Essays in macroeconomic forecasting with linear and non-linear time series models using Bayesian techniques

by

Georgios Kontogeorgos

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

University of Strathclyde

Department of Economics

January 2023

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:
Date:

Acknowledgements

First and foremost, I feel deeply indebted to my primary supervisor Gary Koop for his invaluable guidance and constructive discussions throughout by PhD years. His immense knowledge and plentiful experience have been of extraordinary importance.

I would also like to thank my secondary supervisor and postgraduate research director, Julia Darby, for her support and excellent comments on my thesis. I am also grateful to the University of Strathclyde, Economics department academic staff, fellow PhD students, and administrators for fostering a community which cultivates and promotes an excellent research culture. Especially during the difficult times of Covid-19, this has been a crucial aspect.

This research was supported financially by the Scottish Graduate School of Social Science, which I gratefully acknowledge. I would also like to thank them for the research training and support fund, which allowed me to attend various conferences, seminars and summer schools.

I should also thank professors Mattias Villani, Nikolas Kantas, Simo Särkkä, Arnaud Doucet and Nicolas Chopin for their advice on crucial computational aspects in the first and second essays.

Furthermore, I would like to thank my current employer Allstate Corporation for providing the relevant work schedule flexibility to accommodate my needs with respect to fulfilling PhD studies requirements successfully.

Finally, yet not less importantly, completing this thesis would not have been feasible without the love and support of my family and friends, whose patience, encouragement and endless understanding have kept me moving forward and motivated.

Synopsis

This PhD thesis comprises three essays which explore novel approaches to modelling and forecasting macroeconomic time series using Bayesian techniques. In short, the first essay examines the impact of long-run expectations in a conditional forecasting setting. The second essay proposes the extension of the non-linear class of threshold autoregressive models with a parsimonious set of time-varying parameters. Finally, the third essay proposes multivariate time series forecasting methods which incorporate time-frequency analysis.

In the first essay, I investigate whether incorporating survey-based long-run expectations via the steady-state prior in Bayesian vector autoregressions (BVARs) can improve conditional forecasts. The long-run behaviour of the conditional forecasts is disciplined by employing the steady-state prior, whose mean is set equal to the Survey of Professional Forecasters' long-run expectations. Using US real-time data since 1980 in an out-of-sample forecast evaluation (1997-2013) and assuming the realization of short-term interest rates is known ex-ante, I find that the steady-state prior improves accuracy for long-term yields consistently and further improves GDP and unemployment forecasts with the inclusion of stochastic volatility relative to a benchmark-Minnesota prior BVAR. However, under a model's reality check assuming that the future paths of the variables in the VAR, except for one, are known ex-ante, survey expectations of unemployment improve conditional forecasts accuracy in the period following the financial crisis, whereas, for the CPI and the FFR, they do not. Finally, for homoscedastic steady-state prior BVARs, the hyperparameters are estimated for each forecasting exercise separately by marginal likelihood maximization employing particle swarm optimisation, and it is found that the financial crisis has a minor impact on their optimal values.

In the second essay, I propose the extension of the logistic smooth transition autoregressive model in the univariate setting by allowing the threshold and the constant to be time-varying. Using Monte Carlo simulation, I show that the model's parameters can be estimated successfully via a combination of particle filtering and Markov chain Monte Carlo algorithms. In the empirical application part, using US data for GDP, CPI, FFR, and 10-year bond yield, the proposed model outperforms the linear benchmark model in an out-of-sample forecasting exercise over the period 1985-2018 on many occasions. Following the Great Recession, the time variation of the results shows stronger evidence in favour of the proposed model, and in many instances, the difference versus the linear benchmark model is found to be significant. The inclusion of stochastic volatility does not improve the forecasting performance except for the FFR at the 1-quarter ahead forecast horizon.

Finally, in the third essay, by employing wavelet analysis which allows the time localisation of time series frequency characteristics, I extend the existing literature on wavelet-based univariate time series modelling and forecasting into a multivariate setting and under a Bayesian framework both for single and mixed-frequency time series. Regarding the single-frequency time series, the first approach uses discrete wavelet transform (DWT) based denoising and subsequent conventional Bayesian vector autoregression (BVAR) forecasting. The second approach employs the Haar Maximal Overlap Discrete Wavelet Transform (MODWT), which is a time series multiscale additive decomposition describing fluctuations over different frequency bands. Separate scale BVARs are formed across each scale, and forecasts are estimated by aggregating the separate scale forecasts. The third approach extends the Multiscale Autoregressive model of Renaud et al. (2003) into the Multiscale BVAR employing a Minnesota-inspired prior, which allows a varying degree of shrinkage across different scales, as well as an SSVS prior. In an out-of-sample forecasting exercise using US

macroeconomic variables, the three forecasting approaches are found to outperform a conventional BVAR on many occasions. In particular, wavelet-based denoising forecasting presents merits for density forecasts, and the Multiscale BVAR outperforms the benchmark across all variables for medium to long-term forecasts. Regarding modelling mixed-frequency time series using wavelet analysis, the proposed wavelet-based MF-VAR model, which comprises separate scale MF-VARs in a single system, exhibits increased in-sample forecast accuracy for known monthly time series in a statistical sense compared to the standard MF-VAR; however, this behaviour reverses during recessionary periods like the latest COVID-19 recession.

Contents

List of tables

List of figures

‘The only function of economic forecasting is to make astrology look respectable.’

Credited to
Ezra Solomon (1985)

Essay 1

1. Bayesian vector autoregression conditional forecasting: Incorporating survey-based long-run expectations via the steady-state prior

1.1 Introduction

Conditional forecasting has a prominent role in central banks and fiscal authorities since it is common practice that the projections they publish are conditional on a set of assumptions. Central banks produce the macroeconomic variables’ projections conditional mainly on the main policy instrument, the short-term interest rates[1] under conventional times, as well as other assumptions over fiscal variables and the evolution of the external economic environment[2].

Conditional forecasts can be briefly characterised across three main dimensions. The first dimension evolves around the corresponding underlying models used to produce the conditional forecasts and refers to the two main classes of competing macroeconomic models: structural models such as the micro-founded dynamic stochastic equilibrium (DSGE) and time-series models involving structural VARs and factor models[3]. For the DSGE-relevant literature, see Adolfson, Laséen, Lindé, and Villani (2005), Maih (2010), Del Negro and Schorfheide (2012) and Wieland and Wolters (2013).

The second dimension categorises conditional forecasts according to the conditioning information. Following Antolin-Diaz, Petrella, and Rubio Ramírez (2021), conditional forecasts can be: *i*) conditional on the observables: either directly on the future path of some variable as introduced by Doan, Litterman, and Sims (1984) or by using all the identified structural shocks to generate the desired values on the particular variable(s) as in Waggoner and Zha (1999), *ii*) conditional on specific structural shocks, or *iii*) conditional on a

[1] Galí (2011) provides a critical review of the three main approaches used by central banks on short-term rate assumptions (constant path, market expectations, central banks’ own expectations) in relation to forward-looking New Keynesian models.
[2] For example, ECB (2016) provides the complete set of the assumptions used by the European Central Bank.
[3] Factor models have been found to be outperformed in terms of forecast accuracy by large BVARs (Bańbura, Giannone, & Reichlin, 2010; Koop, 2013).

combination of future paths of some variables and a subset of the structural shocks which allows performing structural scenario analysis. In the later case, conditional forecasting can be performed simultaneously with sign restrictions (Rubio-Ramírez, Waggoner, & Zha, 2010; Baumeister & Hamilton, 2015), narrative restrictions (Antolin-Diaz & Rubio-Ramírez, 2018), or a combination of zero and sign restrictions (Arias, Rubio-Ramírez, & Waggoner, 2018).

With regard to the third dimension, conditional forecasts can be characterised according to the uncertainty of the conditioning information. If there does not exist any uncertainty over the conditioning assumptions, following Waggoner and Zha (1999) terminology, they are 'hard conditioned', whereas when uncertainty is present, they are 'soft conditioned'. Waggoner and Zha (1999) algorithm for soft conditioning assumes that the uncertainty over the conditioning variables(s) falls within a certain range around a central tendency, but Andersson, Palmqvist, and Waggoner (2010) generalise the algorithm further, which allows to condition on the first and second moment of the restricted path.

Another alternative approach competitive to conditional forecasting pertains to the ex-post tilting of the multivariate unconditional predictive density produced by any model according to a set of restrictions that the tilted distribution should satisfy. This approach was first introduced by Robertson, Tallman, and Whiteman (2005) and has grown in popularity for combining unconditional forecasts with informational content from other sources. These sources can involve, amongst others, nowcasts/short-term forecasts produced by other models or survey-based expectations over specific horizons and/or the long run (Altavilla, Giacomini, & Ragusa, 2017; Krüger, Clark, & Ravazzolo, 2017; Knotek II & Zaman, 2019; Tallman & Zaman, 2019).

The second important element that this essay is based on is the use of the steady-state prior, as suggested by Villani (2009), in the context of conditional forecasting. Steady-state prior, which allows imposing beliefs on the VAR process unconditional mean, has been found to improve unconditional forecasting accuracy significantly, especially over long horizons (Beechey & Österholm, 2010; Clark, 2011; Wright, 2013). There are two main approaches when employing steady-state priors. The first involves the provision of information on the VAR's steady-state as estimated from economic theory[4], DSGE models or appropriate calibration (Villani, 2009; Jarociński & Smets, 2008; Beechey & Österholm, 2010; Clark, 2011; Louzis, 2019), while the second sets the steady-state's prior mean equal to survey-based long-run expectations[5] (Wright, 2013; Bańbura & van Vlodrop, 2018).

I treat the steady-state prior along the lines of the second approach, i.e. I use survey-based expectations over the long run as the steady-state prior mean for the BVAR and consider the extent to which the information provided over the steady-state of the VAR improves its performance, not only for unconditional but for conditional forecasts as well. While the applied forecasting literature for unconditional BVAR forecasts is voluminous and spans indicatively over different prior specifications (Carriero, Clark, & Marcellino, 2015a), the use of stochastic volatility (Chiu, Mumtaz, & Pinter, 2017; Carriero, Clark, & Marcellino, 2019) and time-varying parameters (D'Agostino, Gambetti, & Giannone, 2013), the case of conditional forecasting accuracy has not attracted this attention despite its importance for the policy maker. It has to be noted that Andersson, Palmqvist, and Waggoner (2010) also employ a steady-state

[4] Giannone, Lenza, and Primiceri (2019) proposed conjugate priors that discipline the long-run VAR forecasts and can be elicited with respect to economic theory - DSGE models.

[5] Bańbura and van Vlodrop (2018) and Louzis (2019) have shown that allowing for time variation in the steady-state can further improve forecasting accuracy.

BVAR alongside their proposed soft-conditioning estimation algorithm for conditional forecasts and, consequently, the research question I am trying to address comes naturally as to which prior provides superior conditional forecasts.

In more detail, for the purpose of this research, I estimate the conditional forecasts of a set of macroeconomic variables using BVARs where the conditioning information is the future realised path of the key policy interest rate, i.e. conditional forecasts 'hard conditioned' directly on the observable-values of the policy rate and not on the underlying structural shocks. The contribution of this essay lies in using BVARs with the steady-state prior of Villani (2009), whose mean is set equal to the Survey of Professional Forecasters' long-run expectations with the aim of integrating beliefs over the convergence of the time series' conditional forecasts simultaneously with the assumptions used as conditioning information. In that sense, when performing conditional forecasting, the policy-maker will take into account the future path of the conditioning variable jointly with the expectations over the long run.

Finally, another contribution of this essay lies in proposing a new approach for optimising prior hyperparameters for the homoscedastic steady-state prior BVAR. Very few instances in the literature of steady-state prior address this problem; for example, Ankargren, Unosson, and Yang (2018) use an adaptive grid search and Gustafsson, Villani, and Stockhammar (2020) propose Bayesian optimisation algorithms. I take a different approach and use the derivative-free stochastic optimisation algorithm of the Particle Swarm Optimisation (PSO) algorithm to maximise the steady-state BVAR marginal likelihood estimated with the Chib (1995) approximation method with respect to the prior hyperparameters.

The essay proceeds as follows. Section 1.2 describes the real-time data. Section 1.3 discusses the impact of the VAR steady-state for long-run forecasts, the steady-state prior for the homoscedastic BVAR and its marginal likelihood estimation, as well as the steady-state prior BVAR with stochastic volatility. Section 1.4 deals with the estimation of the conditional forecasts, and Section 1.5 discusses the priors' details and the benchmark models. Section 1.6 reports the results, and finally, Section 1.7 concludes.

1.2 Data description

All BVARs specifications are estimated with the following six variables at quarterly frequency[6]: real GDP growth, CPI inflation, Federal Funds Rates (FFR), 10-year bond yield (10-Y), unemployment rate and nonfarm payroll employment. Variables are transformed into annualised quarter-on-quarter growth rates except for the FFR, 10-year bond yield and unemployment rate, which remain in percentage points. See Table *1.1*. Working with a mixed-frequency BVAR with a steady state prior, as in Ankargren, Unosson, and Yang (2020), which takes into account the monthly frequency of some of the variables, would be possible. However, such an approach would be more beneficial for the 'nowscaster' interested in producing monthly updates and short-term forecasts. This issue does not apply in our case

[6] The variables whose original frequency is monthly are first transformed into quarterly frequency by calculating each quarter's respective average.

since we are interested in producing long terms forecasts and using long-term survey expectations.

