep MIDAS* follows the architecture of the *deep quantile* estimator, but instead of using  $x_t$  in eq. (4.6), we use  $x_t^*$ .

#### 4.2.4 Regularized Non-Linear Quantile Regression

Neural networks have a great capacity to estimate non-linear relationships from the data, but this comes at a cost, since they are prone to over-fitting. This can lead to a severe drop in their forecasting performance, especially in small samples. There is a variety of commonly used techniques in ML, see e.g. [Gu, Kelly, and Xiu \(2020a\)](#) for a good summary, that can be used to ease this impact, originally coming from the high-dimensional statistical literature. The reader is also referred to [Goodfellow, Bengio, and Courville \(2016\)](#) for an excellent summary of different topics on the implementation of neural networks, including regularization.

#### 4.2.4.1 Regularization

A common solution to this caveat is regularization, where a penalty term is imposed on the weights of the neural network and is appended in the loss function. Regularization, generally improves the out-of-sample performance of the network by decreasing the in-sample noise from over-parameterization, utilising the bias-variance trade-off. Further, another benefit of regularization is that it provides computational gains in the optimization algorithm. The penalised loss function, for a given quantile  $\tau$ , can be written as:

$$L(G_\tau(\mathbf{x}_t, \mathbf{w}), y_t) = \frac{1}{T} \sum_{t=1}^T \rho_\tau(y_t - \hat{G}_\tau(\mathbf{x}_t, \mathbf{w})) + \phi(\mathbf{w}), \quad (4.10)$$

where the penalty term is

$$\phi(\mathbf{w}) = \begin{cases} \lambda \|\mathbf{w}\|_1, & \text{Lasso} \\ \lambda \|\mathbf{w}\|_2^2, & \text{Ridge} \\ \lambda(1 - \alpha)\|\mathbf{w}\|_1 + \lambda\alpha\|\mathbf{w}\|_2^2, & \text{Elastic Net} \\ 0, & \text{otherwise} \end{cases}$$

and  $\lambda$  and  $\alpha$  are tuning parameters, for which we discuss their selection below. Generally, there is a plethora of loss functions, and the choice among them, depends mainly on the task at hand. In this Chapter we use the quantile loss function. The different penalisation schemes on  $\phi(\mathbf{w})$  work as follows: *deep LASSO* or  $l_1$ -norm penalisation, is a regularization method that shrinks uniformly all the weights to zero, and some at exactly zero. The latter is referred to as the variable selection property of the *deep LASSO*. *Deep Ridge* works in a similar manner to the *deep LASSO*, by shrinking the weights, uniformly to zero, but not at exactly zero. Finally, the *deep Elnet*<sup>2</sup> is a combination of *deep LASSO* and *deep Ridge*, that has been shown to retain good features from both methods, see e.g. [Zou and Hastie \(2005\)](#).

#### 4.2.4.2 Cross Validation

We use Cross Validation (CV) to calibrate all the different (hyper)-parameters outlined above, and aim to maximise the out-of-sample (forecasting) performance of the network.

---

<sup>2</sup>Deep Elastic net

Our CV scheme consists of choices on: i) the total number of layers ( $L$ ) and neurons ( $J$ ), ii) the learning rate ( $\gamma$ ) for the Stochastic Gradient Decent (SGD), iii) the batch size, dropout rate and the level of regularization. Regarding the choice on the activation function, we use ReLU for the hidden layers and a linear function for the output layer. Overall, our aim is to build a neural network that has the best pseudo-out-of-sample (POOS) performance. To achieve this, we need to evaluate the model, select the optimal parameters and hyper-parameters and test its POOS behaviour. It is clear that tuning all these different architectures, parameters and hyper-parameters increases the computational cost a lot.

For this reason we tune the learning rate for the optimiser,  $\gamma$ , from five discrete values in the interval  $[0.01, 0.001]$ . For the width and depth of the neural network we tune the hyper-parameters from the following grids  $[1, 5, 10]$  and  $[10, 30, 50]$ , respectively. The batch size is selected via the following grid  $[10, 20]$ .<sup>3</sup> Furthermore, we tune the regularization parameter,  $\lambda$ , from five discrete values in the interval  $[0.01, 0.001]$ , both for *deep LASSO* and *deep Ridge*, and for the case of the elastic net we choose  $\alpha$  from a grid  $[0.1, 0.5, 0.9]$ . We also use dropout regularization, where the dropout probability is up to 20%, see e.g. [Gu, Kelly, and Xiu \(2020b\)](#).

For the non-linear MIDAS, we also cross validate  $\theta_1$  from eight discrete values in the interval  $[-1, 0.5]$  and for  $\theta_2$ , we use six discrete values in  $[-0.5, 0.5]$ .

We split the whole sample into three distinct subsamples, the training, validation and test subsamples. These subsamples are consequential to maintain the time series structure of the data. The training subsample consists of the first 60% of the sample, the validation is the next 20% of sample and the test is final 20% of sample. First, we use the training sample to estimate (i.e. train) the network parameters. Then, the second subsample or validation is used to tune hyper-parameters by constructing the fitted/-forecasted values given the parameters from the training sample. We proceed with the calculation of the quantile loss function as in eq. (4.10) and evaluate the models' POOS performance on this subset. We repeat the same process  $\Delta$  number of times, where  $\Delta$  is the number of all possible combinations of points across quantiles. We store the quantile loss values and select the parameters and hyper-parameters that minimise the quantile loss based on the POOS forecasts. In the validation step we wish to find

---

<sup>3</sup>We have also considered batch normalisation and find that overall, results exhibit similar pattern with and without it.

the optimal parameters and hyper-parameters that capture complex non-linear relations and produce reliable POOS forecasts.

Finally, in the test subsample we use the optimal parameters and hyper-parameters from the validation step and evaluate the out-of-sample performance of the network.

#### 4.2.4.3 Optimisation

The estimation of neural networks is generally a computational cumbersome optimization problem due to non-linearities and non-convexities. The most commonly used solution utilises stochastic gradient descent (SGD) to train a neural network. SGD uses a batch of a specific size, that is, a small subset of the data at each iteration of the optimization to evaluate the gradient, to alleviate the computation hurdle. The step of the derivative at each epoch is controlled by the learning rate,  $\gamma$ . We use the adaptive moment estimation algorithm (ADAM) proposed by [Kingma and Ba \(2014\)](#)<sup>4</sup>, which is a more efficient version of SGD.

## 4.3 Monte Carlo

### 4.3.1 Setup

In this Section we present Monte Carlo (MC) experiments, in order to study the finite sample performance of the *deep quantile* estimator proposed in Section 4.2, for the different penalisation schemes. We generate artificial data  $\{y_t\}$  using a single predictor  $\{x_t\}$ , according to the following model

$$y_t = h_\tau(x_t) + u_t, \quad (4.11)$$

where  $u_t$  is the realisation of a random variable  $u$  distributed as,  $u_t \sim iidN(-\sigma\Phi^{-1}(\tau), \sigma^2)$ ,  $\sigma = 0.1$  and  $\Phi^{-1}$  is the quantile function of the standard normal distribution.  $h_\tau(\cdot)$  is the general non-linear function that we wish to approximate via the *deep quantile* estimator.

---

<sup>4</sup>ADAM is using estimates for the first and second moments of the gradient to calculate the learning rate.

All the experiments are based on the following values:  $\tau \in (1\%, 2.5\%, 5\%, 10\%, 20\%)$ ,  $T \in (100, 300, 500, 1000, 2000, 5000)$  and the number of MC replications is 100. We consider the following four *data generating mechanisms* (DGM) to assess the finite sample properties of the *deep quantile estimator*:

**Case I:** We consider the case of a  $N(0, 1)$  simulated single predictor that is generated as

$$y_t = h_\tau(x_t) + u_t, \quad h_\tau(x_t) = \sin(2\pi x_t), \quad x_t \sim N(0, 1).$$

This is the simplest design in our Monte Carlo experiments. We use this simple case to showcase that linear methods, as expected, cannot produce reasonable performance under a sigmoid type of a non-linear function  $h_\tau(\cdot)$ .

**Case II:** We consider an AR(1) simulated single predictor as follows

$$y_t = h_\tau(x_t) + u_t, \quad h_\tau(x_t) = \sin(2\pi x_t),$$

where  $x_t$  is simulated as

$$x_t = 0.8x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1).$$

In this design we increase the complexity by introducing a correlated predictor.

**Case III:** We consider the case of a single predictor generated via a GARCH(1,1) model

$$y_t = h_\tau(x_t) + u_t, \quad h_\tau(x_t) = \sin(2\pi x_t),$$

where  $x_t$  is simulated as:

$$x_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = 1 + 0.7x_{t-1}^2 + 0.2\sigma_{t-1}^2.$$

In this design, we wish to examine how the proposed estimator performs, when the regressor is conditionally heteroskedastic, following a GARCH(1, 1) model. A GARCH type of assumption on the distribution of asset returns is one commonly used in the literature.



**Case IV:** We consider the case of a single predictor that is generated as follows:

$$y_t = h_\tau(x_t) + u_t, \quad h_\tau(x_t) = G_\tau(\mathbf{x}_t, \mathbf{w}), \quad x_t \sim N(0, 1).$$

In this case we simulate  $h_\tau(x_t)$  to reflect a function composition, commonly used in neural networks. We simulate it with 3 hidden layers and a specific number of neurons, such as

$$G_\tau(\mathbf{x}_t, \mathbf{w}) = \left( \mathbf{W}^{(3)} \left( \sin \left( \mathbf{W}^{(2)} \left( \sin \left( \mathbf{W}^{(1)} \left( \sin \left( \mathbf{W}^{(0)} x_t' + \mathbf{b}^{(1)} \right) \right) + \mathbf{b}^{(2)} \right) \right) + \mathbf{b}^{(3)} \right) \right) \right)',$$

where  $\mathbf{w} = (\text{vec}(\mathbf{W}^{(0)})', \dots, \text{vec}(\mathbf{W}^{(3)})', \mathbf{b}^{(1)'}, \dots, \mathbf{b}^{(3)'})'$ ,  $\mathbf{W}^{(0)}$  is  $50 \times 1$ ,  $\mathbf{W}^{(1)}$  is  $10 \times 50$ ,  $\mathbf{W}^{(2)}$  is  $8 \times 10$  and  $\mathbf{W}^{(3)}$  is  $1 \times 8$ . Further, we simulate the weights,  $\mathbf{w}$ , so that every entry  $w_{ij}$  is simulated as,  $w_{ij} = \delta_{ij} 1(\delta_{ij} > 0.5)$ , where  $\delta_{ij} \sim U(0, 1)$ , allowing for some sparsity.

Across all cases, we estimate  $h_\tau(x_t)$  using our proposed estimator with different penalisation schemes. Let  $\hat{h}_{\tau, pen} = \hat{G}_{\tau, pen}(x_t, \mathbf{w})$  denotes the estimate, where *pen* corresponds to no regularization, *deep LASSO*, *deep Ridge* and *deep Elnet*. We use the following metrics in order to evaluate the small sample properties, of our *deep quantile* estimator across  $R = 100$ , MC replications: i) the average mean squared error of the true residuals,  $AMSE_{u_t} = \frac{1}{R} \frac{1}{T} \sum_{i=1}^R \left( \sum_{t=1}^T u_t^2 \right)_i$ , ii) the average mean squared error of the estimated residuals,  $AMSE_{\hat{u}_t, pen} = \frac{1}{R} \frac{1}{T} \sum_{i=1}^R \left( \sum_{t=1}^T (y_t - \hat{y}_{t, pen})^2 \right)_i$  and finally, iii) the average absolute bias  $ABIAS_{\hat{h}_{\tau, pen}} = \frac{1}{R} \frac{1}{T} \sum_{i=1}^R \left( \sum_{t=1}^T |(h_\tau(x_t) - \hat{G}_{\tau, pen}(x_t, \mathbf{w}))| \right)_i$ . We report results only for  $AMSE_{\hat{u}_t, pen}$  below, since results for the alternative metrics exhibit similar patterns and are available upon request.

Figures 4.1 – 4.4 about here

### 4.3.2 Results

We present our Monte Carlo results for Cases I – IV in Figures 4.1 – 4.4 respectively. In Figure 4.1, we can see that the linear quantile estimator, under a non-linear setup doesn't work as expected and the MSE remains constant as the sample size increases. Next we present the asymptotic properties for our proposed estimator across different penalization schemes, namely *deep quantile*, *deep LASSO*, *deep Ridge* and *deep Elastic*

*Net*, and find that the proposed non-linear estimators have good finite sample properties.

When  $\tau = 1\%$  it appear that our estimator works well for sample sizes larger than  $T = 300$ , but in comparison with the linear one it generally works better. In Case II our non-linear estimators depict fine finite sample properties and their performance is better than the linear one. In this case the non-regularized estimator performs better than the regularized ones. Next, similar behaviour appears in Case III. In Case IV, where we allow for some sparsity in the weights, we find, as expected, that the linear quantile regression estimator does not work under non-linearity, while the non-linear one works as expected.

Overall, our Monte Carlo results suggest that the *deep quantile* estimator has good finite sample properties, and can approximate non-linear functions. We further find, as expected, that the linear quantile regression estimator does not work under non-linearity. Finally, we find evidence in favour of the penalisation schemes proposed in Section 4.2. Specifically, the penalised *deep quantile* estimators also have good finite sample properties, and in some cases, perform better than the non-regularized one; a finding in favour of weight regularization.

## 4.4 Empirical Setup

In this Section we outline our empirical application setup, where we use the proposed *deep quantile* estimator to forecast *VaR*. We examine the predictive ability of the proposed estimator and other non-parametric models, relative to the linear one, using the quantile encompassing test of [Giacomini and Komunjer \(2005\)](#). We further examine the predictive performance of the different methods by testing their forecasting accuracy, using the [Diebold and Mariano \(1995\)](#), [Giacomini and White \(2006\)](#) and quantile score tests.

### 4.4.1 Deep Quantile *VaR* forecasting

The data used in our empirical application consist of around 36 years of daily prices on the S&P500 index (source: Bloomberg), from September 1985 to August 2020 ( $T = 9,053$  observations). We use daily log returns, defined as  $r_t = \log(P_t/P_{t-1})$  for

our forecasting analysis. We use four different classes of *VaR* models and produce forecasts for  $\tau = (1\%, 5\%, 10\%)$  empirical conditional quantiles, using the *deep quantile* estimator.

The first *VaR* specification we consider is the GARCH(1,1) model that has been proposed by [Bollerslev \(1986\)](#), in which  $\sigma_{1,t}^2 = \omega_0 + \omega_1\sigma_{1,t-1}^2 + \omega_2r_{t-1}^2$ , see eq. (4.12). The second *VaR* specification we consider, is RiskMetrics, proposed by [J.P. Morgan \(1996\)](#), which assumes  $\sigma_{2,t}^2 = \lambda\sigma_{2,t-1}^2 + (1 - \lambda)r_{t-1}^2$ , where for daily returns,  $\lambda = 0.94$ , see eq. (4.13).

The last two specifications we consider follow the *Conditional Autoregressive Value-at-Risk* model (CAViaR), proposed by [Engle and Manganelli \(2004\)](#), where a specific quantile is analysed, rather than the whole distribution. Specifically, the CAViaR model corrects the past  $VaR_{j,t-1}$  estimates in the following way: it increases  $VaR_{j,t}$  when  $VaR_{j,t-1}$  is above the  $\tau^{th}$  quantile, while, when the  $VaR_{j,t-1}$  is less than the  $\tau^{th}$  quantile, it reduces  $VaR_{j,t}$ . Thus, the third *VaR* we examine is the Symmetric absolute value (SV) that responds symmetrically to past returns, see eq. (4.14) and lastly, we consider the Asymmetric slope value (ASV) as it offers a different response to positive and negative returns, see eq. (4.15). For ease of exposition, we refer to the above specification as  $VaR_{1,t}, \dots, VaR_{4,t}$ , respectively. Below we summarise their specifications:

$$VaR_{1,t} = \beta_0 + \beta_1\sigma_{1,t} \quad (4.12)$$

$$VaR_{2,t} = \beta_0 + \beta_1\sigma_{2,t} \quad (4.13)$$

$$VaR_{3,t} = \beta_0 + \beta_1 VaR_{3,t-1} + \beta_2|r_{t-1}| \quad (4.14)$$

$$VaR_{4,t} = \beta_0 + \beta_1 VaR_{4,t-1} + \beta_2 r_{t-1}^+ - \beta_3 r_{t-1}^- \quad (4.15)$$

where  $\beta_i, i = 0, \dots, 3$  are parameters to be estimated. We use these specifications following [Giacomini and Komunjer \(2005\)](#). Under the mixed frequency setup, we consider the following equation

$$VaR_{i,t}^{(MIDAS)} = B(L_\varphi; \vartheta) VaR_{i,t}, \quad (4.16)$$

where  $B(L_\varphi; \vartheta)$  is defined in eq. (4.9),  $i = 1, \dots, 4$  and  $\vartheta$  are parameters to be estimated. For a more detailed summary of MIDAS we refer the reader to [Ghysels, Santa-Clara, and Valkanov \(2004\)](#). As discussed in Section 4.2, the linear association between *VaR* and the covariates can be restrictive. Instead we assume that the relationship between the response variable, *VaR*, and the covariates has an unknown non-linear form for a

given  $\tau$ , that we wish to approximate with our proposed *deep quantile* estimator as

$$VaR_{1,t} = G_{\tau}(\sigma_{1,t}, \mathbf{w}) \quad (4.17)$$

$$VaR_{2,t} = G_{\tau}(\sigma_{2,t}, \mathbf{w}) \quad (4.18)$$

$$VaR_{3,t} = G_{\tau}(VaR_{3,t-1}, |r_{t-1}|, \mathbf{w}) \quad (4.19)$$

$$VaR_{4,t} = G_{\tau}(VaR_{4,t-1}, r_{t-1}^+, r_{t-1}^-, \mathbf{w}), \quad (4.20)$$

where  $VaR_{j,t}$ ,  $j = 1, \dots, 4$  is indexed at (day)  $t = 1, \dots, T$ . The dimension  $p$  of covariates that we use in our analysis depends on the specification chosen for  $VaR$ . Specifically, if  $j = 1, 2$  then  $p = 1$ , if  $j = 3$ ,  $p = 2$  and finally if  $j = 4$  then  $p = 3$ .

In the Online Appendix, we briefly delineate the model specifications for the quantile B-splines, quantile polynomial and quantile MIDAS estimators.

#### 4.4.2 Forecasting Exercise Design

This Section presents our forecasting exercise design. First we split our sample in three distinct parts; the training sample, which is used for the estimation of the weights, the validation sample which is used for tuning the hyperparameters of the models, and the test sample which is used for the evaluation of different models. We use a 60%, 20%, 20% split<sup>5</sup>, which corresponds to 5,053 observations in the training sample, 2,000 in the validation and 2,000 in the test sample.

This specific split is used because we follow [Giacomini and Komunjer \(2005\)](#) and want the power of the Conditional Quantile Forecast Encompassing (CQFE) test to be comparable with this exercise. We use the CV scheme described in Section 4.2 and tune the width and depth of the neural network, the batch size, the learning rate, the dropout rate and the regularization of hyper-parameters. Generally, a forecasting exercise is performed either via a recursive or rolling window, see e.g. [Ghysels, Plazzi, Valkanov, Rubia, and Dossani \(2019\)](#), yet in either setting to produce all one step ahead forecasts for the last 2,000 observations and to tune the hyper-parameters can be computationally challenging. Instead, we follow [Giacomini and Komunjer \(2005\)](#) and perform a fixed forecast window exercise, in which we estimate our models once.

<sup>5</sup>The split is approximately equal to 60%, 20%, 20%. We have examined alternative training, validation and test splits, which give similar patterns to the presented empirics and are available upon request.

For our forecasting design we use a fixed forecast window exercise and predict the ten-day-ahead  $VaR$  as:

$$\widehat{VaR}_{1,t+10}|\mathcal{F}_t = G_\tau(\sigma_{1,t}, \mathbf{w}^*), \quad (4.21)$$

where  $\mathcal{F}_t$  denotes the information set up to time  $t$ ,  $\mathbf{w}^*$  denotes the optimal weights obtained from the CV. Eq. (4.21) illustrates how forecasts for the first  $VaR$  specification were obtained via the *deep quantile* estimator. In a similar manner forecasts can be obtained for other  $VaR$  specifications and alternative models, using eq. (4.12) – (4.20).

We evaluate the forecasting performance of  $VaR$  models with the proposed *deep quantile* estimator as in Section 4.2. Further, we consider ten-day compounded  $VaR$  forecasts, which we relegate to the Online Appendix.

### 4.4.3 Forecast Evaluation

In this Section we discuss the various tests we have considered, in order to evaluate the predictive ability of the *deep quantile* estimator and present the testing results.

#### 4.4.3.1 Diebold Mariano Test

We perform a quantitative forecast comparison across different methods and test their statistical significance. To do so, we calculate the *Root Mean Squared Forecast error* (RMSFE) for each method and perform the [Diebold and Mariano \(1995\)](#) (DM) test, with the [Harvey, Leybourne, and Newbold \(1997\)](#) adjustment to gauge the statistical significance of the forecasts. With the DM test, we assess the forecast accuracy of the *deep quantile* estimator relative to the benchmark linear quantile regression model. In this exercise we set  $\tau$  equal to 1%, 5% and 10%.

In general, RMSFE is used to measure the accuracy of point estimates and is defined as

$$\text{RMSFE} = \sqrt{\frac{\sum_{t=1}^T (y_{t+h} - \widehat{G}_\tau(\mathbf{x}_{t+h}, \mathbf{w}))^2}{T}},$$

where  $h$  denotes the forecasting horizon and  $\widehat{G}_\tau(\mathbf{x}_{t+h}, \mathbf{w})$  is the solution to the eq. (4.8) after selecting the optimal  $\mathbf{w}$  via CV at the  $\tau^{\text{th}}$  quantile. Results from the DM test are reported in Table 4.1, where asterisks denote the statistical significance of rejecting the null hypothesis of the test at 1%, 5% and 10% level of significance, for all quantiles and

models we consider. These results suggest that forecasts produced from the non-linear estimator outperform, for the majority of cases, forecasts obtained from the linear and non-parametric quantile regression estimators.

#### 4.4.3.2 Giacomini White Test

In a similar manner and to complement the DM test, we follow [Carriero, Kapetanios, and Marcellino \(2009\)](#) and further calculate the [Giacomini and White \(2006\)](#) test of equal forecasting accuracy, that can handle forecasts based on both nested and non-nested models, regardless of the estimation procedures used for the derivation of the forecasts, including our proposed *deep quantile* estimator. Table 4.1 illustrates the results for [Giacomini and White \(2006\)](#) test, where daggers denote the statistical significance of rejecting the null hypothesis of the test at 1%, 5% and 10% level of significance, for all quantiles and different models we consider. Similarly to the DM forecasting accuracy test, the [Giacomini and White \(2006\)](#) test is again significant at 1% in most cases, with the following exceptions.

Quantile polynomial regression forecasts are only significant at the 10% level of significance for SV model. In quantile splines, forecasts for the GARCH specification at  $\tau = 5\%$  and RM at  $\tau = 10\%$  are not significant. Forecasts from the linear MIDAS, under the GARCH specification, at  $\tau = 1\%$  are insignificant and under the ASV specification, at  $\tau = 5\%$ , are significant at the 5% significance level. Results for the ASV with *Deep Ridge* estimator at  $\tau = 1\%$  are significant only for the [Giacomini and White \(2006\)](#) test. For the ASV *deep MIDAS Ridge* estimator and at  $\tau = 1\%$ , the forecasts are significant only based on the DM test. Forecasts from *deep Elnet* model under SV specification and at  $\tau = 5\%$  are significant at 5% level of significance. Finally, forecasts from *deep Elnet* under SV specification and  $\tau = 5\%$  are significant at the 5% level of significance.

Overall, results from both the DM and [Giacomini and White \(2006\)](#) tests suggest that the non-linear estimators outperform, for the majority of cases, competing linear and non-parametric estimators in *VaR* forecasting.

Table 4.1 about here

### 4.4.3.3 Conditional Quantile Forecast Encompassing (CQFE)

We present the implementation of the CQFE test as proposed by [Giacomini and Komunjer \(2005\)](#) and the Generalized Method of Moments (GMM) estimation as proposed by [Hansen \(1982\)](#). Let  $\hat{q}_{1,t}$  be a vector of the  $\tau^{th}$  quantile forecasts produced from model 1 and  $\hat{q}_{2,t}$  be the competing forecasts produced from model 2. The basic principle of CQFE is to test whether  $\hat{q}_{1,t}$  conditionally encompasses  $\hat{q}_{2,t}$ . Encompassing occurs when the second set of forecasts fails to add new information to the first set of quantile forecasts (or vice versa) in which case the first (second) quantile forecast is said to encompass the second (first).