For all variables, when producing the forecasts at quarter t, the last observed value is for quarter $t-1$. This holds during the last month of each quarter t: quarterly indicators of the previous quarter $t-1$ have been released, and the monthly indicators available for the two first months of the corresponding quarter are disregarded since the BVARs are specified in quarterly frequency only.

For the BVARs using the steady-state prior, the US Survey of Professional Forecasters (SPF) dataset is used. SPF is conducted at a quarterly frequency over several macroeconomic and financial variables at different forecast horizons and is published during the second month of each quarter. SPF amongst the surveyed information also asks participants' expectations for the long-run expectations (10-year annual average) of CPI inflation in every quarter, and real GDP growth, 3-month Treasury bill and 10-year bond yield during the first quarter of each calendar year. Moreover, for the survey carried out in the third quarter of each calendar year, SPF asks for respondents' estimates of the natural rate of unemployment. Following Tallman and Zaman (2019), the 3-month Treasury bill and the natural rate of unemployment long-run expectations are treated in the same way as the FFR and the unemployment rate respectively. See Table 1.2: SPF data

To respect the real-time information flow, since the BVARs are re-estimated in each quarter, the values used for the steady-state prior mean are those from the latest SPF survey conducted during the first or the third quarter with respect to the corresponding variables. Similar to Tallman and Zaman (2019), I use the median response from the SPF dataset, which is more robust to outliers compared to the mean response as the steady-state prior mean. The first quarter for which SPF long-run expectations exist for all of the five variables using an informative steady-state prior is 1997Q1.

The out-of-sample evaluation period is 1997Q1-2014Q1, i.e. 69 vintages in total. The expanding sample window with recursive re-estimation of the BVARs for the out-of-sample evaluation always starts in 1980Q1. For the purpose of forecast evaluation, the outcome used is the one that became available four quarters after the first release[7,8] as in Kontogeorgos and Lambrias (2022). In order to estimate the conditional forecasts, the actual path of the FFR over the evaluation period 1997Q1-2016Q4 is used (see also the introduction of Section 1.3).

For example, for the 1997Q1 forecasting exercise, all models' estimation sample is 1980Q1-1996Q4. The 1-quarter ahead forecast refers to 1997Q1, and the 12-quarters ahead recursive forecast refers to 1999Q4. The conditional forecasts use the actual path of the FFR over 1997Q1-1999Q4 accordingly. With regard to the outcome used for the estimation of the forecast error, for the 1-quarter ahead forecast of 1997Q1 and the 12-quarters ahead forecast of 1999Q4, their corresponding realizations are taken from the 1998Q2 and the 2001Q1 vintage of data respectively.

[7] For the case of FFR and 10-year bond yield this is inconsequential as they are not subject to revisions.
[8] Similar practises are followed by studies using real-time macroeconomic data. First release, releases two and three quarters ahead of the reference period, up to two years ahead are used in El-Shagi, Giesen, and Jung (2016), Champagne, Poulin-Bellisle, and Sekkel (2020), Tulip (2009) and Faust and Wright (2008), respectively.

Table 1.1: VAR variables

Variable	Transformation	Source
Real GDP	$100\ ((y_t / y_{t-1})^4 - 1)$	RTDSM
CPI	$100\ ((y_t / y_{t-1})^4 - 1)$	RTDSM
FFR	-	ALFRED
10-year bond yield	-	ALFRED
Unemployment rate	-	RTDSM
Nonfarm payroll employment	$100\ ((y_t / y_{t-1})^4 - 1)$	RTDSM

Note. 'RTDSM' is the Real Time Data Set for Macroeconomists database of the Federal Reserve Bank of Philadelphia. 'ALFRED' is the Archival Federal Reserve Economic Data of the Federal Reserve Bank of St. Louis.

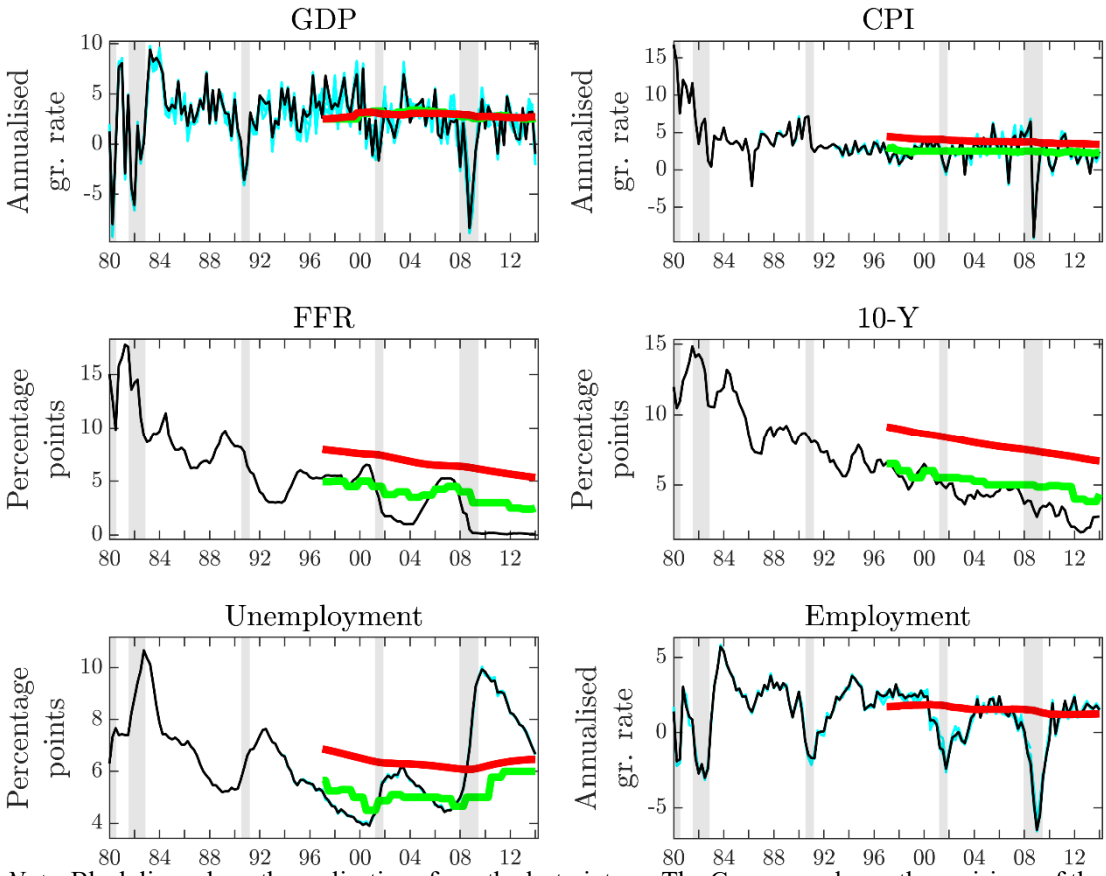
Table 1.2: SPF data

Variable	Available since	Survey freq.
Real GDP	1992Q1	Annually, Q1
CPI	1991Q4	Quarterly
3-M Treasury bill rate	1992Q1	Annually, Q1
10-year bond yield	1992Q1	Annually, Q1
Natural rate of unemployment	1996Q3	Annually, Q3

Note. Survey of Professional Forecasters available from the Federal Reserve Bank of Philadelphia. 10-year average forecasts.

The following two figures show the data. In more detail, Figure 1.1 plots the real-time data since the beginning of the estimation sample in 1980Q1 in black, and the SPF long-run expectations since 1997Q1 in green, while the red lines show the recursively estimated mean since 1980Q1. With the exception of GDP, for the remaining variables overall, it can be observed that the SPF expectations lie below the recursive mean. This is a sensible finding given that as time progresses in the post-mid-1980 great moderation period, the expectations are anchored to lower values compared to the early 1980s. This is particularly evident in the downward trend of the interest rates, whilst lesser for CPI. The effect of the 2008 financial crisis is profound and timely for FFR expectations; however, for the 10-year bond yield and the natural rate of unemployment, the decrease and increase in expectations, respectively, occur somewhat delayed after 2010-11 whilst the economy entered the recovery phase.

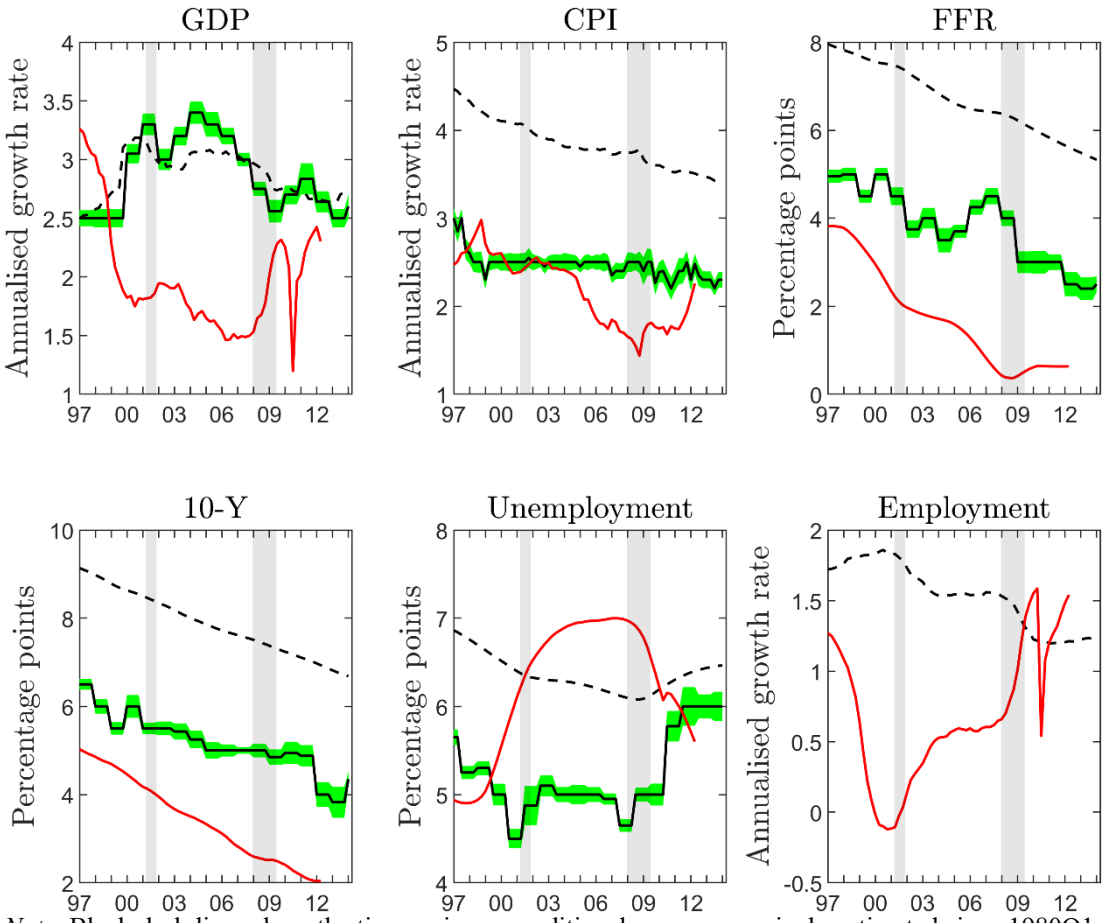
Figure 1.1: Real-time data and SPF long-run expectations



Note. Black lines show the realisations from the last vintage. The Cyan area shows the revisions of the real-time data over time. Green lines show the median SPF long-run expectations (10-year average), and red lines show the recursively estimated mean using the real-time data since 1980Q1.

Figure 1.2 shows over the evaluation period (1997Q1-2014Q1) the SPF long-run expectations accompanied by their confidence intervals constructed from the surveys' individual responses in green, as well as the ex-post realised 10-year average of the time series in red. With the exception of unemployment, for the remaining variables for the post-2000 period, it is clear that the realised 10-year average values stood much below expectations. The impact of the COVID-19 recession data is particularly evident in the realised 10-year average for GDP and employment. Regarding CPI and unemployment, in particular, expectations seem to follow an inverse relationship, i.e. CPI trending downwards whilst unemployment upwards, resembling some sort of an underlying Philips curve behaviour; however, this relationship appears to be much more evident when observing the ex-post realised values in red.

Figure 1.2: SPF long-run expectations and their dispersion



Note. Black dash lines show the time series unconditional means recursively estimated since 1980Q1. Black solid lines show the median SPF long-run expectations (10-year average), and the green area covers the 95% confidence interval after excluding the lowest and highest 5 percentiles from the raw survey data. Red lines show the 10-year ahead average of the realisation, and they stop in 2012 due to data availability.

1.3 BVAR models

This section presents in detail the homoscedastic and heteroscedastic steady-state BVARs models, which are used to assess the impact of incorporating long-term SPF expectations for the purpose of conditional and unconditional forecasting.

1.3.1 VAR steady-state

This section illustrates the importance of the VAR's steady-state (unconditional mean) and how it influences the forecasts directly. A VAR(P) process is defined as:

$$Y_t = B_c + \sum_{p=1}^{P} B_p Y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,\Sigma) \tag{1.1}$$

Where $Y_t = \left[Y_{1,t}, \cdots, Y_{N,t} \right]'$ is $N \times 1$, B_c is $N \times 1$, and B_p is $N \times N$. Σ is a full matrix.