The aim of the CQFE test is to test the null hypothesis, that  $\hat{q}_{1,t}$  performs better than any linear combination of  $\hat{q}_{1,t}$  and  $\hat{q}_{2,t}$ . Under the null hypothesis, it holds

$$E_t(\rho_\tau(y_{t+1} - \hat{q}_{1,t})) \leq E_t(\rho_\tau(y_{t+1} - \theta_0 - \theta_1 \hat{q}_{1,t} - \theta_2 \hat{q}_{2,t})), \quad (4.22)$$

that is satisfied if and only if the weights  $(\theta_1, \theta_2)$  are equal to  $(1, 0)$ . The objective function of the GMM is:

$$J_T = g_T(\boldsymbol{\theta})' \mathbf{W}_T g_T(\boldsymbol{\theta}).$$

The optimal weights are computed as:

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} g_T(\boldsymbol{\theta})' \mathbf{W}_T g_T(\boldsymbol{\theta}), \quad g_T(\boldsymbol{\theta}) = \frac{\sum_{t=1}^T (\tau - \mathbb{1}_{\tau}\{y_{t+1} - \hat{q}_t < 0\}) z_T}{T},$$

where  $\mathbf{W}_T$  is a positive definite matrix,  $g_T(\boldsymbol{\theta})$  is the sample moment condition,  $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)'$  is a set of weights,  $\boldsymbol{\theta}^* = (\theta_0^*, \theta_1^*, \theta_2^*)'$  denotes the optimal weights,  $\hat{q}_t = (1, \hat{q}_{1,t}, \hat{q}_{2,t})'$  is a vector with the forecasted values based on the pairwise models 1, and 2 in the CQFE test,  $m$  denotes the out-of-sample size and  $z_T$  is a vector of instruments. [Hansen \(1982\)](#) showed that by setting  $\mathbf{W}_T = \mathbf{S}_T^{-1}$  i.e the inverse of an asymptotic covariance matrix, is optimal as it estimates  $\boldsymbol{\theta}^*$  with as small as possible asymptotic variance.  $\mathbf{S}$  is also known as the spectral density matrix of  $g_T$ . We follow [Newey and West \(1987\)](#) and use a heteroskedasticity robust estimate  $\hat{\mathbf{S}}_T$ , of  $\mathbf{S}$  defined as:

$$\hat{\mathbf{S}}_T = \hat{\mathbf{S}}_0 + \sum_{j=1}^m \left(1 - \frac{j}{m+1}\right) (\hat{\mathbf{S}}_j + \hat{\mathbf{S}}_j'), \quad \text{where} \quad \hat{\mathbf{S}}_j = \frac{1}{T} \sum_{t=j+1}^T g_t(\hat{\boldsymbol{\theta}}) g_{t-j}(\hat{\boldsymbol{\theta}}).$$

$\hat{\mathbf{S}}_0$  is the estimated spectral density matrix evaluated at frequency zero. The GMM estimation is performed recursively, i.e. i) minimize  $J_T$  using an identity weighting

matrix to get  $\theta^*$ , which gives  $\mathbf{W}_T$  via  $\widehat{\mathbf{S}}_T$  and ii) minimize  $J_T$  using  $\mathbf{W}_T = \widehat{\mathbf{S}}_T^{-1}$  from step i).

Consequently, we consider two separate test  $H_{10} : (\theta_1^*, \theta_2^*) = (1, 0)$  versus  $H_{1a} : (\theta_1^*, \theta_2^*) \neq (1, 0)$  and  $H_{20} : (\theta_1^*, \theta_2^*) = (0, 1)$  versus  $H_{2a} : (\theta_1^*, \theta_2^*) \neq (0, 1)$ , which correspond to testing whether forecast  $\widehat{q}_{1,t}$  encompasses  $\widehat{q}_{2,t}$  or  $\widehat{q}_{2,t}$  encompasses  $\widehat{q}_{1,t}$ . Then the CQFE statistics are defined as:

$$\begin{aligned} \text{ENC}_1 &= T((\theta_1^*, \theta_2^*) - (1, 0))' \widehat{\mathbf{\Omega}}((\theta_1^*, \theta_2^*) - (1, 0))' \\ \text{ENC}_2 &= T((\theta_1^*, \theta_2^*) - (0, 1))' \widehat{\mathbf{\Omega}}((\theta_1^*, \theta_2^*) - (0, 1))', \end{aligned}$$

where  $\widehat{\mathbf{\Omega}} = g_T(\theta)' \mathbf{S}^{-1} g_T(\theta)$ . The asymptotic distribution of the GMM estimates of  $\theta$  requires the moment conditions to be once differentiable. To satisfy this requirement, we follow [Giacomini and Komunjer \(2005\)](#) and replace the moment condition with the following smooth approximation:

$$g_\tau(\theta) = \frac{\sum_{t=1}^T [\tau - (1 - \exp((y_{t+1} - \theta' \widehat{q}_t) / \eta))] \mathbb{1}\{y_{t+1} - \widehat{q}_t < 0\} z_T}{T},$$

where  $\eta$  is the smoothing parameter. We choose the critical values,  $c_{crit}$  of the test from a  $\chi_2^2$  distribution, in which  $\widehat{q}_{i,t}$  encompasses  $\widehat{q}_{j,t}$ , if  $\text{ENC}_i \leq c_{crit} \forall i \neq j = 1, 2$ . In the empirical application, the vector of instruments,  $z_T$ , is  $(1, r_t, \text{VaR}_{i,t}, \text{VaR}_{j,t})$ ,  $\forall i \neq j = 1, 2$ .

We select  $\eta$  to be 0.005, following the CQFE test rejection probabilities in [Giacomini and Komunjer \(2005\)](#), since our POOS size is 2,000 observations. We consider the following five blocks: i) the non-parametric, ii) the non-linear, iii) the non-linear MIDAS, iv) the linear and v) the linear MIDAS blocks. The non-parametric block consists of the quantile polynomial and quantile splines estimators, the non-linear block consists of the *deep quantile* estimators for the different regularization schemes and the non-linear MIDAS block consists of the *deep MIDAS* estimators for the different regularization schemes. Finally, the linear and linear MIDAS blocks consist of the linear quantile and linear quantile MIDAS estimators, respectively.

We examine each block of models across different quantiles. Specifically, we consider how many times the models within a specific block outperform models from other blocks and present these results in [Table 4.2](#). Under this setting a *win* denotes that the prevailing model encompasses the competing benchmark model, while a *loss* means



that the competing model encompasses the prevailing one. Precisely, we consider a *win* when the computed p-value of the CQFE test fails to reject the null hypothesis, i.e.  $H_{10}$  or  $H_{20}$ . On the contrary, in the case where the CQFE test suggests that there is no encompassing between the forecasts, we consider this as a *loss*, i.e. the null hypothesis is rejected. Furthermore, the CQFE test has a gray zone in which the test can fail to reject both null hypotheses ( $H_{10}$  and  $H_{20}$ ), hence the test is inconclusive. Below we summarise the CQFE testing results for the different quantiles when  $\eta = 0.005$ .

For the 10<sup>th</sup> quantile, the non-linear block encompasses 660 times the competing blocks, in comparison to the linear block, which encompasses the competing blocks 165 times and the non-parametric block that encompasses the others 320 times. The linear block does not encompass other blocks less than 23 times and the non-linear block for 139 times. Additionally, the test is inconclusive 643 times for the non-linear block and 149 times for the linear one. Thus, the non-linear block is ranked first in terms of how many times it encompasses the other blocks and the non-linear MIDAS block is ranked second.

For the 5<sup>th</sup> quantile, the non-linear block encompasses 711 times other blocks, 333 times the non-parametric and the linear 177 times. Further, the linear block does not encompass the other blocks 11 times and the non-linear 88 times. Finally, for the non-linear block, the CQFE test is inconclusive 702 times and 167 times for the linear block. The ranking of the first two blocks is the same as in the 10<sup>th</sup> quantile.

Finally, we examine the 1<sup>st</sup> quantile. In this case, the non-linear block encompasses 723 times the other blocks, 341 times the non-parametric and the linear block 171 times. Furthermore, the linear block does not encompass 17 times the other blocks and the non-linear 76 times. The test is inconclusive 715 times for the non-linear block and 165 times for the linear one. The ranking remains the same as above. Results for different smoothing parameters  $\eta$  suggest similar patterns and are available upon request.

Table 4.2 about here

## 4.5 Semi-Structural analysis

A general issue in ML is the trade-off between accuracy and interpretability; where the output of a highly complicated model, e.g. a deep neural network, can have great

accuracy or forecasting performance, but cannot be easily interpreted. In this Section we first discuss the details of two methods that can be used to make ML methods interpretable. The first one is the Shapley Additive Explanation Values (SHAP), that has received a lot of attention recently, and the second is partial derivatives. Further we make a formal comparison on the output of both methods, based on the output of the *deep quantile* estimator that illustrates, i) that both methods can be used to make the impact of each covariate in neural networks interpretable and ii) perhaps surprisingly that the use of partial derivatives, offers more stable results at a fraction of the computational cost.

### 4.5.1 Shapley values

Shapley values (SHAP) are a general class of additive attribution methods, based on the initial work of [Shapley \(1953\)](#) where the goal was to determine how to fairly split a pay-off among players in a cooperative game. In the context of ML, the goal of SHAP values is to explain the prediction of the dependent variable by estimating the contribution of each covariate to the prediction. SHAP values, following the exposition in [Lundberg and Lee \(2017\)](#) and [Lundberg, Erion, and Lee \(2018\)](#) can be constructed as follows.

Let  $f(\mathbf{x}_t) = \widehat{G}(\mathbf{x}_t, \mathbf{w})$  be the output of the estimated model we wish to interpret, given a  $p \times 1$  vector of covariates  $\mathbf{x}_t$ , and  $\widehat{f}$  the explanation model, to be defined below. Further, let  $\mathbf{x}_t^\dagger$  be the  $M \times 1$  subset (vector) of  $\mathbf{x}_t$  that contains simplified covariates. These simplified covariates, can be mapped to the original through a mapping function  $h_{\mathbf{x}_t}(\cdot)$ , such that  $\mathbf{x}_t = h_{\mathbf{x}_t}(\mathbf{x}_t^\dagger)$ . Then under the local accuracy property of [Lundberg and Lee \(2017\)](#), if there exists a vector,  $\mathbf{z}_t^\dagger$ , with binary inputs, such that  $\mathbf{z}_t^\dagger \approx \mathbf{x}_t^\dagger$ , then  $\widehat{f}(\mathbf{z}_t^\dagger) \approx f(h_{\mathbf{x}_t}(\mathbf{z}_t^\dagger))$ , where the explanation model (i.e. the additive attribution function) is

$$\widehat{f}(\mathbf{z}_t^\dagger) = \phi_0 + \sum_{i=1}^M \phi_i z_{t,i}^\dagger, \quad (4.23)$$

and  $\widehat{f}(\mathbf{z}_t^\dagger)$  represents the linear decomposition of the original ML model, where  $\phi_0$  is the intercept,  $\phi_i \in \mathbb{R}$  is the effect to each dependent variable  $z_{t,i}^\dagger \in (0, 1)$ , that provides local and global inference at the same time. If  $z_{t,i} = 1$  then the covariate is observed, on the contrary, if  $z_{t,i} = 0$  then the covariate is unknown. Under the following three properties: i) local accuracy i.e. the explanation function should match the original model, ii) missingness, which ensures that input variable have no attributed effect

and iii) consistency, under which, if an input variables is important, then the effect to each dependent variable should not decline, the SHAP value is

$$\phi_i = \sum_{M \subseteq p \setminus \{i\}} \frac{|M|!(p - |M| - 1)!}{p!} [f_{M \cup \{i\}}(x_{M \cup \{i\}}) - f_M(x_M)], \quad (4.24)$$

where  $p$  is the set of all predictors,  $|M|$  is the number of non-zero elements in  $x_i^\dagger$ ,  $f_M(x_M)$  is the model's output using except from the  $i^{\text{th}}$  covariate, and  $f_{M \cup \{i\}}(x_{M \cup \{i\}})$  is the output of the model, when  $\{i\}$  is included in the covariate set.

The calculation of SHAP values can be computationally expensive, as it requires  $2^N$  possible permutations of the predictors. For the case of deep neural networks [Lundberg and Lee \(2017\)](#), and [Shrikumar, Greenside, and Kundaje \(2017\)](#), have shown that *DeepLIFT* can be used as an approximation of the deep SHAP that is computationally feasible<sup>6</sup>, preserving the three properties above. *DeepLIFT* is a recursive prediction explanation method for deep learning. The Additive feature attribution methods analogy of *DeepLIFT* is called the summation-to-delta property is

$$\sum_{i=1}^p C_{\Delta x_{t,i} \Delta o} = \Delta o. \quad (4.25)$$

Then the SHAP values can be obtained as

$$\phi_i = C_{\Delta x_{t,i} \Delta o},$$

where  $C_{\Delta x_{t,i} \Delta o}$ , represents the impact of a covariate to a reference value relative to the initial value, is assigned to each  $x_{t,i}$  covariate,  $o = f(\cdot)$  is the output of the model,  $\Delta o = f(x) - f(r)$ ,  $\Delta x_{t,i} = f(x_{t,i}) - r_{t,i}$  and  $r$  the reference value. Eq. (4.25) matches eq. (4.23), if in  $\Delta o$  we set  $\phi_0 = f(r_{t,i})$  and  $\phi_i = C_{\Delta x_{t,i} \Delta o}$ .

## 4.5.2 Partial Derivatives

The use of partial derivatives for the interpretation of a model is straight forward in econometrics, with various uses, ranging from the simple linear regression model to

<sup>6</sup>There are other methods that can be used to achieve this, such as Tree Explainer, Kernel Explainer, Linear Explainer, Gradient Explainer.

impulse response analysis. In this Section we show how partial derivatives can be used even in highly non-linear deep neural networks. Before we start the analysis, note that while the deep neural networks are highly non-linear, their solution/output via SGD optimization methods, can be treated as differentiable function, as the majority of activation functions are differentiable. Let's consider the case of ReLU, that is not differentiable at 0, whereas it is in every other point. From the point of gradient descent, heuristically, it works well enough to treat it as a differentiable function. Further, [Goodfellow, Bengio, and Courville \(2016\)](#) argue that this issue is negligible and ML softwares are prone to rounding errors, which make it very unlikely to compute the gradient at a singularity point. Note that even in this extreme case, both *SGD* and *ADAM*, will use the right subgradient at 0.

For a general  $\mathbf{x}_t \in \mathbb{R}^p$ , let

$$d_{j,i,t} = \frac{\partial \widehat{G}_{j,\tau}(\mathbf{x}_t, \mathbf{w})}{\partial x_{j,i,t-1}}, \quad (4.26)$$

denote the partial derivative of covariate  $x_i = x_{it}$ , for  $i = 1, \dots, p$  at time  $t = 1, \dots, T$ ,  $\widehat{G}_{j,\tau}(\mathbf{x}_t, \mathbf{w})$  is the forecasted  $VaR_{j,t}$ , across the  $j$  different  $VaR$  specifications we consider. We assess the partial derivative in time, since, following [Kapetanios \(2007\)](#), we expect it to vary in time, due to the inherent non-linearity of the neural network. Our covariate(s)  $\mathbf{x}_t$  are the conditional volatility for GARCH and RM,  $VaR$  lagged values, the absolute *S&P500* daily return and the positive and negative *S&P500* daily returns for SV and ASV, respectively. It is evident that under the classic linear regression problem, or linear quantile regression model, the effect of the covariates  $\mathbf{x}_t$  in the dependent variable  $y_t$  is constant, time invariant, and corresponds to  $\widehat{\beta}(\tau)$ .

### 4.5.3 Results

In this application we use the whole sample size i.e. around 36 years of daily returns on the *S&P500* index to provide an accurate interpretation of our *deep quantile* estimator. Figures 4.5 – 4.8 illustrate the partial derivatives and SHAP values evaluated in time on the output of our *deep quantile*<sup>7</sup> estimator, for a specific quantile  $\tau$ . Further, we compare the partial derivatives of the *deep quantile* estimator relative to the linear quantile regression partial derivative, i.e. the  $\beta(\tau)$  coefficient. Both partial derivatives

<sup>7</sup>In this Section we limit our attention in the output of the best performing model, in terms of its forecasting capacity, as reflected by the forecast gains measure in Section 4.4, for each model, based on the different penalisation schemes. Results from all the different penalisation schemes suggest similar patterns to the ones discussed above and are available upon request.

and SHAP values seem to identify interesting patterns that can be linked to some well known events. Below we discuss our results for all models we have considered in our empirical application.

The results for the first two models, i.e. *GARCH* and *RM* can be summarised together, since in both models there is only one covariate, that is the conditional volatility, but with a different specification. The results from this model are illustrated in Figures 4.5. We find that the partial derivative appears to be more stable over time, fluctuating around the constant partial derivative,  $\beta(\tau)$ , of the linear quantile estimator. When there is a crisis or a stressful event in the financial markets, they increase. As an example, we see significant spikes in the partial derivatives, both in March 2020 as well as in 2008, which stand for the onset of the COVID-19 pandemic and the *Great Recession* respectively. We also find that the biggest increase occurs in 1987, the year when *Black Monday* happened, and also significant variation during the U.S. government shutdown in 2019. The values for the partial derivatives generally increase, as  $\tau$  decreases. SHAP values have a similar behaviour with the partial derivatives, but are more volatile across time. For the first two models, there are some events, e.g. during the 1991, where the values for both SHAP and partial derivatives do not increase a lot. We view this finding as an inability of these two models, to properly account for this crisis.

In the last two models, the merit of SHAP values and partial derivatives becomes clear, since in these models we have more than one covariates and both methods can provide an indication on the effect of each covariate on the final output. Overall, we find that increasing the number of covariates, allow these models to account for all crises within the sample. For the case of the *SV* model, we find that the important covariate is the lagged values of *VaR*, rather than the absolute values of *S&P500*. Similar to the one covariate models, we find that the partial derivatives are more stable than SHAP values, fluctuating closely around  $\beta(\tau)$  and picking up when there are crisis or distress in the economy or financial markets. The SHAP values again appear to be more volatile with a wider range. Similar to the findings of the one covariate models, the higher the values for the partial derivative and SHAP, the lower the  $\tau$  quantile.

For the case of the *ASV* model, we find that again the lagged values of *VaR* is the most significant covariate, the negative *S&P500* returns have some impact and the positive *S&P500* returns are almost insignificant. Similar to the cases above, we find that the partial derivative is more stable than SHAP values, fluctuating closely around  $\beta(\tau)$

and picking up when there is a crisis or distress in the economy or financial markets. The SHAP values again appear to be more volatile with a wider range. Again and same as before, lower quantiles have higher partial derivatives. The results for these two models are illustrated in Figures 4.6, 4.7 and 4.8.

Different penalization schemes maintain the aforementioned results, with a lower magnitude. Overall, we observe that the linear quantile regression shows a fixed pattern across time and it is evident that this model does not anticipate shocks in the economy. As [Engle and Manganelli \(2004\)](#) suggest, *SV* and *ASV* react more to negative shocks and in stressful events their spike is larger than the *GARCH* and *RM* models. Finally, covariates with the minimum contribution on the forecasted values, such as the positive *S&P500* returns have negligible impact on both SHAP and partial derivatives values.

Figures 4.5 – 4.8 about here

## 4.6 Conclusion

In this Chapter we contribute to the expanding literature on the use of ML in finance and propose a *deep quantile* estimator that has the potential to capture the non-linear association between asset returns and predictors. In Section 4.2, we lay out the exact workings of our proposed estimator, and illustrate how it generalises linear quantile regression.

In the Monte Carlo exercise in Section 4.3, we study the finite sample properties of the *deep quantile* estimator, based on a number of data generating processes. We present extensive evidence the estimator gives good finite sample performance, that is a function of  $T$ , uniformly across different regularization schemes.

We use the *deep quantile* estimator, with various penalization schemes, to forecast *VaR*. We find that our estimator gives considerable predictive gains, up to 74%, relative to the *VaR* forecasts produced by the linear quantile regression. This result is backed by the forecasting accuracy tests, i.e. the [Diebold and Mariano \(1995\)](#), the [Giacomini and White \(2006\)](#) and the quantile score tests. Further, results from the CQFE test of [Giacomini and Komunjer \(2005\)](#) suggest that forecasts obtained from the non-linear estimators encompass forecasts from the linear and non-parametric models with a higher

frequency. These findings are in support of the non-linear association between the conditional quantile of asset returns and covariates, hence suggesting a new avenue in forecasting in finance and in macroeconomics during extreme events.

In addition, we do a semi-structural analysis to examine the contribution of the predictors in *VaR* over time. We consider, following the ML literature, SHAP values and further partial derivatives. Our findings suggests that our non-linear estimator reacts more in stressful events and exhibits time-variation, while the linear quantile estimator presents, as expected, a constant time invariant behaviour. We conclude that financial variables are characterised by non-linearities, that our proposed *deep quantile* estimator can approximate quite well.

Finally, we make a formal comparison between SHAP and partial derivatives, and interestingly find that partial derivatives can be used to make ML methods interpretable, are less volatile, easier to interpret and can be computed at a fraction of time used in the calculation of SHAP values.

## **A.1 Appendix (Chapter 4)**



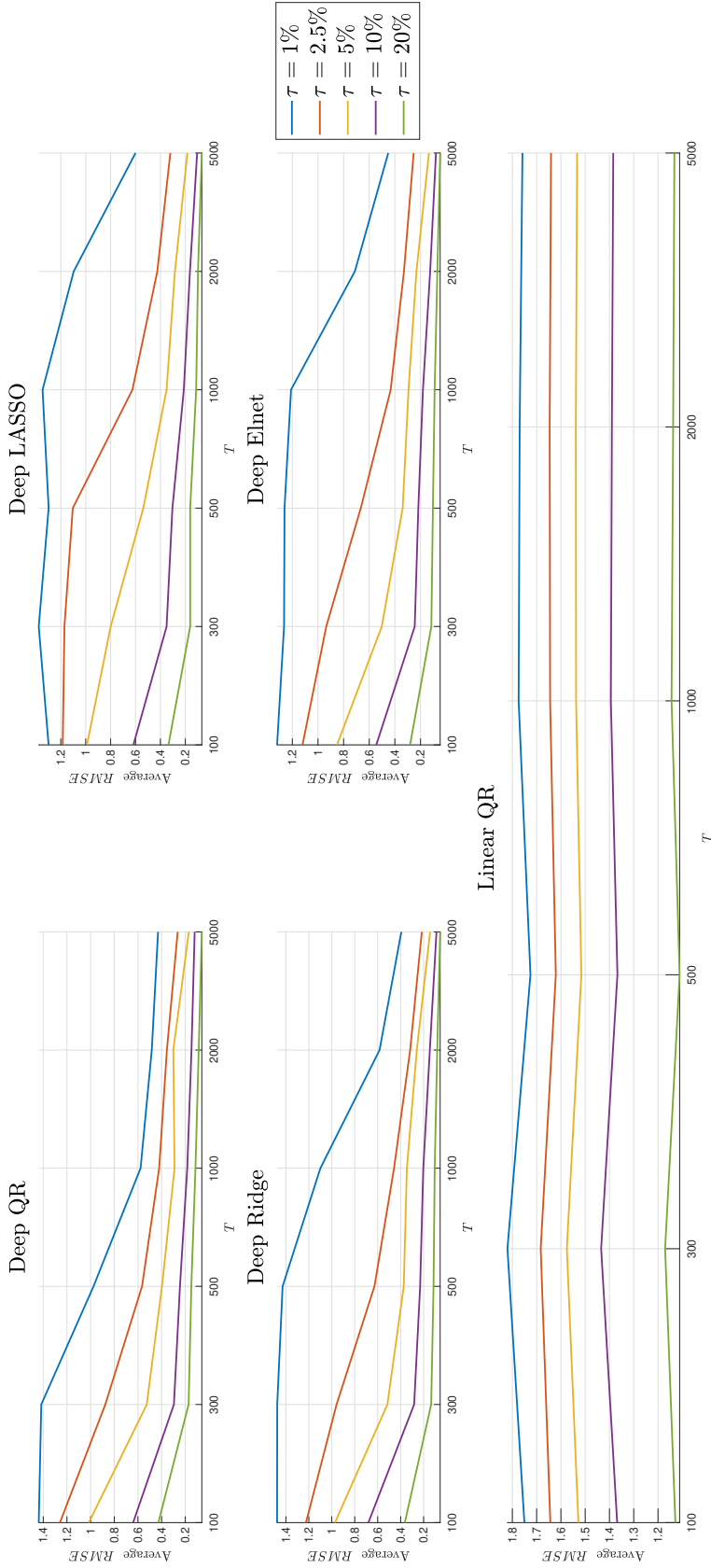


FIGURE 4.1: Monte Carlo results for Case I. Model:  $y_t = h_\tau(x_t) + u_t$ ,  $h_\tau(x_t) = \sin(2\pi x_t)$ ,  $x_t \sim N(0,1)$ ,  $u_t \sim iidN(-\sigma\Phi^{-1}(\tau), \sigma^2)$ ,  $\sigma = 0.1$  and  $\Phi^{-1}$  is the quantile function of the standard normal distribution. Figure presents the average mean squared error of the estimated residuals,  $AMSE_{\hat{u}_t, pen}$  for the different penalization schemes (Model),  $T = 100, 300, 500, 1000, 2000, 5000$  and different quantiles,  $\tau = (1\%, 2.5\%, 5\%, 10\%, 20\%)$ .