Stacking all the VAR coefficients in $B = \left[B_c, B_1, \cdots, B_P \right]'$ and defining $X_t = \left[1, Y'_{t-1}, \cdots, Y'_{t-P} \right]$ the VAR in compact notation is given by: $Y'_t = X'_t B + \varepsilon_t$ or in matrix notation:

$$Y = XB + E \tag{1.2}$$

Where Y is $T \times N$, X is $T \times (1 + NP)$, B is $(1 + NP) \times N$ and E is $T \times N$.

Taking expectations, equation (1.1) yields:

$$E\left[Y_t \right] = B_c + \sum_{p=1}^{P} B_p E\left[Y_{t-p} \right] + E\left[\varepsilon_t \right]$$

Under the assumption of a covariance-stationary VAR process, where $E\left[Y_t \right] = E\left[Y_{t-p} \right] = \mu$ denotes the VAR process steady-state of dimensions $N \times 1$, the previous equations yields: $\mu = B_c + \sum_{p=1}^{P} B_p \mu$. Or equivalently:

$$B_c = \left(I_N - B_1 - \ldots - B_P \right) \mu \tag{1.3}$$

By simple reordering of equation (1.3), the steady-state of the VAR (Hamilton (1994), p. 258) is given by:

$$\mu = \left(I_N - B_1 - \ldots - B_P \right)^{-1} B_c \tag{1.4}$$

By substituting equation (1.3) in equation (1.1) and rearranging, the VAR process can be reparametrized equivalently as:

$$Y_t - \mu = \sum_{p=1}^{P} B_p \left(Y_{t-p} - \mu \right) + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \Sigma \right) \tag{1.5}$$

By using this representation, it is easy to show how the VAR process expressed in a gap from its steady-state will converge asymptotically to the steady-state under iterative forecasting. To keep the analysis tractable, a VAR(1) process is assumed. The $h = 1$ step ahead deterministic component of the forecast will be $Y_{T+1|T} - \mu = B_1 \left(Y_T - \mu \right)$, and by backward substitution, the h-step ahead forecast will be $Y_{T+1|t} - \mu = B_1^h \left(Y_T - \mu \right)$ or equivalently:

$$Y_{T+h|T} = \mu + B_1^h \left(Y_T - \mu \right) \tag{1.6}$$

From equation (1.6), it is obvious that asymptotically the forecasts will converge to the steady-state: $\lim_{h \to \infty} Y_{T+h|T} = \mu + \lim_{h \to \infty} B_1^h \left(Y_T - \mu \right) = \mu$. What controls the speed of convergence to the steady-state is the overall persistence of the VAR process influenced by the $(B_1)_{m,n}$ elements. For example, for the simplistic case that the off-diagonal elements of B_1 are zero, implying no explanatory power from each variable to the others, obviously, the persistence is controlled by how close the diagonal elements of B_1 are to unity.

1.3.2 Steady-state BVAR

This section presents more details of the homoscedastic BVAR implemented with the steady-state prior. Firstly, all the details of the steady-state BVAR's prior and posterior estimation are presented analytically, and afterwards, the step-by-step procedure of the marginal likelihood estimation is set out.

1.3.2.1 Steady-state BVAR estimation

The usual parametrisation of a stationary VAR, as given in equation (1.1), does not allow to impose directly a prior over its steady-state (or unconditional mean) μ, and the standard approach when using a Minnesota type prior is to set the prior diffuse for the constant[9]. For this reason, Villani (2009) suggested using the VAR parametrization of equation (1.5), where a prior for the steady-state μ can be easily applied[10]:

$$B(L)(Y_t - \mu) = \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma) \tag{1.7}$$

Where $B(L) = I_N - B_1 L - \cdots - B_P L^P$ with L being the backshift operator: $L^p Y_t = Y_{t-p}$.

Using the above VAR parametrisation in a fully Bayesian setting requires the use of priors for the autoregressive coefficients B, the variance-covariance matrix Σ, and the steady-state μ as follows in more detail.

Firstly, the coefficients B of the steady-state prior BVAR follow a diagonal multivariate normal distribution which is akin to Litterman (1986) and follow a Minnesota-style shrinkage.

$$\text{vec}(B) \sim \mathcal{N}(\underline{\text{M}}_B, \underline{\Omega}_B)^{11} \tag{1.8}$$

The first moment of coefficients B prior is set as:

$$E\left[\left(B_p\right)_{m,n}\right] = \begin{cases} c_n & \text{if } p = 1, m = n \\ 0 & \text{otherwise} \end{cases} \tag{1.9}$$

The variance of the coefficients' prior is set as:

$$\text{var}\left[\left(B_p\right)_{m,n}\right] = \begin{cases} \dfrac{\lambda_1^2}{p^{\lambda_3}}, & m = n \\ \lambda_2 \dfrac{\lambda_1^2}{p^{\lambda_3}} \dfrac{\sigma_m^2}{\sigma_n^2}, & \text{otherwise} \end{cases} \tag{1.10}$$

---

[9] Carriero, Clark, and Marcellino (2015a) is a notable exception where the prior of the constant is informative but rather set in a way that does not have any direct interpretation with the steady-state.

[10] Villani (2009) allows for the more generic case that the steady-state can be affected by deterministic trends as well: $B(L)(Y_t - \Psi d_t) = \varepsilon_t$, where$\Psi$ is an $N \times q$ matrix and d_t is column vector $q \times 1$ of deterministic variables. In this study obviously $\Psi d_t = \mu$.

[11] vec(·) vectorizes a matrix $m \times n$ into a column vector $mn \times 1$ by stacking all of its columns.

Secondly, the prior for the variance-covariance matrix Σ is inverse Wishart:

$$\Sigma \sim \mathcal{IW}\left(\underline{S}_{\Sigma}, \underline{d}_{\Sigma}\right) \tag{1.11}$$

Where $\underline{S}$ is the prior scale matrix, and $\underline{d}$ are the degrees of freedom.

Finally, in addition to the prior for the case of a standard BVAR with an intercept, now an additional prior for the steady-state μ is defined:

$$\mu \sim \mathcal{N}\left(\underline{\mathrm{M}}_{\mu}, \underline{\Omega}_{\mu}\right) \tag{1.12}$$

Where $\underline{\mathrm{M}}_{\mu}$ is the prior mean, and $\underline{\Omega}_{\mu}$ is the variance for the steady-state. $\underline{\Omega}_{\mu}$ is assumed to be diagonal. More details on how the steady-state prior mean – $\underline{\mathrm{M}}_{\mu}$, and variance – $\underline{\Omega}_{\mu}$ are set using the SPF dataset are presented in Section 1.5.

Bayesian inference in the steady-state BVAR can be performed using a Gibbs sampler which comprises the following three blocks – see Villani (2009):

1. $p\left(B^{(r)} \middle| \Sigma^{(r-1)}, \mu^{(r-1)}, Y\right)$
2. $p\left(\Sigma^{(r)} \middle| B^{(r)}, \mu^{(r-1)}, Y\right)$
3. $p\left(\mu^{(r)} \middle| B^{(r)}, \Sigma^{(r)}, Y\right)$

Where (r) denotes the $r = 1,\cdots, R$ Gibbs sampler iteration.

The first and the second blocks are estimated similar to the case for the independent normal - inverse Wishart prior but for the VAR process, which has been demeaned with μ, i.e. Y_t is substituted with $\tilde{Y}_t = Y_t - \mu$. To keep the VAR in matrix notation consistent with equation (1.2), it follows that $\tilde{Y}$ and $\tilde{X}$ refer to the demeaned time series. The first block, in more detail, draws the coefficients B from the conditional posterior distribution.

$$vec\left(B\right) \sim \mathcal{N}\left(\overline{\mathrm{M}}_{B}, \overline{\Omega}_{B}\right) \tag{1.13}$$

Where $\overline{\Omega}_{\mathbf{B}} = \left(\Sigma^{-1} \otimes \left(\tilde{\mathrm{X}}'\tilde{\mathrm{X}}\right) + \underline{\Omega}_{B}^{-1}\right)^{-1}$, $\overline{\mathrm{M}}_{B} = \overline{\Omega}_{B}\left(\underline{\Omega}_{B}^{-1}\underline{\mathrm{M}}_{B} + \left(\Sigma^{-1} \otimes \tilde{\mathrm{X}}'\right)vec\left(\tilde{Y}\right)\right)$

The second block draws the variance-covariance matrix Σ from its conditional posterior distribution:

$$\Sigma \sim \mathcal{IW}\left(\overline{S}_{\Sigma}, \overline{d}_{\Sigma}\right) \tag{1.14}$$

Where $\overline{S}_{\Sigma} = \underline{S}_{\Sigma} + \left(\tilde{Y} - \tilde{X}B\right)'\left(\tilde{Y} - \tilde{X}B\right)$ and $\overline{d}_{\Sigma} = \underline{d}_{\Sigma} + T$

Finally, the third block concerning the posterior draws of the steady-state samples from the following posterior – see Villani (2009):

$$\mu \sim \mathcal{N}\left(\overline{\mathrm{M}}_{\mu}, \overline{\Omega}_{\mu}\right) \tag{1.15}$$

Where $\overline{\Omega}_\mu = \left(U'\left(\left(D'D\right)\otimes\Sigma^{-1}\right)U + \underline{\Omega}_\mu^{-1}\right)^{-1}$, $\overline{\mathrm{M}}_\mu = \overline{\Omega}_\mu\left(U'vec\left(\Sigma^{-1}\left(\mathrm{B}\left(L\right)Y\right)'D\right) + \underline{\Omega}_\mu^{-1}\underline{\mathrm{M}}_\mu\right)$,

with $D = \left[1_{T\times 1} \mid -1_{T\times P}\right]$ and $U = \left[I_N \mid B_1', \cdots, B_P'\right]'$.

In order to estimate both the unconditional and conditional forecasts, as suggested in Blake and Mumtaz (2017), the steady-state demeaned VAR is reparametrized back into a VAR with a constant B_c by reordering equation (1.7) as a standard VAR process in equation (1.1) with the constant given by equation (1.3).

1.3.2.2 Steady-state BVAR marginal likelihood estimation

The marginal likelihood for conjugate priors like the original Minnesota prior, sum-of-coefficients and the 'dummy-initial-observation'' implementation of the natural conjugate normal-inverse Wishart prior can be estimated analytically in a closed form (Giannone, Lenza, & Primiceri, 2015). The usual approach in the literature is to use a grid search over a sensible range of hyperparameters in order to select them optimally (Dieppe, Legrand, & Van Roye, 2016). More recently, Chan, Jacobi, and Zhu (2020) have suggested a more computationally efficient way to perform this optimisation problem by using automatic differentiation.

However, for BVARs which have non-conjugate priors or include latent states where MCMC methods are required for the estimation of the posterior, an analytical solution of the marginal likelihood does not exist[12]. There are a few notable exceptions in the literature where the optimal hyperparameters are estimated, including Schorfheide and Song (2015), who use a grid search with the modified harmonic mean estimator (Gelfand & Dey, 1994; Geweke, 1999) for a mixed-frequency VAR, and Ankargren, Unosson, and Yang (2018) who use an adaptive grid search with the improved Chib (1995) estimator of Fuentes-Albero and Melosi (2013) for the case of a mixed-frequency BVAR with a steady-state prior. For the same case of a homoscedastic BVAR with steady-state prior, Gustafsson, Villani, and Stockhammar (2020) propose Bayesian optimisation algorithms for hyperparameters optimisation as an alternative to an exhaustive grid search.

In my case, I estimate the optimal hyperparameters' values λ_1, λ_2, and λ_3 by maximising the marginal likelihood of the corresponding VAR models. Since an analytical, closed-form solution does not exist, I resort to the approximation method of Chib (1995). I repeat this process for each vintage of the forecasting exercise, rather than estimating the optimal values given the full sample, and subsequently, examine whether the optimal hyperparameters' values fluctuate significantly over time (see Figure 1.5). The optimisation is done using particle swarm optimisation (PSO). More information on the PSO algorithm can be found in Appendix A.1.1 Particle swarm optimisation.

Below follows a demonstration of approximating the log marginal likelihood for a homoscedastic BVAR with a steady-state prior. Starting from Bayes' rule:

[12] Koop, McIntyre, Mitchell, and Poon (2020) propose a Metropolis-Hastings step within the Gibbs sampler to estimate optimal hyperparameters values for heteroscedastic BVARs using Minnesota prior without the need to directly estimate or approximate the marginal likelihood.

$$p\left(\theta|Y,\mathcal{M}\right)=\frac{p\left(Y|\theta,\mathcal{M}\right)p\left(\theta,\mathcal{M}\right)}{p\left(Y|\mathcal{M}\right)}$$

By simple re-ordering, the log marginal likelihood is equal to:

$$\ln p\left(Y|\mathcal{M}\right)=\ln p\left(Y|\theta,\mathcal{M}\right)+\ln p\left(\theta,\mathcal{M}\right)-\ln p\left(\theta|Y,\mathcal{M}\right)$$

This formula is valid at any point of the posterior distribution of parameters θ; however, for the approximation to be as accurate as possible, a point of support $\tilde{\theta}$ with high density is chosen, where $\tilde{\theta}$ is the mean of the posterior's distribution. In the following equations, the index $\mathcal{M}$ representing the model hyperparameters is dropped for notational simplicity.