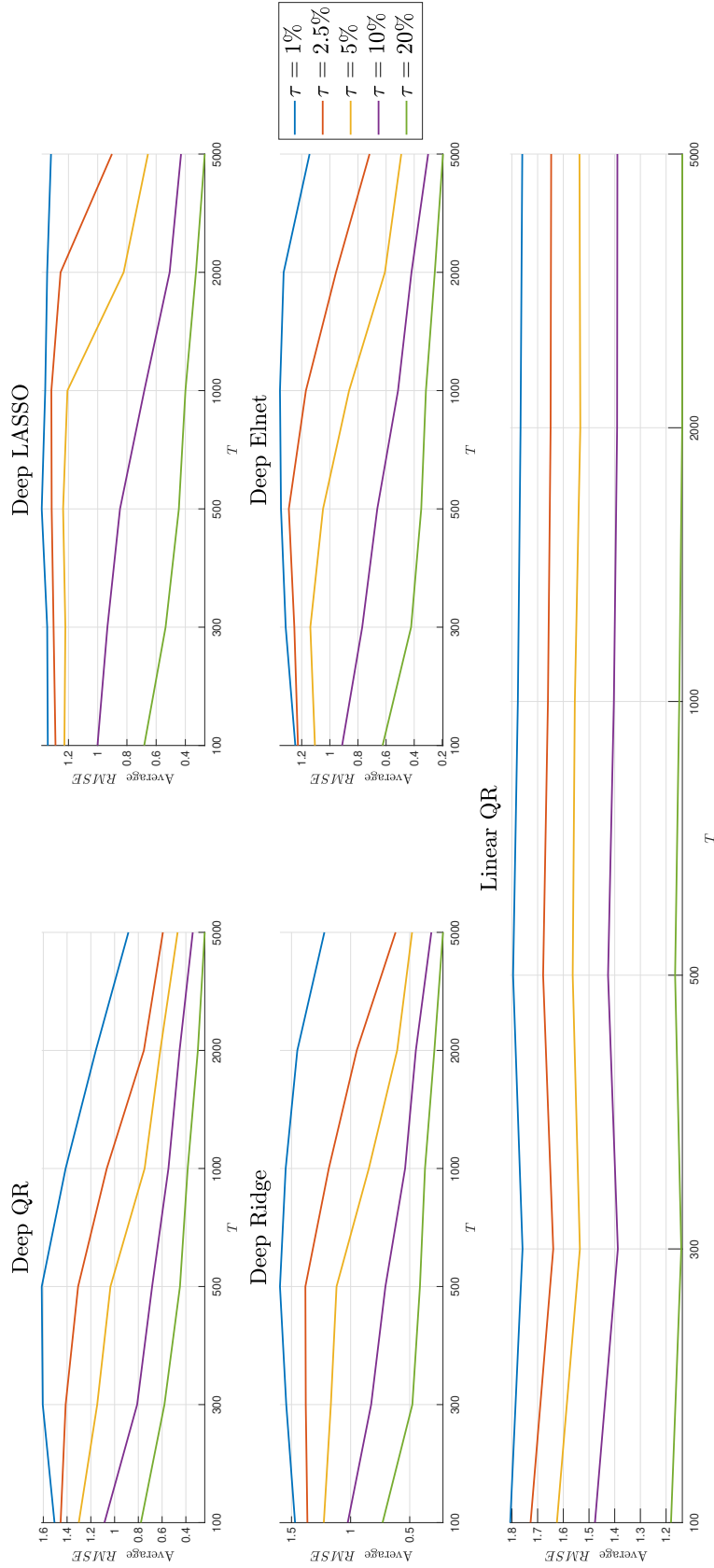


FIGURE 4.2: Monte Carlo results for Case II. Model:  $y_t = h_\tau(x_t) + u_t$ ,  $h_\tau(x_t) = \sin(2\pi x_t)$ ,  $x_t = 0.8x_{t-1} + \varepsilon_t$ ,  $\varepsilon_t \sim N(0, 1)$ ,  $u_t \sim iidN(-\sigma\Phi^{-1}(\tau), \sigma^2)$ ,  $\sigma = 0.1$  and  $\Phi^{-1}$  is the quantile function of the standard normal distribution. Figure presents the average mean squared error of the estimated residuals,  $AMSE_{iid, pen}$  for the different penalization schemes (Model),  $T = 100, 300, 500, 1000, 2000, 5000$  and different quantiles,  $\tau = (1\%, 2.5\%, 5\%, 10\%, 20\%)$

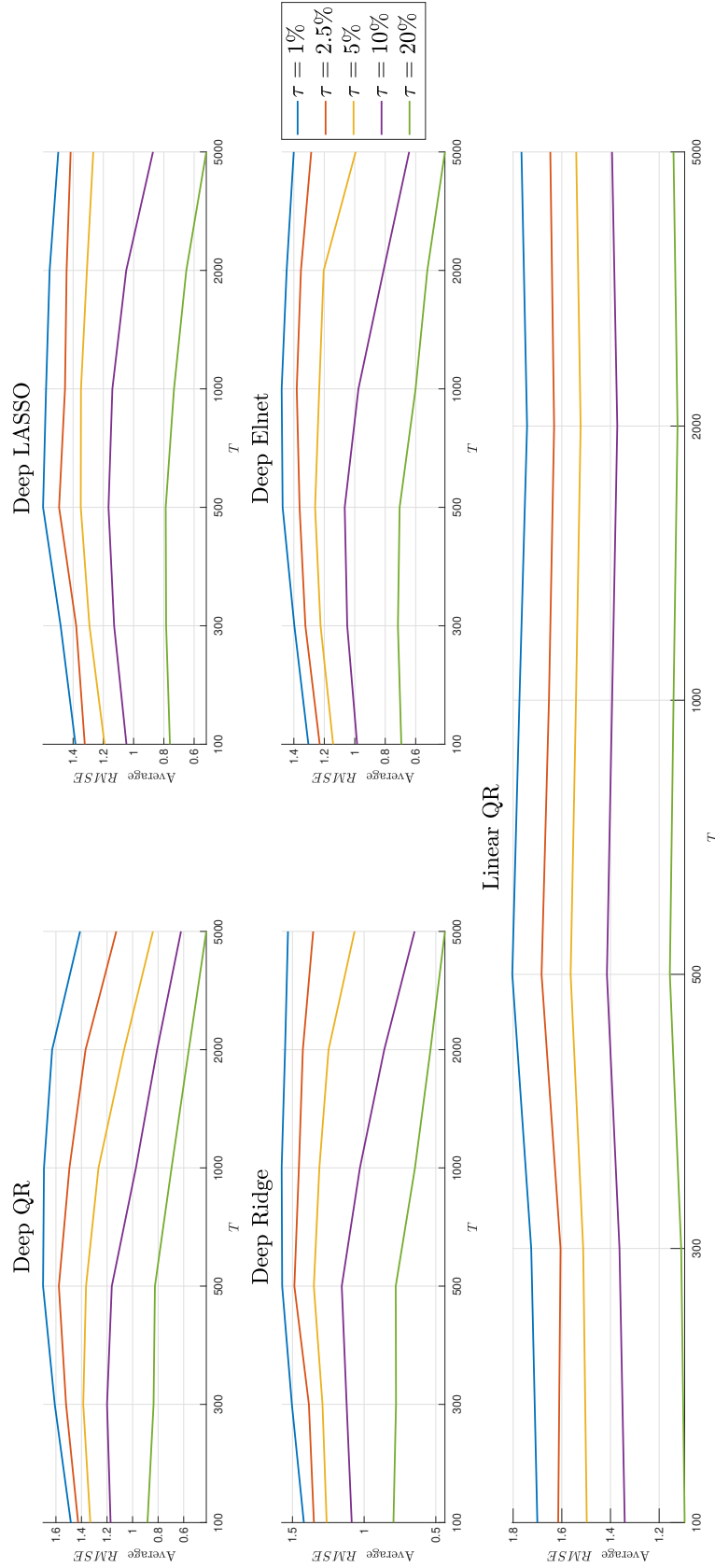


FIGURE 4.3: Monte Carlo results for Case III. Model:  $y_t = h_\tau(x_t) + u_t$ ,  $h_\tau(x_t) = \sin(2\pi x_t)$ ,  $x_t = \sigma_t \varepsilon_t$ ,  $\sigma_t^2 = 1 + 0.7x_{t-1}^2 + 0.2\sigma_{t-1}^2$ ,  $\varepsilon_t \sim N(0,1)$ ,  $u_t \sim iidN(-\sigma\Phi^{-1}(\tau), \sigma^2)$ ,  $\sigma = 0.1$  and  $\Phi^{-1}$  is the quantile function of the standard normal distribution. Figure presents the average mean squared error of the estimated residuals,  $AMSE_{\hat{u}_t, pen}$  for the different penalization schemes (Model),  $T = 100, 300, 500, 1000, 2000, 5000$  and different quantiles,  $\tau = (1\%, 2.5\%, 5\%, 10\%, 20\%)$

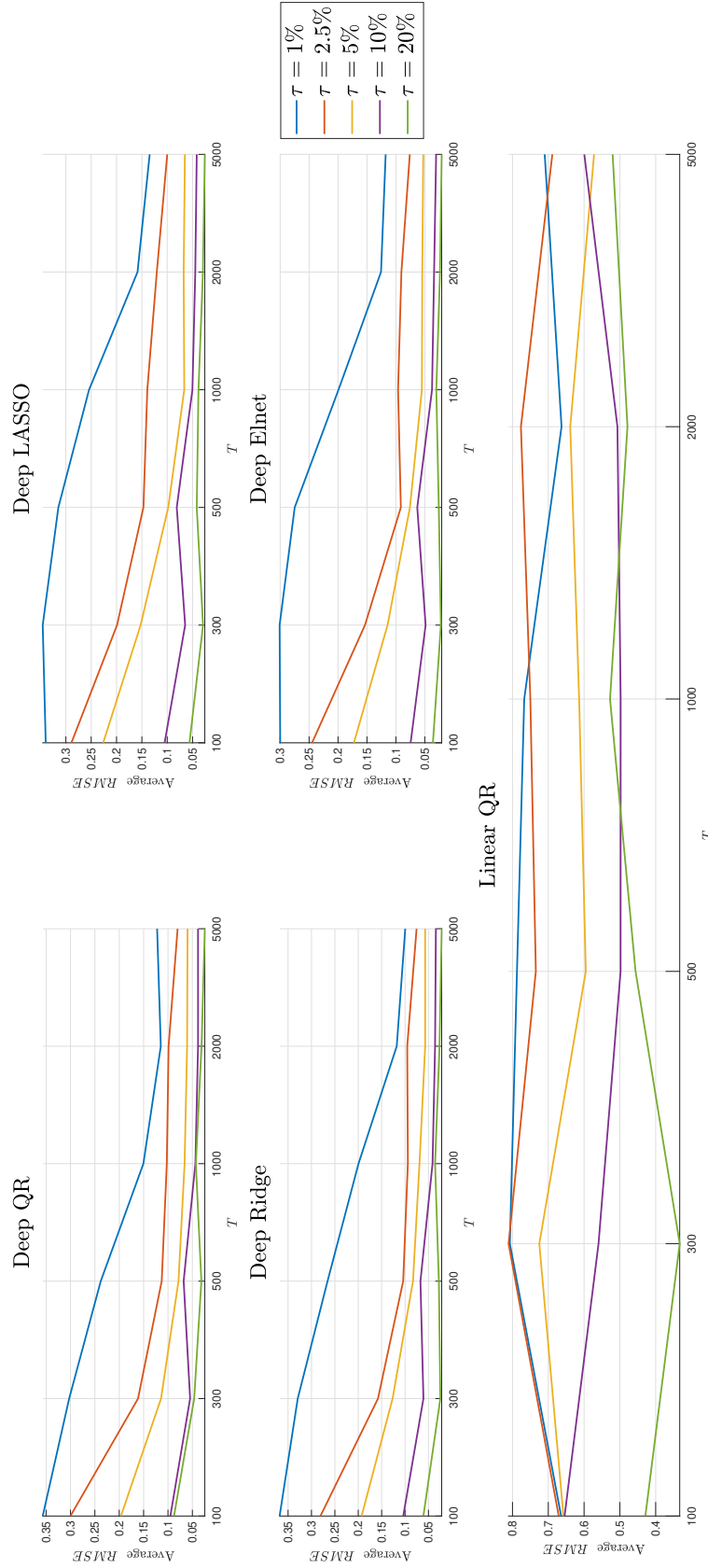


FIGURE 4.4: Monte Carlo results for Case IV. Model:  $y_t = h_\tau(x_t) + u_t$ ,  $h_\tau(x_t) = G(x_t, \boldsymbol{w})$ ,  $x_t \sim N(0, 1)$ ,  $w_{i,j} = \delta_{i,j} \mathbb{1}(\delta_{i,j} > 0.1)$ ,  $\delta_{i,j} \sim U(0, 1)$ ,  $u_t \sim iidN(-\sigma\Phi^{-1}(\tau), \sigma^2)$ ,  $\sigma = 0.1$  and  $\Phi^{-1}$  is the quantile function of the standard normal distribution. Figure presents the average mean squared error of the estimated residuals,  $AMSE_{\hat{u}_t, pen}$  for the different penalization schemes (Model),  $T = 100, 300, 500, 1000, 2000, 5000$  and different quantiles,  $\tau = (1\%, 2.5\%, 5\%, 10\%, 20\%)$

Model	$\tau$	Polynomial	Spline	MIDAS	Deep QR	Deep MIDAS	Deep LASSO	Deep MIDAS LASSO	Deep Ridge	Deep MIDAS Ridge	Deep Elnet	Deep MIDAS Elnet
GARCH	1%	0.808	0.631***/+++	1.015	0.572***/+++	0.454***/+++	0.723***/+++	0.502***/++	0.550***/+++	0.554***	0.410***/+++	0.731***/+++
	5%	1.030	0.919	1.230***/+++	0.659***/+++	0.504***/+++	0.674***/+++	0.852***/+++	0.793***/+++	0.600***/+++	0.548***/+++	0.541***/+++
	10%	1.103	1.161***/++	1.276***/+++	0.552***/++	0.563***/+++	0.655***/++	0.644	0.598***/++	0.586***/+++	0.667***/+++	0.598***/++
RM	1%	0.771	0.503***/+++	1.059***/+++	0.366***/+++	0.399***/+++	0.346***/+++	0.399***/+++	0.261***/+++	0.242***/+++	0.386***/+++	0.399***/+++
	5%	0.918	0.704***/+++	1.195***/+++	0.371***/+++	0.302	0.629***/+++	0.333***/+++	0.415***/+++	0.497***/+++	0.396***/+++	0.818***/+++
	10%	1.089	1.011	1.367***/+++	0.433***/+++	0.422***/+++	0.500***/+++	0.433***/+++	0.511***/+++	0.367***/+++	0.433***/+++	0.422***/+++
SV	1%	0.988	1.308***/+++	0.605***/+++	0.661***/+++	0.631***/+++	0.655***/+++	0.643***/+++	0.732***/+++	0.756***/+++	0.679***/+++	0.637***/+++
	5%	0.992	1.473***/+++	0.645***/+++	0.629***/+++	0.702	0.698***/+++	0.722***/+++	0.645***/+++	0.576***/+++	0.812***/+++	0.669***/+++
	10%	0.986*	1.175***/++	0.650***/+++	0.657***/+++	0.664	0.762***/+++	0.699***/+++	0.720***/+++	0.748***/+++	0.783***/+++	0.734***/+++
ASV	1%	0.998	1.030***/+++	3.524***/++	0.671***/+++	0.695***/+++	0.709***/+++	0.651***/+++	0.884	0.787**	0.815***/+++	0.647***/+++
	5%	0.997	1.024***/+++	3.313***/+	0.648***/+++	0.645***/+++	0.687***/+++	0.648***/+++	0.627***/+++	0.648***/+++	0.687***/+++	0.696***/+++
	10%	1.000	1.236***/+++	0.977***/+++	0.644***/+++	0.639***/+++	0.657***/+++	0.648***/+++	0.634***/+++	0.630***/+++	0.639***/+++	0.653***/+++

TABLE 4.1: Comparison of the forecasting methods. Table reports relative RMSE. The smaller the entry ( $< 1$ ) the better the forecast. \*, \*\*, and \*\*\* denote results from Diebold and Mariano (1995) test with the Harvey, Leybourne, and Newbold (1997) adjustment for predictive accuracy, indicating rejection of the null hypothesis that the models have the same predictive accuracy at the 10%, 5%, and 1% levels of significance, respectively. +, ++, and +++ denote results from the Giacomini and White (2006) test, indicating rejection of the null hypothesis of equal forecasting accuracy of the models at the 10%, 5%, and 1% levels of significance, respectively.