The term representing the VAR's likelihood is:

$$p\left(Y|\tilde{\theta}\right)=\left(2\pi\right)^{-0.5TN}\left|\tilde{\Sigma}\right|^{-0.5T}\exp\left(-0.5vec\left(E\right)\left(\tilde{\Sigma}\otimes I_{T}\right)^{-1}vec\left(E\right)\right)$$

For the steady-state prior marginal likelihood estimation, the VAR model's prior and posterior probability is comprised of the three terms corresponding to the priors of B, Σ and μ. Consequently, the steady-state prior term is decomposed as:

$$\ln p\left(\tilde{\theta}\right)=\ln p\left(\tilde{B}\right)+\ln p\left(\tilde{\mu}\right)+\ln p\left(\tilde{\Sigma}\right)$$

In the same manner, the posterior distribution is comprised of the following three parts:

$$\ln p\left(\tilde{\theta}\middle|Y\right)=\ln p\left(\tilde{B},\tilde{\mu},\tilde{\Sigma}\middle|Y\right)=\ln p\left(\tilde{B}\middle|\tilde{\mu},\tilde{\Sigma},Y\right)+\ln p\left(\tilde{\mu}|\tilde{\Sigma},Y\right)+\ln p\left(\tilde{\Sigma}\middle|Y\right)$$

Chib (1995) showed that the first term $p\left(\tilde{B}\middle|\tilde{\mu},\tilde{\Sigma},Y\right)$ can be estimated by evaluating the pdf of the posterior of B as defined for $\tilde{\mu}$ and $\tilde{\Sigma}$ at the point $\tilde{B}$:

$$p_{\mathcal{N}}\left(\tilde{B}\middle|\overline{\mathrm{M}}_{B},\overline{\Omega}_{B}\right)$$

Where: $\overline{\Omega}_{B}=\left(\tilde{\Sigma}^{-1}\otimes\left(\tilde{\mathrm{X}}'\tilde{\mathrm{X}}\right)+\underline{\Omega}_{B}^{-1}\right)^{-1}$, $\overline{\mathrm{M}}_{B}=\overline{\Omega}_{B}\left(\underline{\Omega}_{B}^{-1}\underline{\mathrm{M}}_{B}+\left(\tilde{\Sigma}^{-1}\otimes\tilde{\mathrm{X}}'\right)vec\left(\tilde{Y}\right)\right)$

The third term can be approximated following the 'Rao-Blackwellization' technique as follows:

$$p\left(\tilde{\Sigma}\middle|Y\right)\simeq\frac{1}{R}\sum_{r=1}^{R}p\left(\tilde{\Sigma}\middle|B^{(r)},\mu^{(r)},Y\right)$$

Where $r = 1,\cdots, R$ represent Gibbs sampler retained draws.
Each term of the summation $p\left(\tilde{\Sigma}\middle|B^{(r)},\mu^{(r)},Y\right)$ is the pdf of the inverse Wishart with degrees of freedom: $\overline{d}=\underline{d}+T$ and scale matrix: $\overline{S}=\underline{S}+\left(\tilde{Y}-\tilde{X}B^{(r)}\right)'\left(\tilde{Y}-\tilde{X}B^{(r)}\right)$ evaluated at $\tilde{\Sigma}$: $p_{\mathcal{IW}}\left(\tilde{\Sigma}\middle|\overline{S},\overline{d}\right)$.

However, the second term $p\left(\tilde{\mu}\middle|\tilde{\Sigma},Y\right)$, as suggested in Chib (1995), requires a further additional reduced Gibbs sampler's run with retained draws $r'=1,\cdots,R'$. These reduced runs are performed simply by running the Gibbs sampler with two blocks conditional on $\tilde{\Sigma}$. In more detail, the Gibbs sampler is based on the following two blocks:

1. $p\left(B^{(r')}\middle|\mu^{(r'-1)},\tilde{\Sigma},Y\right)$
2. $p\left(\mu^{(r')}\middle|B^{(r')},\tilde{\Sigma},Y\right)$

Consequently, the term $p\left(\tilde{\mu}\middle|\tilde{\Sigma},Y\right)$ can be approximated as:

$$p\left(\tilde{\mu}\middle|\tilde{\Sigma},Y\right)\simeq\frac{1}{R'}\sum_{r'=1}^{R'}p\left(\tilde{\mu}\middle|B^{(r')},\tilde{\Sigma},Y\right)$$

Each term of the summation $p\left(\tilde{\mu}\middle|B^{(r')},\tilde{\Sigma},Y\right)$ evaluates the pdf of the normal distribution at $\tilde{\mu}$ as defined below. The variance-covariance matrix is $\overline{\Omega}_{\mu}=\left(U'\left(\left(D'D\right)\otimes\tilde{\Sigma}^{-1}\right)U+\underline{\Omega}_{\mu}^{-1}\right)^{-1}$, and the mean is defined as $\overline{\mathrm{M}}_{\mu}=\overline{\Omega}_{\mu}\left(U'vec\left(\tilde{\Sigma}^{-1}\left(\mathrm{B}(L)^{(r')}Y\right)'D\right)+\underline{\Omega}_{\mu}^{-1}\underline{\mathrm{M}}_{\mu}\right)$, where $D=\left[1_{T\times1}\ \middle|\ -1_{T\times P}\right]$ and $U=\left[I_N\ \middle|\ B_1^{(r')},\cdots,B_P'^{(r')}\right]'$. Thus, each term of the summation is equal to $p_{\mathcal{N}}\left(\tilde{\mu}\middle|\overline{\mathrm{M}}_B,\overline{\Omega}_B\right)$.

1.3.3 Heteroscedastic steady-state BVAR with stochastic volatility

By further relaxing the assumption of homoscedasticity and adding stochastic volatility (SV), the previous model is now defined as:

$$B(L)(Y_t-\mu)=\varepsilon_t,\quad \varepsilon_t\sim\mathcal{N}(0,\Sigma_t) \tag{1.16}$$

The inclusion of stochastic volatility, as in Cogley and Sargent (2005), transforms the previous model in equation (1.7) with the constant variance-covariance matrix Σ to innovations whose distribution has a time-varying structure Σ_t defined as:

$$\Sigma_t=A^{-1}L_tA^{-1\prime} \tag{1.17}$$

Matrix A is a lower triangular matrix $N\times N$ with ones in the diagonal[13]:

[13] Primiceri (2005) allows matrix A to be time-varying, i.e. A_t, with individual elements components $\alpha_{m,n}$ below the diagonal following a random walk: $\alpha_{m,n,t}=\alpha_{m,n,t-1}+\varphi_t,\ \ \varphi_t\sim\mathcal{N}(0,\sigma^2)$.

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \alpha_{2,1} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \alpha_{N,1} & \alpha_{N,2} & \cdots & \alpha_{N,N-1} & 1 \end{bmatrix} \tag{1.18}$$

Elements $\alpha_{m,n}$ below the diagonal, similar to Cogley and Sargent (2005), follow a normal prior distribution. Since elements $\alpha_{m,n}$ are estimated for each row of matrix A independently; it is more convenient to define the prior for all elements $\alpha_{m,n}$, $2 \leq m \leq N$, $1 \leq n \leq m-1$ per row m:

$$\alpha_{m,\bullet} \sim \mathcal{N}\left(\underline{\mathrm{M}}_{\alpha_{m,\bullet}}, \underline{\Omega}_{\alpha_{m,\bullet}}\right) \tag{1.19}$$

The term L_t in equation (1.17) is a diagonal matrix $N \times N$, and its diagonal elements $L_{n,n}$ for $n = 1,\cdots, N$ are defined as $L_{n,n,t} = h_{n,t}$, where the natural logarithm of $h_{n,t}$ follows a random walk[14]:

$$\ln h_{n,t} = \ln h_{n,t-1} + \eta_{n,t}\,,\ \eta_{n,t} \sim \mathcal{N}\left(0, \Phi_{n,n}\right) \tag{1.20}$$

Stacking all equations' $n = 1,\cdots, N$ stochastic volatilities $h_{n,t}$ and their corresponding innovations, it follows that $\ln h_t = \ln h_{t-1} + \eta_t$. Cogley and Sargent (2005) define matrix Φ as a diagonal matrix[15] whose diagonal elements $\Phi_{n,n}$ follow an inverse gamma prior with scale matrix $\underline{S}$ and degrees of freedom $\underline{d}$:

$$\Phi_{n,n} \sim \mathcal{IG}\left(\underline{S}_{\Phi_{n,n}}, \underline{d}_{\Phi_{n,n}}\right) \tag{1.21}$$

The steady-state BVAR with stochastic volatility is simply a combination of the building blocks from the previous model by further adding the blocks relevant to the stochastic volatility. Thus, the iteration (r) from the Gibbs sampler is based now on the following five blocks:

1. $p\left(B^{(r)} \middle| A^{(r-1)}, h^{(r-1)}, \Phi^{(r-1)}, \mu^{(r-1)}, Y\right)$
2. $p\left(A^{(r)} \middle| B^{(r)}, h^{(r-1)}, \Phi^{(r-1)}, \mu^{(r-1)}, Y\right)$
3. $p\left(h^{(r)} \middle| B^{(r)}, A^{(r)}, \Phi^{(r-1)}, \mu^{(r-1)}, Y\right)$
4. $p\left(\Phi^{(r)} \middle| B^{(r)}, A^{(r)}, h^{(r)}, \mu^{(r-1)}, Y\right)$
5. $p\left(\mu^{(r)} \middle| B^{(r)}, A^{(r)}, h^{(r)}, \Phi^{(r)}, Y\right)$

[14] Clark and Ravazzolo (2015) find the random walk stochastic volatility specification superior in terms of forecast accuracy with respect to BVARs with alternative time-varying volatilities.

[15] Primiceri (2005) allows the variance-covariance matrix Φ to be a full matrix following an inverse Wishart prior: $\Phi \sim \mathcal{IW}(\underline{S}_{\Phi}, \underline{d}_{\Phi})$. This specification allows all variables' stochastic volatilities components to be hit by a common shock via their common innovation term η_t, however with the drawback that as the n dimension of the VAR increases, the prior on the variance-covariance of η_t becomes highly informative.

The first block draws the coefficients B from the conditional posterior distribution:

$$vec(B) \sim \mathcal{N}\left(\overline{\mathrm{M}}_B, \overline{\Omega}_B\right)$$

Where $\overline{\Omega}_B = \left(\underline{\Omega}_B^{-1} + \sum_{t=1}^{T} \Sigma_t^{-1} \otimes \left(\tilde{X}_t \tilde{X}_t'\right)\right)^{-1}$, $\overline{\mathrm{M}}_B = \overline{\Omega}_B \left(\underline{\Omega}_B^{-1} \underline{\mathrm{M}}_B + vec\left(\sum_{t=1}^{T} \tilde{X}_t' \tilde{Y}_t \Sigma_t^{-1}\right)\right)$

The second block draws from the posterior of $\alpha_{m,\bullet}$ for $2 \le m \le N$ and is given by the following transformed regressions:

$$\varepsilon_{m,t} h_{m,t}^{-1/2} = \sum_{k=1}^{m-1} \alpha_{m,k} \left(-\varepsilon_{k,t} h_{m,t}^{-1/2}\right) + h_{m,t}^{-1/2} \xi_{m,t}, \quad \xi_{m,t} \sim \mathcal{N}(0,1)$$

Defining as Z_m and z_m the left-hand and right-hand side variables for each equation $2 \le m \le N$ respectively, the conditional posterior distributions of elements $\alpha_{m,\bullet}$ is given:

$$\alpha_{m,\bullet} \sim \mathcal{N}\left(\overline{\mathrm{M}}_{\alpha_{m,\bullet}}, \overline{\Omega}_{\alpha_{m,\bullet}}\right)$$

Where $\overline{\Omega}_{\alpha_{m,\bullet}} = \left(\underline{\Omega}_{\alpha_{m,\bullet}}^{-1} + \mathrm{Z}_m \mathrm{Z}_m\right)^{-1}$, $\overline{\mathrm{M}}_{\alpha_{m,\bullet}} = \overline{\Omega}_{\alpha_{m,\bullet}} \left(\underline{\Omega}_{\alpha_{m,\bullet}}^{-1} \underline{\mathrm{M}}_{\alpha_{m,\bullet}} + \mathrm{Z}_m' z_m\right)$

The third block is estimated following the algorithm[16] of Kim, Shephard, and Chib (1998). More details can be found in Appendix A.1.3 Stochastic volatility estimation.

The fourth block draws from the posterior of Φ independently for each equation *i*:

$$\Phi_{n,n} \sim \mathcal{IG}\left(\overline{S}_{\Phi_{n,n}}, \overline{d}_{\Phi_{n,n}}\right)$$

Where $\overline{S}_{\Phi_{n,n}} = \underline{S}_{\Phi_{n,n}} + \sum_{t=1}^{T} \left(\Delta \ln h_{n,t}\right)^2$ and $\overline{d}_{\Phi_{n,n}} = \underline{d}_{\Phi_{n,n}} + T$

For the fifth block, the posterior draws $\mu^{(r)}$ of the steady-state are defined as follows – see Clark (2011):

$$\mu \sim \mathcal{N}\left(\overline{\mathrm{M}}_\mu, \overline{\Omega}_\mu\right)$$

Where $\overline{\Omega}_\mu = \left(U' \left(\sum_{t=1}^{T} \left(D_t' D_t \otimes \Sigma_t^{-1}\right)\right) U + \underline{\Omega}_\mu^{-1}\right)^{-1}$,

$\overline{\mathrm{M}}_\mu = \overline{\Omega}_\mu \left(U' vec\left(\sum_{t=1}^{T} \Sigma_t^{-1} \left(\mathrm{B}(L) Y_t\right) D_t\right) + \underline{\Omega}_\mu^{-1} \underline{\mathrm{M}}_\mu\right)$, with $D = \left[1_{T\times 1}, -1_{T\times P}\right]$ and

$U = \left[I_N, \mathrm{B}_1', \cdots, \mathrm{B}_P'\right]'$.

[16] Cogley and Sargent (2005) estimate stochastic volatility following Jacquier, Polson, and Rossi (1994), while Primiceri (2005) and Carriero, Clark, and Marcellino (2019) amongst others follow Kim, Shephard, and Chib (1998). For a comparison of these two algorithms refer to Geweke, Koop, and van Dijk (2011), p. 476.

1.4 Conditional forecasting using Kalman filtering

Conditional forecasts are produced following the Bańbura, Giannone, and Lenza (2015) approach. This approach is an extension of the Clarida and Coyle (1984) original method, which casts a reduced form VAR into a state-space model and then uses Kalman filtering to infer the conditional forecasts as states by treating them as missing observations in the observation equation[17]. The further improvement by Bańbura, Giannone, and Lenza (2015) on this part involves the use of a simulation smoother for the state-space model rather than the simple Kalman filter and smoother proposed in Clarida and Coyle (1984), allowing the estimation of the entire predictive density of the conditional forecasts and not only the point estimates of the conditional forecasts.

The methodology used for the conditional forecasts is presented analytically below, starting by defining the state-space model representation of the VAR process:

$$Y_t = GS_t + e_t \tag{1.22}$$

$$S_t = C + FS_{t-1} + v_t \tag{1.23}$$

For the observation equation (1.22):

G is an $N \times NP$ matrix defined as $G = \left[I_N, 0_{N \times N(P-1)} \right]$ and $e_t \sim \mathcal{N}\left(0, R_t\right)$ with $R_t = 0_{N \times N}$. The state S_t is a column vector $NP \times 1$ and is defined as $S_t = \left[Y_t', Y_{t-1}', \cdots, Y_{t-P+1}' \right]'$.

For the state equation (1.23):

Column vector C is a $NP \times 1$ column vector defined as $C = \left[B_c', 0_{1 \times N(P-1)} \right]'$, and F is the $NP \times NP$ VAR companion matrix:

$$F = \left[\begin{array}{c|c} B_1, \cdots, B_{P-1} & B_P \\ \hline I_{N(P-1)} & 0_{N(P-1) \times N} \end{array} \right] \tag{1.24}$$

The innovation term v_t is distributed as $v_t \sim \mathcal{N}(0, Q_t)$. Q_t is $NP \times NP$ with the upper left elements $N \times N$ equal to Σ or Σ_t, depending on whether the underlying VAR model is homoscedastic or heteroscedastic with stochastic volatility, respectively:

$$Q_t = \left[\begin{array}{c|c} \Sigma_t & 0_{N \times N(P-1)} \\ \hline 0_{N(P-1) \times N} & 0_{N(P-1) \times N(P-1)} \end{array} \right] \tag{1.25}$$

Bańbura, Giannone, and Lenza (2015) estimate the computational time it takes for different algorithms to run[18] and conclude that the most efficient is the Durbin and Koopman (2002)

[17] For a detailed textbook treatment of missing observations refer to Harvey (1989), p. 143 and Durbin and Koopman (2012), p. 110.

[18] The algorithms being compared are: Carter and Kohn (1994), Durbin and Koopman (2002) implemented with the Harvey (1989) Kalman smoother, Durbin and Koopman (2002) implemented with the de Jong (1988) Kalman smoother and the Waggoner and Zha (1999) implemented as in

algorithm implemented using the de Jong (1988) Kalman smoother. I follow the same approach, although the estimated VAR models are relatively small ($N = 6$, $P = 4$, $H = 12$ step-ahead recursive forecasts) in comparison to Bańbura, Giannone, and Lenza (2015), who estimate conditional forecasts using a large $n = 26$ variable VAR for $H = 60$ conditioning periods.

In particular, I follow algorithm '2.a' of the Durbin and Koopman (2002) simulation smoother, as clarified explicitly in Jarociński (2015). The algorithm is comprised of 3 steps:

1. Follow the recursion of the state-space model equations (1.23) and (1.22) by making draws for the state-space model's innovations e_t and v_t from their respective distributions[19]. For each $t = T+1,\cdots, T+H$, I estimate S_t^+, Y_t^+ and finally save $Y^+ = \left[Y_{T+1}^+,\cdots,Y_{T+H}^+\right]'$ and $S^+ = \left[S_{T+1}^+,\cdots,S_{T+H}^+\right]'$. The recursion is initiated with: $S_0 = \left[y_T', y_{T-1}',\cdots,y_{T-p+1}'\right]'$. Conditional forecasts begin at $T+1$, and thus there is no need to estimate the states S_t for $t = 1,\cdots, T$.
2. Apply the Kalman filter and de Jong (1988) Kalman smoother to $Y^* = Y - Y^+$ and estimate the states $\hat{S}^*$. The intercept of the state equation (1.23) C is set to zero, as suggested in Jarociński (2015). The Kalman filter is initialised with $S_0 = 0_{NP\times 1}$, $\text{var}(S_0) = 10^{-9}$. For more information, refer to Appendix A.1.2 Kalman filtering and smoothing.
3. Finally, estimate a draw of the state as: $\tilde{S} = \hat{S}^* + S^+$. Conditional forecasts are given by taking the first n columns of the matrix $\tilde{S}$.

When performing the conditional forecasts, two approaches are employed. The first approach, henceforth named as 2-step estimation, uses the available data $Y_1,\cdots,Y_T$ to estimate the models' unknown parameters θ, i.e. $p(\theta^{(r)} \mid Y_1,\cdots, Y_T)$ and then generates conditional forecasts $\hat{Y}_{T+1},\cdots,\hat{Y}_{T+H}$ based on the conditioning information on some observable(s). For the case of models with stochastic volatility, conditional forecasting is performed using the last estimated variance-covariance matrix of the VAR innovations Σ_T.

The second approach, henceforth named as joint estimation, appends the conditional forecasts for each iteration (r) of the Gibbs sampler $\hat{Y}_{T+1},\cdots,\hat{Y}_{T+H}$ to the available data $Y_1,\cdots,Y_T$, i.e. $[Y_1,\cdots, Y_T, Y_{T+1},\cdots,Y_{T+H}]$ and then estimates the entire set of the unknown parameters based on this extended set of data: $p(\theta^{(r)} \mid Y_1,\cdots, Y_T, Y_{T+1},\cdots, Y_{T+H})$. For the models specified with stochastic volatility, this implies that the stochastic volatility[20] of the VAR innovation terms is also estimated for the conditional forecasts, i.e. $\Sigma_{T+1},\cdots, \Sigma_{T+H}$. This approach was originally described in Waggoner and Zha (1999) Gibbs sampler[21] and allows the finite-sample parameter uncertainty to be accounted for the conditional forecasts.

In terms of the conditioning variables, I focus only on the interest rates, which for central banks is the main policy instrument for conducting monetary policy, at least prior to the deployment of unconventional monetary policy tools (quantitative easing, forward guidance). Waggoner

---

Jarociński (2010). Other simulation smoothers include Frühwirth-Schnatter (1994), de Jong and Shephard (1995) and Chan and Jeliazkov (2009) .

[19] State equation innovations v_t are drawn using only the upper left $n \times n$ block of Q_t in equation (1.25).

[20] The Gibbs sampler in the second iteration uses the value Σ_T for $\Sigma_{T+1},\cdots,\Sigma_{T+H}$.

[21] Waggoner and Zha (1999) employ homoscedastic VAR models only, without stochastic volatility.

and Zha (1999), in their empirical application part, estimate forecasts conditional on the actual path of short-term interest rates for a single forecasting exercise, while Andersson, Palmqvist, and Waggoner (2010), in a similar fashion, estimate the forecasts conditional on the market expectations uncertainty over the short-term interest rates.

For the rest of the essay, I use as a conditioning variable the actual path of the short-term interest rates and perform the logical experiment of which prior would have superior performance assuming the policy-maker would have a perfect foresight over the evolution of the main policy instrument. Of course, this is an ex-post evaluation, while a real-time evaluation that would respect the information-set available to the policy-maker for the specific case at hand would require conditioning the forecasts on an assumed path over the short-term interest rates evolution as given by the central bank's own expectations, for example, market-implied or survey-based expectations.

As documented in Faust and Wright (2008), Clark and McCracken (2017), and McCracken and McGillicuddy (2019), in a frequentist framework, conditional forecasts in order to be assessed under the traditional measures of forecast accuracy and efficiency have to be conditioned on the actual 'true' future paths of a subset of the VAR variables and not on assumed ones. Clark and McCracken (2017) show that the properties of the conditional forecast errors can be decomposed into two separable factors: The properties of the unconditional forecast errors and the properties that are due to the conditioning information. Consequently, by using the ex-post realized values of the conditioning variables, it is feasible to eliminate the second factor that is due to the quality of the conditioning information and attribute the conditional forecasts properties and performance only to the properties of the underlying model that is used. In a Bayesian framework Angelini, Lalik, Lenza, and Paredes (2019) in a similar type of exercise intended to provide evidence that the proposed model indeed captures the statistical interdependencies and salient features of the data correctly, they proceed along the same direction, i.e. evaluating the BVAR's performance under the actual path of the assumptions as observed ex-post.

1.5 Priors and other estimation details

In order to compare the steady-state prior VAR specifications, I proceed to estimate the hyperparameters by maximizing the marginal likelihood for the homoscedastic case and then use these optimal values for the steady-state prior's VAR equivalent heteroscedastic specifications.

Across all BVARs, the mean $\underline{\mathrm{M}}_B$ of the Minnesota type prior in equation (1.9) is set with c_n equal to 0.25 for variables with fast mean-reverting properties (GDP growth and non-farm employment) and 0.8 for all the remaining persistent variables as in Clark (2011). The terms σ_m^2 and σ_n^2 are set equal to the corresponding AR(P)[22] estimated residuals variance by means of OLS.

For the homoscedastic VAR models, the inverse Wishart prior of Σ is set in the following manner: The degrees of freedom $\underline{d}$ are set equal to $N+2$, which are the least degrees of

[22] The number of lags of the AR(P) regressions is set equal to P, the same as for the VAR(P).

freedom such that the prior is properly defined. The scale matrix $\underline{S}$ is diagonal, and its elements $\underline{S}_{ii}$ are equal to the variance of the AR(P) residuals for each variable.

For the VAR models with stochastic volatility, I use a diffuse prior on elements $\alpha_{i,\cdot}$ in equation (1.19) with zero mean 0 and variance-covariance diagonal elements set equal to 10^4. Under the assumptions of a zero covariance between the shocks of the stochastic volatility, the prior of the diagonal elements of $\Phi_{n,n}$ in equation (1.21) is an inverse gamma with 6 degrees of freedom which make the prior relatively non-informative and with a scale set equal to a moderate value 0.1 which does not yield any excessive spikes during the early 80s and the financial crisis that would have increased the uncertainty of the corresponding vintages' forecasts in the midst of the financial crisis excessively.

Regarding the initialisation values for models with stochastic volatility for the first iteration of the Gibbs sampler, these are Σ_t for $t = 1,\cdots, T$ is set equal to Σ estimated for the homoscedastic VAR(P), $\ln h_{n,t}$ for $t = 1,\cdots, T$ and all equations $n = 1,\cdots, N$ is set equal to $\ln(0.8\ diag(\text{var}(Y)))$ and for the Kim, Shephard, and Chib (1998) algorithm I set diffuse initial conditions: $\ln h_0 = 0_{N\times 1}$, $\text{var}(\ln h_0) = 10\ I_N$ and $\Phi = 0.001\ I_N$.

Regarding the steady-state prior in equation (1.12), its mean $\underline{\mathrm{M}}_{\mu}$ is defined in the following manner: for the variables for which survey-based long-run expectations exist[23], these are set as the appropriate steady-state prior mean, while for the variables (e.g. employment) for which long-run expectations do not exist, the steady state prior mean is set equal to their unconditional mean. The variance of the steady-state prior $\underline{\Omega}_{\mu}$ is set according to Wright (2013): for the variables for which survey-based long-run expectations exist, their steady-state prior variance is set equal to 0.05, while for the variables (employment) for which there are no survey-based expectations a diffuse steady state prior is used.

For all models, in order to avoid any draws from the Gibbs sampler that would yield explosive forecasts where the VAR process would be in the non-stationary region, the standard approach in the literature is followed; that is, stability of the VAR is imposed, which also ensures its stationarity, i.e. all draws for which the VAR's companion form has eigenvalues with a modulus less than one[24] are discarded – see Lütkepohl (2005), p. 15.

As a benchmark prior, both for the unconditional and conditional forecasts, I employ the original Minnesota prior proposed in Litterman (1986), assuming the variance-covariance matrix Σ is known and set equal to the VAR OLS estimate $\hat{\Sigma}$. The prior hyperparameter values use the conventional values of $\lambda_1 = 0.2$, $\lambda_2 = 0.25$, and $\lambda_3 = 1$, while a diffuse prior is used for the intercepts.

Both for the homoscedastic and heteroscedastic BVAR models, the Gibbs sampler uses a burn-in sample of 5000 draws for convergence, and the subsequent 5000 draws are retained for inference and forecasting purposes.

[23] As already discussed in Section 1.2, in order to mitigate the impact of outliers, the SPF long-run expectations median response is used instead of the mean response.