$\eta = 0.005$	Block	$\tau = 1\%$			$\tau = 5\%$			$\tau = 10\%$		
		wins	losses	inconclusive	wins	losses	inconclusive	wins	losses	inconclusive
	linear	171	17	165	177	11	167	165	23	149
	non-parametric	341	35	333	333	43	319	320	56	305
	non-linear	<b>723</b>	<b>76</b>	<b>715</b>	<b>711</b>	<b>88</b>	<b>702</b>	<b>660</b>	<b>139</b>	<b>643</b>
	MIDAS	168	20	164	168	20	166	167	21	165
	non-linear MIDAS	685	67	680	660	92	653	613	139	606

TABLE 4.2: Entries of the table present the number of times a block encompasses (wins), does not encompass (losses) and is inconclusive, according to the CQFE test for different quantiles  $\tau$ . Results are reported for  $\eta = 0.005$ .

FIGURE 4.5: Partial Derivative, SHAP and  $\hat{\beta}(\tau)$  for GARCH and RM models.

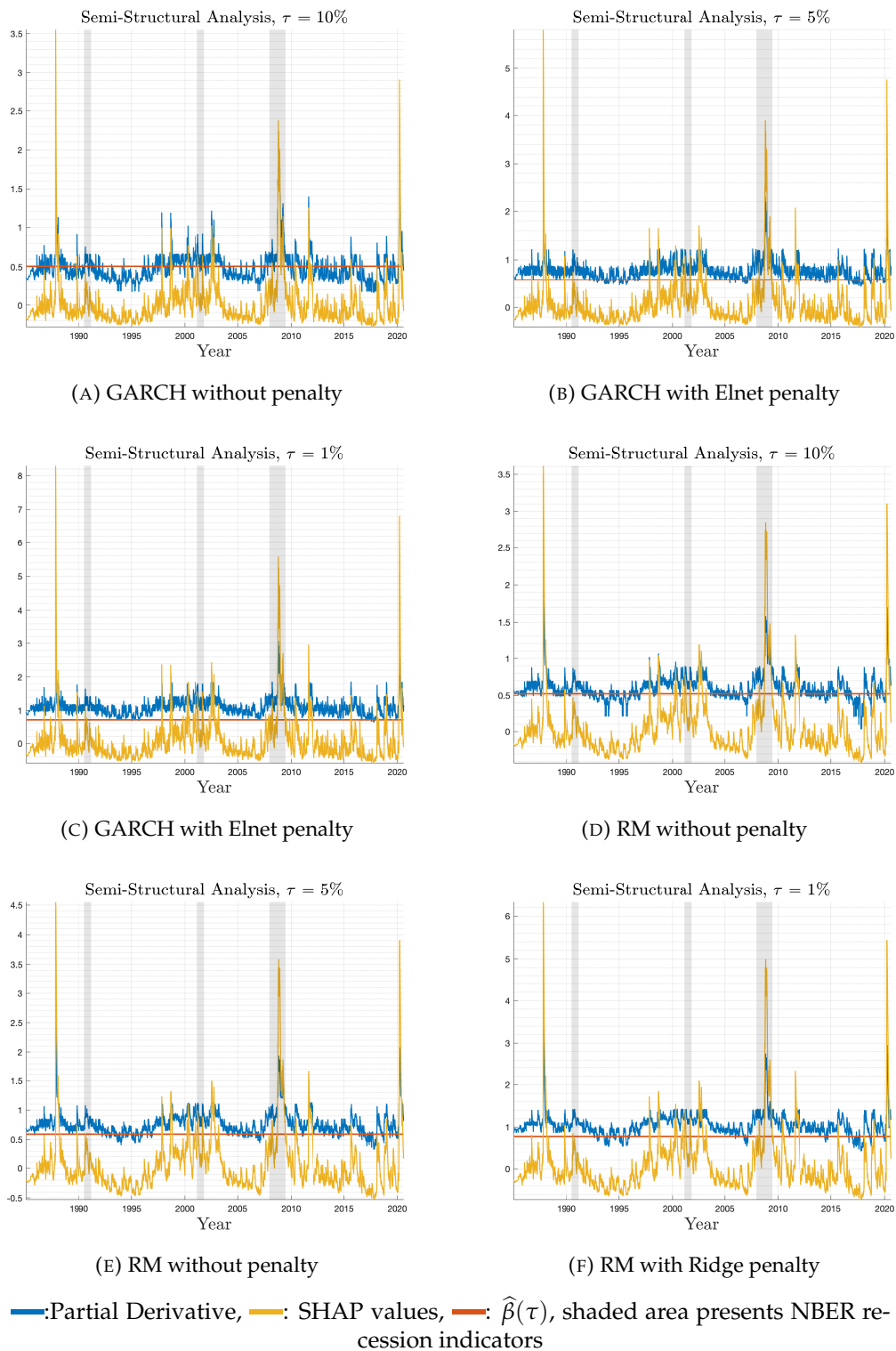
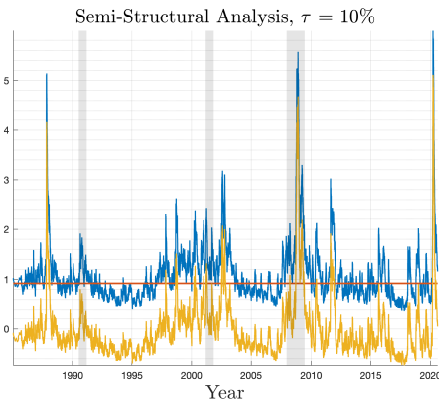
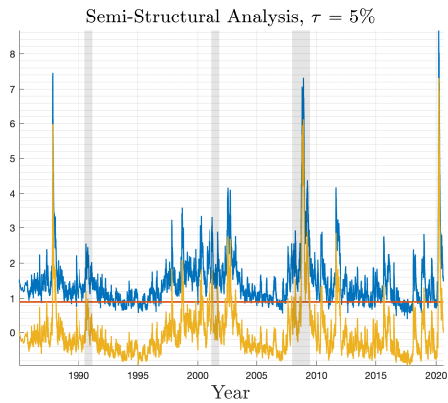


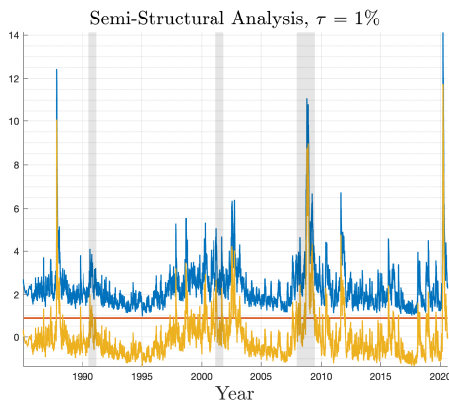
FIGURE 4.6: Partial Derivative, SHAP and  $\hat{\beta}(\tau)$  for SV model.



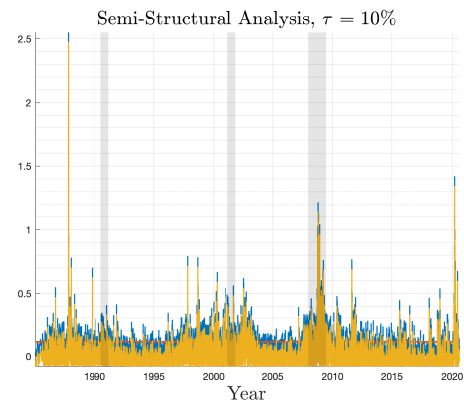
(A) SV without penalty



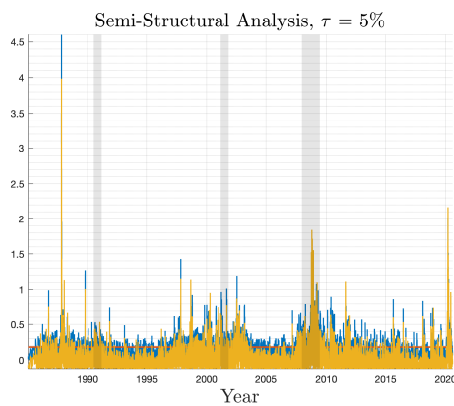
(B) SV without penalty



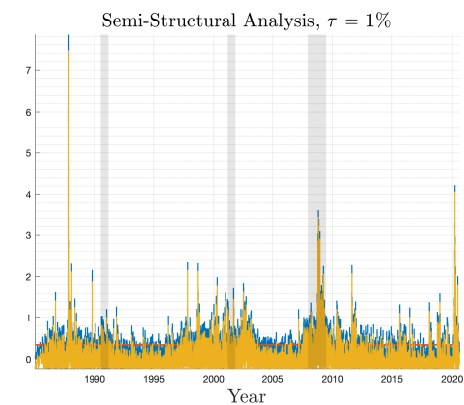
(C) SV with LASSO penalty



(D) VaR lagged values without penalty



(E) VaR lagged values without penalty



(F) VaR lagged values with LASSO penalty

—: Partial Derivative, —: SHAP values, —:  $\hat{\beta}(\tau)$ , shaded area presents NBER recession indicators



FIGURE 4.7: Partial Derivative, SHAP and  $\hat{\beta}(\tau)$  for ASV model.