[24] Formally this is defined as: the solution to $|\lambda\ I_{NP} - F| = 0$ has eigenvalues (roots) λ, whose modulus $|\lambda|$ lies inside the unit circle $|\lambda|<1$ for the generic case that the eigenvalues are complex numbers. F is the VAR companion form matrix defined in equation (1.24).

1.6 Results

This section presents all the empirical results relevant to BVARs with the steady-state prior both for unconditional and conditional forecasts. Firstly, Section 1.6.1 presents the in-sample results. In more detail, Section 1.6.1 deals with the optimal hyperparameter values with respect to maximizing the marginal likelihood for the homoscedastic BVARs, and subsequently, Section 1.6.1.2 explores the sensitivity of the steady-state posterior and long-run unconditional forecasts with regard to the steady-state prior tightness are discussed. In continuation, Section 1.6.1.3 shows the estimates of the stochastic volatility over the full sample.

Section 1.6.2 shows the out-of-sample forecast evaluation metrics. In particular, Section 1.6.2.1 discusses the full sample point forecasts, Section 1.6.2.2 provides the results for density forecasts, whereas the time variation of the results is further discussed in Section 1.6.2.3.

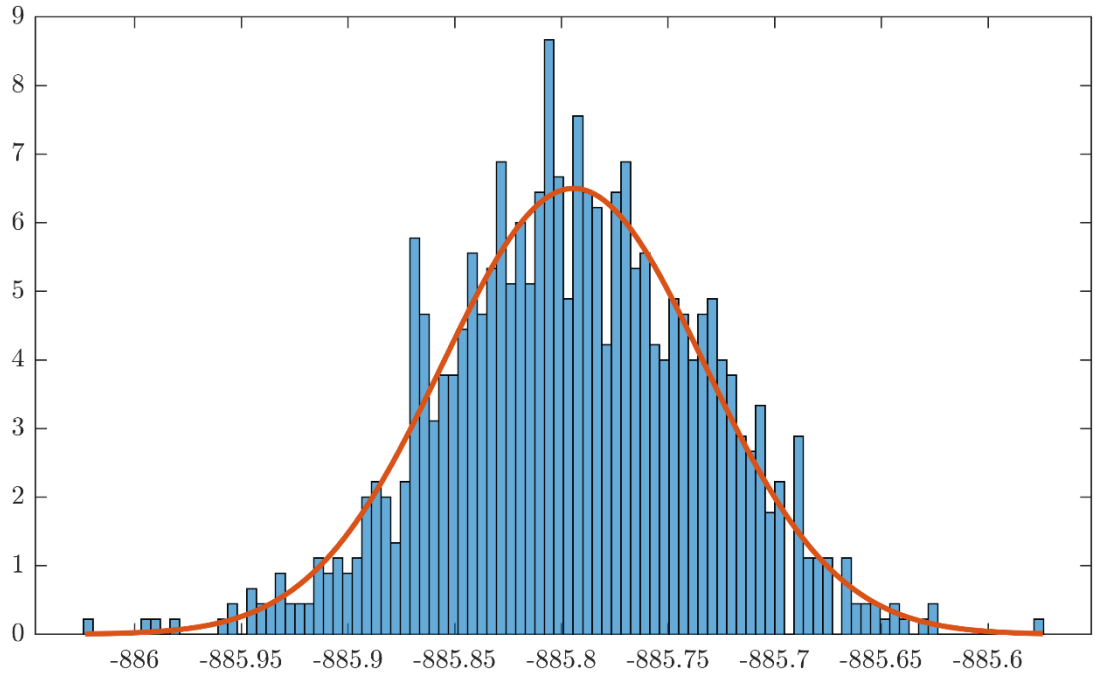
Finally, Section 1.6.3 deals with the models' reality check by examining how plausible the provided results would have been, especially in light of the financial crisis.

1.6.1 In-sample results

1.6.1.1 Homoscedastic BVAR steady-state prior hyperparameters estimation

In order to evaluate how precise the approximation of the marginal likelihood following the Chib (1995) method is, the log marginal likelihood is estimated 1000 times over the full sample 1980Q1-2013Q4 with conventional hyperparameter values where all MCMC starting values and hyperparameters are identical. Clearly, the approximation is relatively precise, and its standard error is 1.94×10^{-3}.

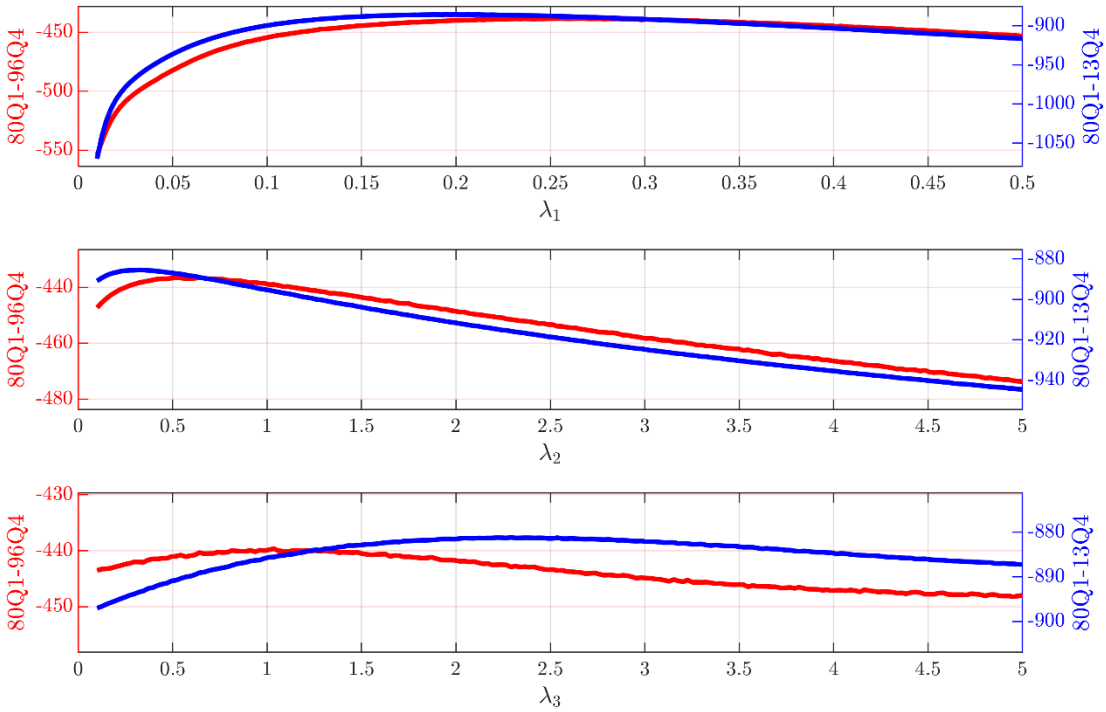
Figure 1.3: Log marginal likelihood histogram



Note. Histogram of 1000 evaluations of the log marginal likelihood with $\lambda_1 = 0.2$, $\lambda_2 = 0.25$, and $\lambda_3 = 1$ over 1980Q1-2013Q4.

The following figure shows the sensitivity of the log marginal likelihood to assumptions made with respect to each hyperparameter separately, while the others are kept constant at conventional values for the first vintage (1980Q1-1996Q4) and the last vintage (1980Q1-2013Q4). What stands out consistently from all the sub-plots is that the log marginal likelihood is concave within a sensible range of hyperparameter values, although for the case of λ_1 and λ_2, the curvature is relatively smooth. This fact, combined with the approximation error of the log marginal likelihood, albeit small, understandably should yield some noise in the optimized hyperparameter values.

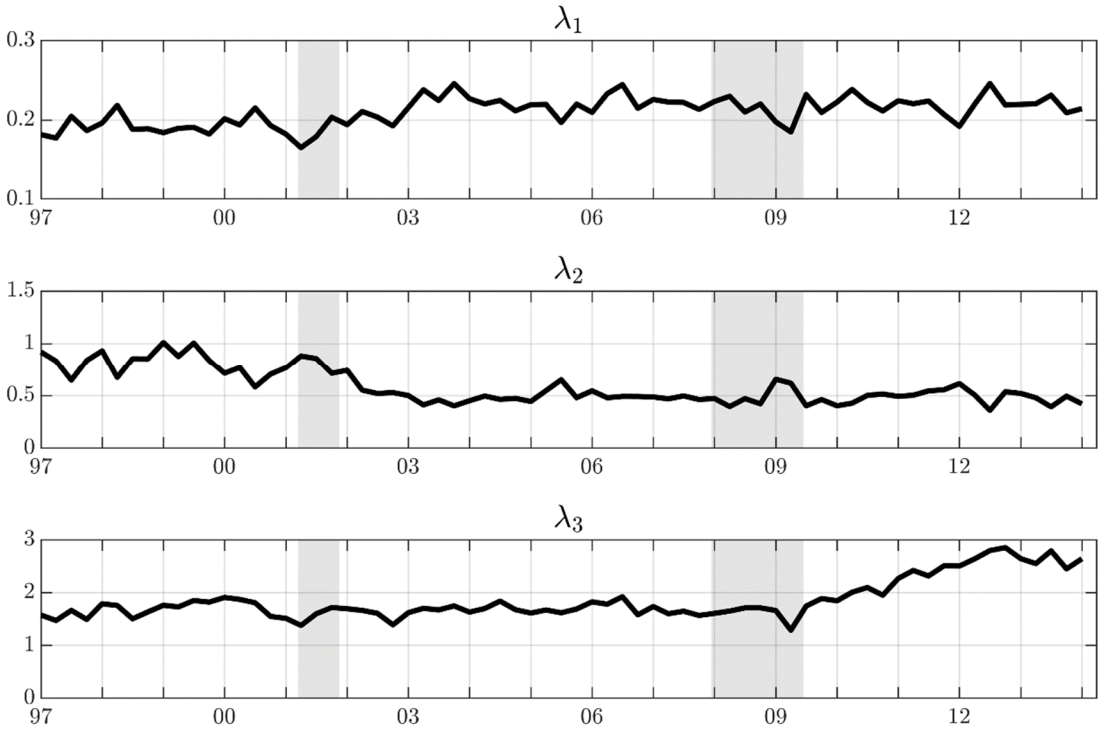
Figure 1.4: Log marginal likelihood sensitivity



Note. Hyperparameter values during log marginal likelihood sensitivity analysis are kept constant at $\lambda_1 = 0.2$, $\lambda_2 = 0.25$, $\lambda_3 = 1$. Left (red) and right (blue) y-axis refer to the first (1980Q1-1996Q4) and the last vintage (1980Q1-2013Q4).

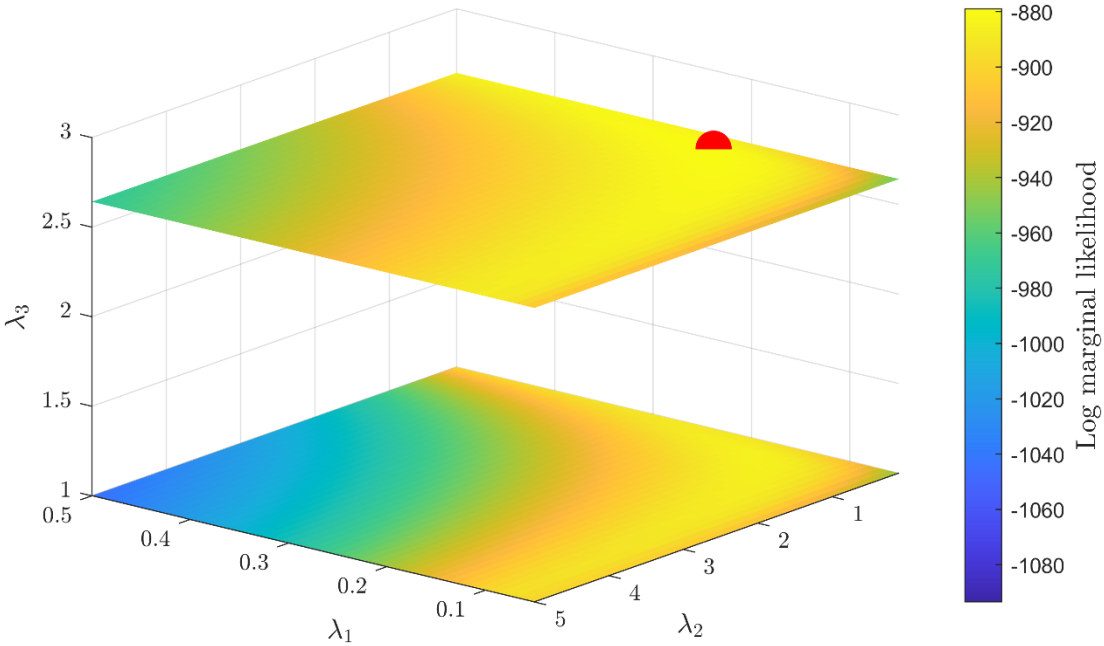
The following figure shows the optimal hyperparameter values as estimated for each vintage of the homoscedastic steady-state prior VAR. The results can be summarized as follows: Optimal λ_1 values fluctuate around the value of 0.2 with a tendency to increase slightly above 0.2 after 2004-05. Optimal λ_2 values fluctuate around the mid-range between 0.5 and 1, while after 2003 stabilize at around 0.5. Finally, optimal λ_3 values fluctuate between 1.5 and 2 until the financial crisis, whereas afterwards increase progressively from 2 to 3 as the estimation sample increases accordingly, implying a lower impact of the more distant lags.

Figure 1.5: Steady-state prior optimal hyperparameters for each vintage



Indicatively, the figure below shows for the steady-state prior how the log marginal likelihood changes over the space of hyperparameters λ_1 and λ_2 for a slice at the optimal value of $\lambda_3 = 2.64$ and the conventional value of $\lambda_3 = 1$ controlling the lag decay.

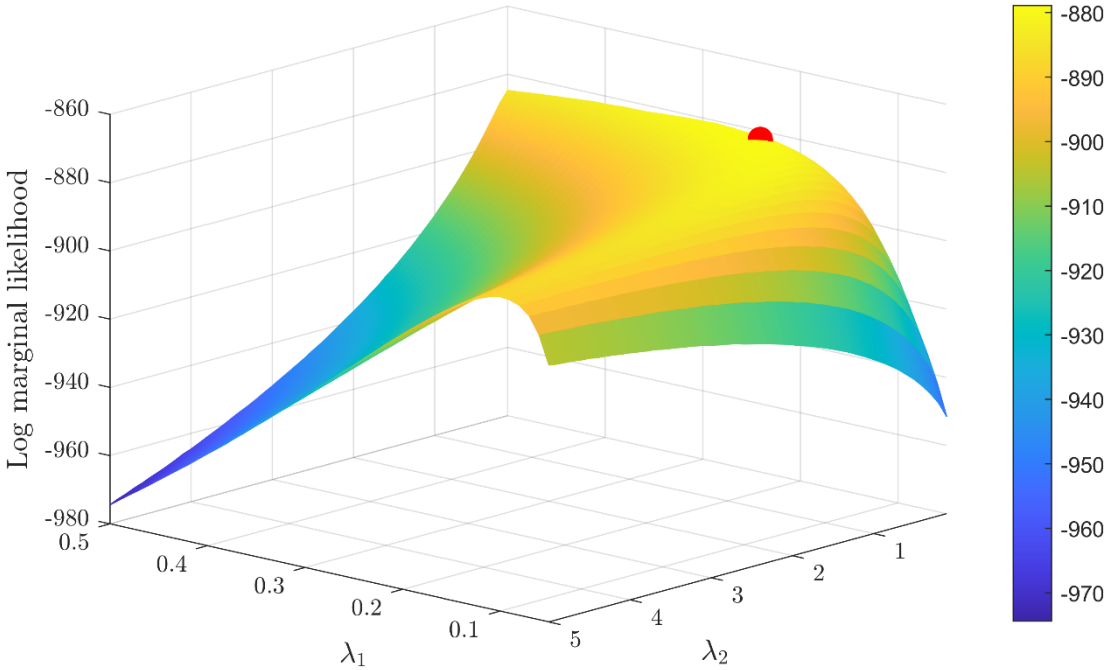
Figure 1.6: Steady-state prior log marginal likelihood heatmap



Note. Heatmap is shown for the optimal $\lambda_3 = 2.64$ and $\lambda_3 = 1$ for reference purposes. Estimation sample 1980Q1-2013Q4. The red dot shows the maximum log marginal likelihood.

Finally, the figure below shows the surface of the log marginal likelihood conditional on the optimal value $\lambda_3 = 2.64$ for the steady-state prior using the last vintage (1980Q1-2013Q4).

Figure 1.7: Steady-state prior log marginal likelihood surface



Note. Log marginal likelihood surface conditional on the optimal $\lambda_3 = 2.64$. The red dot shows the maximum log marginal likelihood.

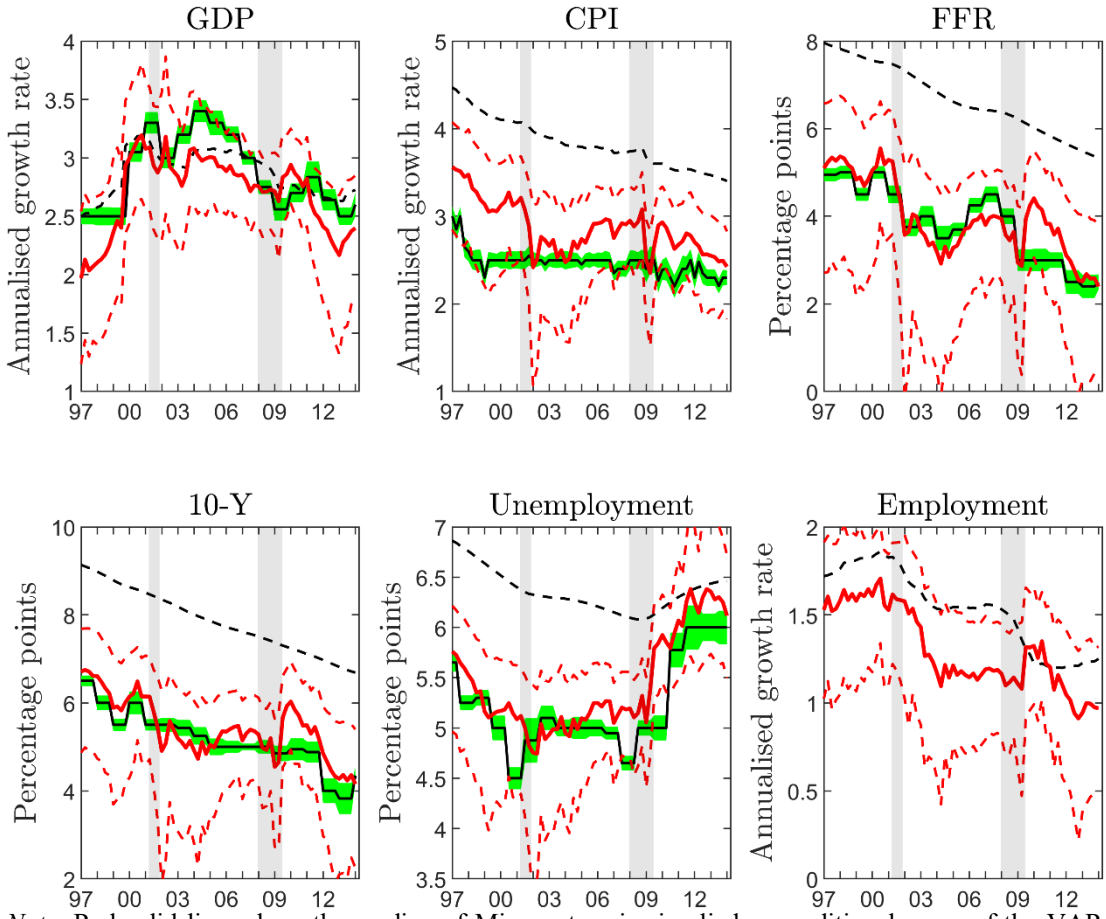
1.6.1.2 Minnesota and steady-state prior's long-run behaviour

In order to evaluate the SPF long-run expectations, firstly, I evaluate the in-sample implied unconditional mean of the VAR process with the benchmark Minnesota prior (which uses a diffuse prior on the constant) and plot them in the following figure. Two facts stand out. Firstly, for all variables in almost the entire evaluation sample, the implied unconditional mean moves in line with the SPF long-run expectations meaning that they capture any increasing or decreasing trends correctly.

Secondly, there is a visible distinction between variables with fast mean-reverting properties (GDP) versus the more persistent series (CPI, FFR, 10-year bond yield and unemployment). For GDP, there are instances where the gap between SPF and the benchmark-Minnesota prior's implied unconditional means differs by 0.5% in annualised growth rates during the 1997-2000 and to a lesser degree during the 2001-2008 period. However, for the most persistent variables the Minnesota prior's implied unconditional mean moves very closely with the SPF long-run expectations, and the series even coincide during some sub-periods. Noteworthily, for the case of CPI inflation, the Minnesota prior's implied unconditional mean remains consistently higher than the SPF long-run expectations throughout the entire sample, except for the recessionary periods.

This finding should not be surprising given the fact that potentially a large part of the SPF respondents may have based their estimates on variants of similar econometric models further augmented by judgemental input to capture informational content from soft indicators and the sometimes fast evolution of economic sentiment which is likely to be reflected in hard data relatively slowly. Nonetheless, although reassuring, the fact that the benchmark-Minnesota prior behaviour is to an extent aligned with SPF naturally poses the question of to what extent the use of the SPF informational content in a tight steady-state prior would yield any significant improvement in results compared to the use of a purely uninformative steady-state prior or the benchmark Minnesota prior which is diffuse for the constant.

Figure 1.8: Minnesota prior implied unconditional mean

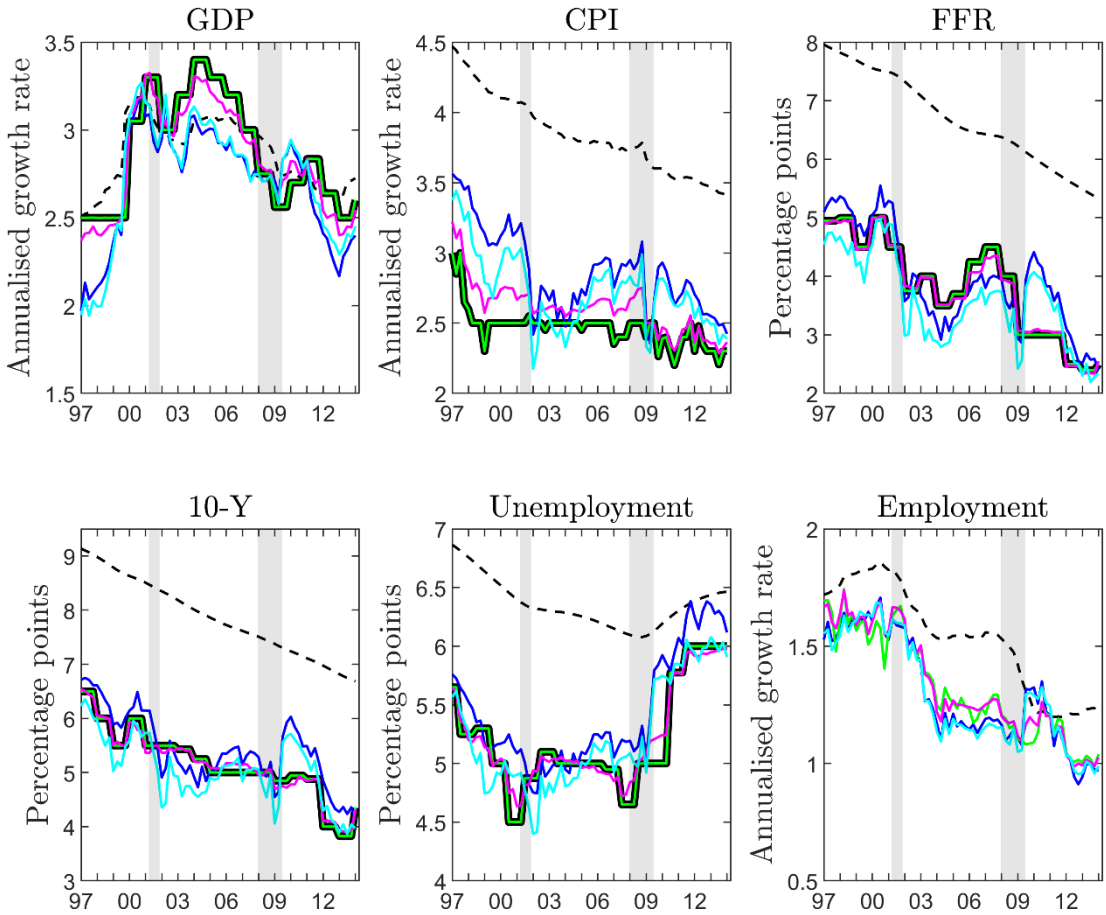


Note. Red solid lines show the median of Minnesota prior implied unconditional mean of the VAR process estimated from Gibbs sampler draws with equation (1.4). Red dash lines show the 68% credible intervals. Black dash lines show the time series unconditional mean recursively estimated since 1980Q1. Black solid lines show the median SPF long-run expectations (10-year average), and the green area covers the 95% confidence interval after having excluded the lowest and highest 5 percentiles from the raw survey data.

Figure 1.9 below illustrates the sensitivity of the steady-state posterior with respect to its prior tightness. For all the variables for which SPF long-run expectations exist, a sensitivity analysis for the three following cases is performed: *i*) a very tight prior with variance equal to 0.0001, *ii*) the conventional value of 0.05 as in Wright (2013), and *iii*) a diffuse prior with large variance for the steady-state set equal to 100. A visual inspection of the figure reveals that the

tighter the steady-state prior is, the more it moves closer to the SPF long-run expectations, as expected. For a steady-state prior's variance of 10^{-4}, the posterior (green) coincides perfectly with the prior (black), whereas for the diffuse value of 100, the posterior of the steady-state (cyan) almost coincides with the Minnesota prior implied unconditional mean (blue) which is estimated as discussed previously. In the case of a steady-state prior variance value of 0.05 (purple), the posterior is a mixture of the prior and the data, although it can be seen to be more closely influenced by the prior itself.