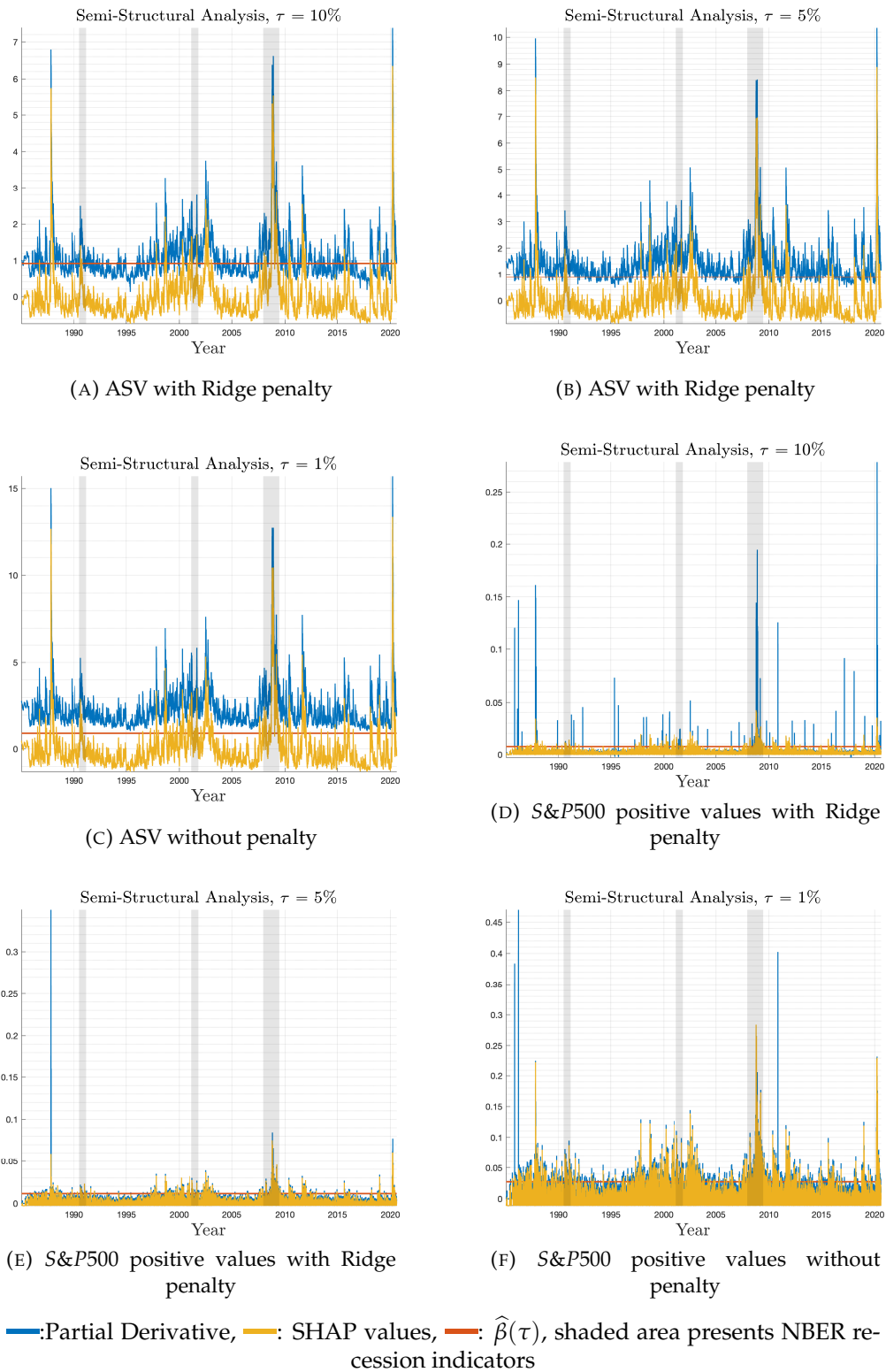
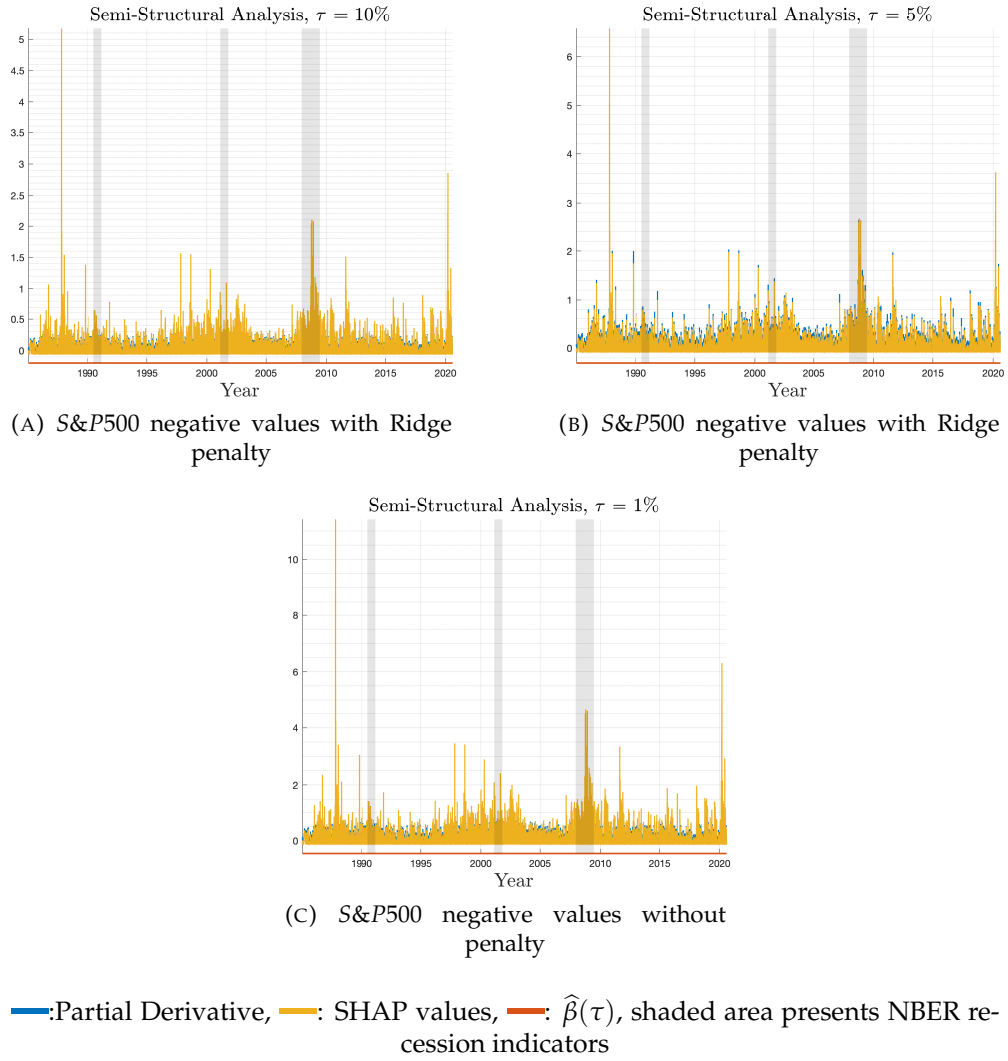


FIGURE 4.8: Partial Derivative, SHAP and  $\hat{\beta}(\tau)$  for ASV model.



## Chapter 5

# Conclusion

This thesis contributes to both the macroeconomic and financial forecasting literatures using Big Data econometric methods in the following ways: first, we discuss possible prior setups using a Mixed frequency Vector Autoregression (MF-VAR) written at the common lower frequency. Second, we use ideas from recent work in estimating high frequency MF -VAR models to discuss how monthly indicators offer more information on producing one-step ahead U.S. GDP by state forecasts. Finally, we drop the linearity assumption and using machine learning, we propose a deep quantile estimator that is both non-linear and non-parametric estimator and forecast Value-at-Risk.

Chapter 2 contributes to the econometric literature and develops models for regional nowcasting. We use Bayesian mixed frequency methods estimated at the common lower frequency. Moreover, we propose a procedure which allows model estimation with stochastic volatility and large datasets. We produce high frequency state-level GDP nowcasts that will assist policymakers in understanding the impact of greater regionalisation on economic growth in the U.S. and evaluate its impact on present and future economic conditions in a more timely fashion. We evaluate the accuracy of point and density forecasts, by making comparisons across models with constant and stochastic volatility. We provide results on the accuracy of nowcasts of realtime economic growth in the U.S. from 2006 to 2018. Empirical results suggest that models with stochastic volatility outperform models with constant volatility at nowcasting.

The second essay (Chapter 3) develops a Mixed Frequency Vector Autoregressive model (MF-VAR) for producing timely monthly nowcasts and historical estimates of

GDP growth at the state level for the U.S. economy. The variables in the MF-VAR include GDP growth at the state and country level, as well as additional monthly variables at the state and country level. The variables are observed at different frequencies, leading to a complicated high-dimensional MF-VAR. A computationally-fast approximate Bayesian Markov Chain Monte Carlo (MCMC) algorithm is proposed for estimating the MF-VAR coefficients and nowcasting. Empirical results explore the nature and magnitude of spillover effects among the U.S. states. Further, the proposed model produces historical estimates at monthly frequency for both the U.S. economy and U.S. states.

The third essay (Chapter 4) proposes a deep quantile estimator, using neural networks and their universal approximation property to examine a non-linear association between the conditional quantiles of a dependent variable and predictors. The proposed methodology is versatile and allows both the use of different penalty functions, as well as high dimensional covariates. We present a Monte Carlo exercise where we examine the finite sample properties of the proposed estimator and show that our approach delivers good finite sample performance. We use the deep quantile estimator to forecast Value-at-Risk and find significant gains over linear quantile regression alternatives, supported by various testing schemes. The Chapter also contributes to the interpretability of neural networks output by making comparisons between the commonly used SHAP values and an alternative method based on partial derivatives.

# Bibliography

- ADAMS, P. A., T. ADRIAN, N. BOYARCHENKO, AND D. GIANNONE (2021): “Forecasting macroeconomic risks,” *International Journal of Forecasting*.
- ALTAVILLA, C., R. GIACOMINI, AND G. RAGUSA (2017): “Anchoring the yield curve using survey expectations,” *Journal of Applied Econometrics*, 32(6), 1055–1068.
- ANKARGREN, S., AND P. JONÉUS (2019): “Simulation smoothing for large mixed-frequency VARs,” *Preprint available on arXiv*.
- ARIAS, M. A., C. S. GASCON, AND D. E. RAPACH (2016): “Metro business cycles,” *Journal of Urban Economics*, 94, 90–108.
- ATHEY, S., AND G. W. IMBENS (2017): “The state of applied econometrics: Causality and policy evaluation,” *Journal of Economic Perspectives*, 31(2), 3–32.
- BABII, A., X. CHEN, E. GHYSELS, AND R. KUMAR (2020): “Binary choice with asymmetric loss in a data-rich environment: Theory and an application to racial justice,” *arXiv preprint arXiv:2010.08463*.
- BAÑBURA, M., D. GIANNONE, AND L. REICHLIN (2010): “Large Bayesian vector autoregressions,” *Journal of Applied Econometrics*, 25(1), 71–92.
- BANBURA, M., D. GIANNONE, AND L. REICHLIN (2010): “Large Bayesian vector autoregressions,” *Journal of Applied Econometrics*, 25, 71–92.
- BAÑBURA, M., AND M. MODUGNO (2014): “Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data,” *Journal of Applied Econometrics*, 29(1), 133–160.
- BATES, J. M., AND C. W. GRANGER (1969): “The combination of forecasts,” *Journal of the Operational Research Society*, 20(4), 451–468.

- BAUMEISTER, C., D. LEIVA-LEÓN, AND E. SIMS (2021): "Tracking weekly state-level economic conditions," *The Review of Economics and Statistics*, pp. 1–45.
- BAUR, D., AND N. SCHULZE (2005): "Coexceedances in financial markets—a quantile regression analysis of contagion," *Emerging Markets Review*, 6(1), 21–43.
- BELLONI, A., V. CHERNOZHUKOV, D. CHETVERIKOV, AND I. FERNÁNDEZ-VAL (2019): "Conditional quantile processes based on series or many regressors," *Journal of Econometrics*, 213(1), 4–29.
- BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2014): "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 81(2), 608–650.
- BHATTACHARYA, A., A. CHAKRABORTY, AND B. K. MALLICK (2016): "Fast sampling with Gaussian scale mixture priors in high-dimensional regression," *Biometrika*, p. asw042.
- BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, 31(3), 307–327.
- BRAVE, S., R. BUTTERS, AND A. JUSTINIANO (2019): "Forecasting economic activity with mixed frequency Bayesian VARs," *Journal of Applied Econometrics*, 35, 1692–1707.
- BUCCI, A. (2020): "Realized Volatility Forecasting with Neural Networks," *Journal of Financial Econometrics*, 18(3), 502–531.
- CARLINO, G. A., R. H. DEFINA, AND K. SILL (2001): "Sectoral shocks and metropolitan employment growth," *Journal of Urban Economics*, 50(3), 396–417.
- CARRIERO, A., T. E. CLARK, AND M. MARCELLINO (2019): "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*.
- CARRIERO, A., G. KAPETANIOS, AND M. MARCELLINO (2009): "Forecasting exchange rates with a large Bayesian VAR," *International Journal of Forecasting*, 25(2), 400–417.
- CARVALHO, C. M., N. G. POLSON, AND J. G. SCOTT (2009): "Handling sparsity via the horseshoe," in *Artificial Intelligence and Statistics*, pp. 73–80. PMLR.

- (2010): “The horseshoe estimator for sparse signals,” *Biometrika*, 97(2), 465–480.
- CHAN, J. C. (2019): “Asymmetric conjugate priors for large Bayesian VARs,” .
- CHAN, J. C., C. Y. HSIAO, ET AL. (2014): *Estimation of stochastic volatility models with heavy tails and serial dependence*. Wiley Online Library.
- CHAN, J. C., AND I. JELIAZKOV (2009): “Efficient simulation and integrated likelihood estimation in state space models,” *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2), 101–120.
- CHEN, X., Y. LIU, S. MA, AND Z. ZHANG (2020): “Efficient estimation of general treatment effects using neural networks with a diverging number of confounders,” *arXiv preprint arXiv:2009.07055*.
- CHERNOZHUKOV, V., AND L. UMANTSEV (2001): “Conditional value-at-risk: Aspects of modeling and estimation,” *Empirical Economics*, 26(1), 271–292.
- CHIB, S. (1995): “Marginal Likelihood from the Gibbs Output,” *Journal of the American Statistical Association*, forthcoming. Also Washington University Olin School of Business working paper, .
- CLARK, T. E. (2011): “Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility,” *Journal of Business & Economic Statistics*, 29(3), 327–341.
- CLARK, T. E., AND F. RAVAZZOLO (2012): “The macroeconomic forecasting performance of autoregressive models with alternative specifications of time-varying volatility,” .
- COGLEY, T., AND T. J. SARGENT (2005): “Drifts and volatilities: monetary policies and outcomes in the post WWII US,” *Review of Economic Dynamics*, 8(2), 262–302.
- CRONE, T. M., AND A. CLAYTON-MATTHEWS (2005): “Consistent economic indexes for the 50 states,” *Review of Economics and Statistics*, 87(4), 593–603.
- CROSS, J. L., C. HOU, AND A. POON (2020): “Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity,” *International Journal of Forecasting*.

- DEL NEGRO, M. (2002): "Asymmetric shocks among US states," *Journal of International Economics*, 56(2), 273–297.
- DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–263, Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar.
- DIEBOLD, F. X., AND K. YILMAZ (2014): "On the network topology of variance decompositions: Measuring the connectedness of financial firms," *Journal of econometrics*, 182(1), 119–134.
- DIEPPE, A., R. LEGRAND, AND B. VAN ROYE (2016): "The BEAR toolbox," .
- DORAN, H. E. (1992): "Constraining Kalman filter and smoothing estimates to satisfy time-varying restrictions," *The Review of Economics and Statistics*, pp. 568–572.
- DU, Z., M. WANG, AND Z. XU (2019): "On Estimation of Value-at-Risk with Recurrent Neural Network," in *2019 Second International Conference on Artificial Intelligence for Industries (AII)*, pp. 103–106. IEEE.
- DURBIN, J., AND S. J. KOOPMAN (2002): "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 89(3), 603–616.
- ENGLE, R. F., AND S. MANGANELLI (2004): "CAViaR: Conditional autoregressive value at risk by regression quantiles," *Journal of Business & Economic Statistics*, 22(4), 367–381.
- FARRELL, M. H., T. LIANG, AND S. MISRA (2021): "Deep neural networks for estimation and inference," *Econometrica*, 89(1), 181–213.
- GALANT, A., AND H. WHITE (1992): "On learning the derivatives of an unknown mapping with multilayer feed forward neural network," *Neural networks*, 5, 129–138.
- GEFANG, D., G. KOOP, AND A. POON (2020): "Computationally efficient inference in large Bayesian mixed frequency VARs," *Economics Letters*, 191, 109120.
- GEORGE, E., D. SUN, AND S. NI (2008a): "Bayesian stochastic search for VAR model restrictions," *Journal of Econometrics*, 142, 553–580.



- GEORGE, E. I., D. SUN, AND S. NI (2008b): "Bayesian stochastic search for VAR model restrictions," *Journal of Econometrics*, 142(1), 553–580.
- GHYSELS, E. (2016): "Macroeconomics and the reality of mixed frequency data," *Journal of Econometrics*, 193(2), 294–314.
- GHYSELS, E., AND M. MARCELLINO (2018): *Applied Economic Forecasting Using Time Series Methods*. Oxford University Press.
- GHYSELS, E., A. PLAZZI, AND R. VALKANOV (2016): "Why invest in emerging markets? The role of conditional return asymmetry," *The Journal of Finance*, 71(5), 2145–2192.
- GHYSELS, E., A. PLAZZI, R. VALKANOV, A. RUBIA, AND A. DOSSANI (2019): "Direct versus iterated multiperiod volatility forecasts," *Annual Review of Financial Economics*, 11, 173–195.
- GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2004): "The MIDAS touch: Mixed data sampling regression models," .
- GIACOMINI, R., AND I. KOMUNJER (2005): "Evaluation and combination of conditional quantile forecasts," *Journal of Business & Economic Statistics*, 23(4), 416–431.
- GIACOMINI, R., AND G. RAGUSA (2014): "Theory-coherent forecasting," *Journal of Econometrics*, 182(1), 145–155.
- GIACOMINI, R., AND H. WHITE (2006): "Tests of conditional predictive ability," *Econometrica*, 74(6), 1545–1578.
- GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2015): "Prior selection for vector autoregressions," *Review of Economics and Statistics*, 97(2), 436–451.
- GNEITING, T., AND A. E. RAFTERY (2007): "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, 102(477), 359–378.
- GOODFELLOW, I., Y. BENGIO, AND A. COURVILLE (2016): *Deep learning*. MIT press.
- GU, S., B. KELLY, AND D. XIU (2020a): "Autoencoder asset pricing models," *Journal of Econometrics*.
- (2020b): "Empirical asset pricing via machine learning," *The Review of Financial Studies*, 33(5), 2223–2273.

- HANSEN, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica: Journal of the Econometric Society*, pp. 1029–1054.
- HARTMANN, S., AND J. SPRENGER (2010): "Bayesian epistemology," .
- HARVEY, D., S. LEYBOURNE, AND P. NEWBOLD (1997): "Testing the equality of prediction mean squared errors," *International Journal of forecasting*, 13(2), 281–291.
- HE, Z., AND A. KRISHNAMURTHY (2013): "Intermediary asset pricing," *American Economic Review*, 103(2), 732–70.
- HORNIK, K. (1991): "Approximation capabilities of multilayer feedforward networks," *Neural networks*, 4(2), 251–257.
- HORNIK, K., M. STINCHCOMBE, AND H. WHITE (1989): "Multi-Layer Feedforward Networks and Universal Approximators," *Neural Network*, 2, 359–366.
- JAROCIŃSKI, M. (2015): "A note on implementing the Durbin and Koopman simulation smoother," *Computational Statistics & Data Analysis*, 91, 1–3.
- JOSEPH, A. (2019): "Shapley regressions: a framework for statistical inference on machine learning models," Discussion paper, Bank of England.
- J.P. MORGAN, M. (1996): "Reuters (1996) RiskMetrics-Technical Document," *JP Morgan*.
- KAPETANIOS, G. (2007): "Measuring conditional persistence in nonlinear time series," *Oxford Bulletin of Economics and Statistics*, 69(3), 363–386.
- KAPETANIOS, G., AND A. P. BLAKE (2010): "Tests of the martingale difference hypothesis using boosting and RBF neural network approximations," *Econometric Theory*, 26(5), 1363–1397.
- KASTNER, G., AND F. HUBER (2020): "Sparse Bayesian vector autoregressions in huge dimensions," *Journal of Forecasting*, 39, 1142–1165.
- KEILBAR, G., AND W. WANG (2021): "Modelling systemic risk using neural network quantile regression," *Empirical Economics*, pp. 1–26.
- KIM, S., N. SHEPHARD, AND S. CHIB (1998): "Stochastic volatility: likelihood inference and comparison with ARCH models," *The review of economic studies*, 65(3), 361–393.

- KINGMA, D. P., AND J. BA (2014): "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*.
- KOENKER, R. (2005): *Quantile Regression*, Econometric Society Monographs. Cambridge University Press.
- KOENKER, R., AND G. BASSETT JR (1978): "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50.
- KOENKER, R., V. CHERNOZHUKOV, X. HE, AND L. PENG (2017): *Handbook of quantile regression*. CRC press.
- KOENKER, R., AND K. F. HALLOCK (2001): "Quantile regression," *Journal of economic perspectives*, 15(4), 143–156.
- KOOP, G., AND D. KOROBILIS (2010): *Bayesian multivariate time series methods for empirical macroeconomics*. Now Publishers Inc.
- KOOP, G., S. MCINTYRE, AND J. MITCHELL (2019): "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)*.
- KOOP, G., S. MCINTYRE, J. MITCHELL, AND A. POON (2020a): "Reconciled Estimates and Nowcasts of Regional Output in the UK," *National Institute Economic Review*, 253, R44–R59.
- (2020b): "Regional Output Growth in the United Kingdom: More Timely and Higher Frequency Estimates From 1970," *Journal of Applied Econometrics*, 35, 176–197.
- (2022): "Using stochastic hierarchical aggregation constraints to nowcast regional economic aggregates," *International Journal of Forecasting*.
- KOOP, G., M. H. PESARAN, AND S. M. POTTER (1996): "Impulse response analysis in nonlinear multivariate models," *Journal of econometrics*, 74(1), 119–147.
- KOOP, G. M. (2013): "Forecasting with medium and large Bayesian VARs," *Journal of Applied Econometrics*, 28(2), 177–203.
- KOROBILIS, D. (2013): "VAR forecasting using Bayesian variable selection," *Journal of Applied Econometrics*, 28, 204–230.

- KRÜGER, F., T. E. CLARK, AND F. RAVAZZOLO (2017): "Using entropic tilting to combine BVAR forecasts with external nowcasts," *Journal of Business & Economic Statistics*, 35(3), 470–485.
- LIANG, S., AND R. SRIKANT (2016): "Why deep neural networks for function approximation?," *arXiv preprint arXiv:1610.04161*.
- LUNDBERG, S. M., G. G. ERION, AND S.-I. LEE (2018): "Consistent individualized feature attribution for tree ensembles," *arXiv preprint arXiv:1802.03888*.
- LUNDBERG, S. M., AND S.-I. LEE (2017): "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*, pp. 4765–4774.
- MCCRACKEN, M. W., AND S. NG (2016): "FRED-MD: A monthly database for macroeconomic research," *Journal of Business & Economic Statistics*, 34(4), 574–589.
- MCCRACKEN, M. W., M. OWYANG, AND T. SEKHPOSYAN (2019): "Real-Time Forecasting and Scenario Analysis using a Large Mixed-Frequency Bayesian VAR," *Working Paper*.
- MEINSHAUSEN, N. (2006): "Quantile regression forests," *Journal of Machine Learning Research*, 7(Jun), 983–999.
- NEAL, R. M. (2003): "Slice sampling," *The annals of statistics*, 31(3), 705–767.
- NEWBY, W. K., AND K. D. WEST (1987): "Hypothesis testing with efficient method of moments estimation," *International Economic Review*, pp. 777–787.
- PADILLA, O. H. M., W. TANSEY, AND Y. CHEN (2020): "Quantile regression with deep ReLU Networks: Estimators and minimax rates," *arXiv preprint arXiv:2010.08236*.
- PARK, J., AND I. W. SANDBERG (1991): "Universal Approximation using Radial-Basis-Function Networks," *Neural Computation*, 3(4), 246–257.
- PESARAN, H. H., AND Y. SHIN (1998): "Generalized impulse response analysis in linear multivariate models," *Economics letters*, 58(1), 17–29.
- POHL, W., K. SCHMEDDERS, AND O. WILMS (2018): "Higher order effects in asset pricing models with long-run risks," *The Journal of Finance*, 73(3), 1061–1111.
- POLSON, N. G., J. G. SCOTT, AND J. WINDLE (2014): "The Bayesian bridge," *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pp. 713–733.

- PRIMICERI, G. E. (2005): "Time varying structural vector autoregressions and monetary policy," *The Review of Economic Studies*, 72(3), 821–852.
- ROBERTSON, J. C., E. W. TALLMAN, AND C. H. WHITEMAN (2005): "Forecasting using relative entropy," *Journal of Money, Credit, and Banking*, 37(3), 383–401.
- SCHMIDT-HIEBER, J. (2020): "Nonparametric regression using deep neural networks with ReLU activation function," *The Annals of Statistics*, 48(4), 1875–1897.
- SCHORFHEIDE, F., AND D. SONG (2015): "Real-time forecasting with a mixed-frequency VAR," *Journal of Business & Economic Statistics*, 33(3), 366–380.
- SHAPLEY, L. S. (1953): "A value for n-person games," *Contributions to the Theory of Games*, 2(28), 307–317.
- SHEN, G., Y. JIAO, Y. LIN, J. L. HOROWITZ, AND J. HUANG (2021): "Deep quantile regression: Mitigating the curse of dimensionality through composition," *arXiv preprint arXiv:2107.04907*.
- SHRIKUMAR, A., P. GREENSIDE, AND A. KUNDAJE (2017): "Learning important features through propagating activation differences," *arXiv preprint arXiv:1704.02685*.
- SIMS, C. A. (1980): "Macroeconomics and reality," *Econometrica: journal of the Econometric Society*, pp. 1–48.
- SMALTER HALL, A., AND T. R. COOK (2017): "Macroeconomic indicator forecasting with deep neural networks," *Federal Reserve Bank of Kansas City Working Paper*, (17-11).
- STOCK, J. H., AND M. W. WATSON (1989): "New indexes of coincident and leading economic indicators," *NBER macroeconomics annual*, 4, 351–394.
- TAMBWEKAR, A., A. MAIYA, S. S. DHAVALA, AND S. SAHA (2021): "Estimation and Applications of Quantiles in Deep Binary Classification," *IEEE Transactions on Artificial Intelligence*.
- TOBIAS, A., AND M. K. BRUNNERMEIER (2016): "CoVaR," *The American Economic Review*, 106(7), 1705.
- WAGER, S., AND S. ATHEY (2018): "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 113(523), 1228–1242.

- WAGGONER, D. F., AND T. ZHA (1999): "Conditional forecasts in dynamic multivariate models," *Review of Economics and Statistics*, 81(4), 639–651.
- YAROTSKY, D. (2017): "Error bounds for approximations with deep ReLU networks," *Neural Networks*, 94, 103–114.
- ZHANG, W., H. QUAN, AND D. SRINIVASAN (2018): "An improved quantile regression neural network for probabilistic load forecasting," *IEEE Transactions on Smart Grid*, 10(4), 4425–4434.
- ZOU, H., AND T. HASTIE (2005): "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, 67(2), 301–320.