Figure 1.9: Steady-state posterior sensitivity

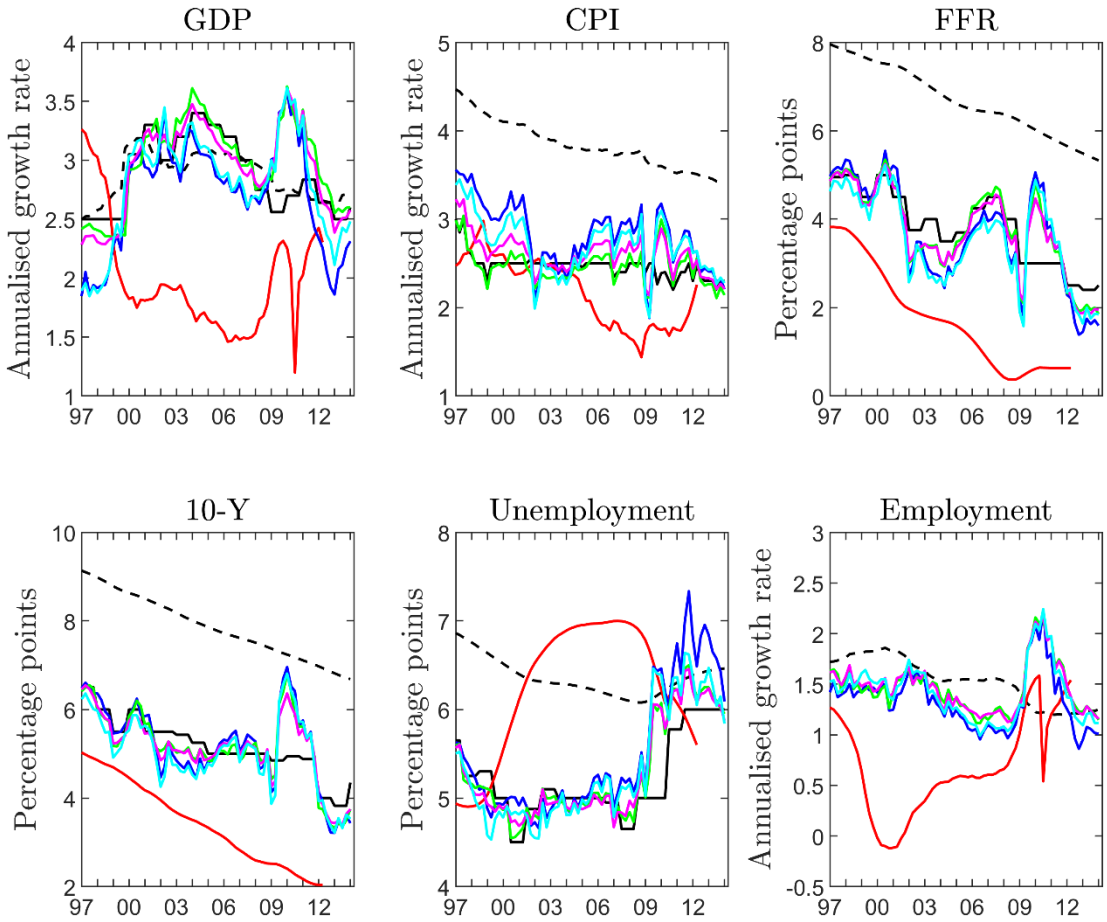


Note. Green, purple and cyan lines show the median of the posterior for SS, where the SS prior variance for the variables that survey data exist is set equal to 0.0001, 0.05, and 100 (diffuse), respectively. Blue lines show the median of the Minnesota prior implied unconditional mean. Black dash lines show the time series unconditional mean recursively estimated since 1980Q1. Black solid lines show the median SPF long-run expectations (10-year average). Red lines show the 10-year ahead average of the realisations, which stop in 2008 due to data availability.

In a similar manner, I proceed further in an out-of-sample evaluation and check the sensitivity of the steady-state prior's forecasts with respect to the tightness of the prior. The results are shown in Figure 1.10. First of all, for the case of the diffuse steady-state prior, again, as expected, the corresponding results (cyan lines) almost coincide with the Minnesota prior (blue lines), which is diffuse on the constant and consequently on the VAR process unconditional-mean. Secondly, for the tightest case (green lines), the 10-year ahead forecasts average is more

close to the SPF long-run expectations compared to the less tight and diffuse case (purple and cyan lines, respectively). However, this fact is more profound for the fast mean-reverting variables and far less evident for the persistent variables.

Figure 1.10: BVAR long-run forecasts



Note. Green, purple and cyan lines show the 10-year ahead forecasts average for the steady-state prior, whose variance for the variables that survey data exist is set equal to 0.0001, 0.05 and 100 (diffuse), respectively. Blue lines show the 10-year ahead forecasts average for Minnesota prior. Black dash lines show the time series unconditional mean recursively estimated since 1980Q1. Black solid lines show the median SPF long-run expectations (10-year average). Red lines show the 10-year ahead average of the realisations, which stop in 2012 due to data availability.

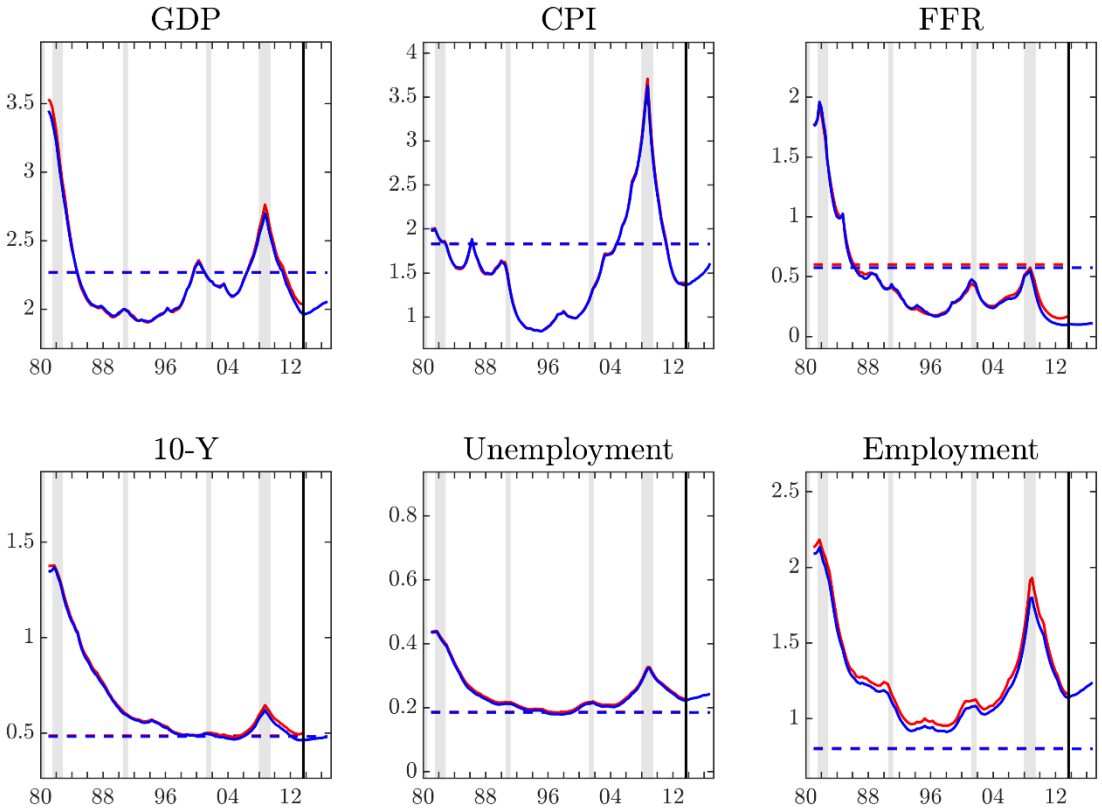
1.6.1.3 Stochastic volatility estimates

This section deals briefly with the estimates of stochastic volatility. In order to have a better understanding of how the volatility varies over time, the following figure shows the median of the standard deviation of the stochastic volatility for each variable, i.e. $(\Sigma_{n,n,t})^{1/2}$ as estimated for the last out-of-sample forecasting exercise, i.e. vintage of 2014Q1. In more detail, the solid red line shows the stochastic volatility in the conventional setting where the model is estimated first on the available data, and subsequently, the conditional forecasts are made (2-step

estimation), while the solid blue line shows the estimates of the stochastic volatility that account for parameters uncertainty by appending the conditional forecasts to the original data sample and estimating the parameters as discussed in Section 1.4 in more detail (joint estimation).

In addition, in order to gauge how the stochastic volatility behaves versus the alternative of the homoscedastic steady-state prior BVAR, the dash lines show the estimates of the standard deviations of the VAR's innovations for each variable, i.e. $(\Sigma_{n,n})^{1/2}$ again following the two approaches discussed previously. Overall, it is visible that the estimates both for heteroscedastic and homoscedastic BVARs with regard to the two approaches in conditional forecasting do not exhibit any material differences and almost coincide. Finally, examining in more detail how stochastic volatility changes over time, it is obvious that the period before the start of the great moderation (circa 1985) is characterised by excessive volatility, as expected. The 2008 financial crisis has resulted in big spikes, especially for CPI and employment, which are commensurate with the data.

Figure 1.11: Stochastic volatility estimates



Note. Red solid and dash lines show the standard deviation of the VAR's innovations for the heteroscedastic (2-step estimation) and homoscedastic specifications, respectively. Blue solid and dash lines show the standard deviation of the VAR's innovations for the heteroscedastic (joint estimation) and homoscedastic specification, respectively. Forecasting exercise 2014Q1: estimation sample 1980-2013, conditional forecasts horizon 2014-2017.

1.6.2 Out-of-sample results – Forecast evaluation metrics

Point forecasts for variable *n* and horizon *h* are evaluated according to the root mean squared error (RMSE) :

$$\mathrm{RMSE}_{n,h} = \left[\frac{1}{T-H-t_0-1} \sum_{t=t_0}^{T-H} \left(Y_{n,t+h} - \hat{Y}_{n,t+h|t} \right)^2 \right]^{1/2}$$

Where t_0 refers to 1996Q4, T refers to 2016Q4, and $H = 12$.

Density forecasts are evaluated according to the continuous ranked probability score (CRPS). CRPS is defined as:

$$\mathrm{CRPS}(F, y) = -\int_{-\infty}^{+\infty} \left(F(x) - \mathbb{I}_{x \geq y}(x) \right)^2 dx$$

Where F is the cumulative density function of the predictive density, y is the outcome, and $\mathbb{I}_{x \geq y}(x)$ is the indicator function, i.e. $\mathbb{I}_{x \geq y}(x) = 1$ if $x \geq y$ and 0 otherwise. CRPS imposes a positive penalty the further away a point of the predictive density is from the outcome, and thus a lower CPRS value implies a better forecasting performance. Following, Gneiting and Raftery (2007), the CPRS can be evaluated in a closed form:

$$\mathrm{CRPS}(F, y) = \frac{1}{2} E_F \left| \hat{y} - \hat{y}' \right| - E_F \left| y - \hat{y} \right|$$

Where $\hat{y}$ are independent draws from the predictive density with cumulative distribution F and y is the outcome.

The exact formula[25] for the retained draws R of the MCMC algorithm for variable n and horizon h is given as:

$$\mathrm{CRPS}_{n,h} = \frac{1}{T-H-t_0-1} \sum_{t=t_0}^{T-H} \left[\frac{1}{R} \sum_{r=1}^{R} \left| \hat{Y}_{n,t+h|t}^{(r)} - Y_{n,t+h} \right| - \frac{1}{2R^2} \sum_{r=1}^{R} \sum_{r'=1}^{R} \left| \hat{Y}_{n,t+h|t}^{(r)} - \hat{Y}_{n,t+h|t}^{(r')} \right| \right]$$

Furthermore, density forecasts are also evaluated using the log predictive score (LPS), and the results are available in Appendix A.2 Additional results – Log predictive score.

In order to provide a rough measure of whether the differences in the forecast accuracy are significant, Diebold and Mariano (1995) test is employed with Harvey, Leybourne, and Newbold (1997) small sample adjustment[26]. The hypothesis testing is always one-sided, i.e. rejection of the null of equal performance versus the benchmark provides evidence of the proposed model outperforming the benchmark model. The test applies in the same manner both for point forecasts (MSE) and density forecasts[27] (CRPS, LPS).

[25] More information can be found in Dieppe, Legrand, and Van Roye (2016).

[26] For the (few) cases were the estimated variance is negative the standard Diebold and Mariano (1995) with Bartlett kernel is applied

[27] The results of Diebold-Mariano tests for density forecasts should be interpreted with care since this type of statistical testing applies to rolling estimation schemes and not to recursive re-estimation of the models under an expanding window (Amisano & Giacomini, 2007).

In what follows, benchmark refers to the benchmark Minnesota prior used either for unconditional or conditional forecasting in any given specification. The only exception to this benchmark definition is when using the Diebold-Mariano test to compare the relative forecasting performance of the conditional versus the unconditional forecasts for the steady-state BVARs in Tables 3 and 5.

1.6.2.1 Point forecasts

Regarding the point forecasts, initially, the relative forecasting performance between the unconditional and the conditional forecasting performance for each specification of the steady-state prior, i.e. homoscedastic and heteroscedastic, is evaluated. In addition, the two conditional forecasting methods, i.e. 2-step estimation and joint estimation, are compared against the unconditional forecasts. The results are presented in the following table.

Overall, in the majority of the cases, the conditional forecasts outperform the unconditional forecasts as expected, with RMSE ratios below unity. For 10-year bond yield and GDP, with few exceptions, the results are found to be consistently significant across all specifications. Moreover, focusing on the impact of accounting for parameter uncertainty and employing joint estimation of the conditional forecasts further improves the results for the majority of the cases (although employment results deteriorate), albeit marginally.

Table 1.3: Steady-state prior conditional versus unconditional forecasts evaluated with RMSEs

	GDP	CPI	FFR	10-Y	Unemployment	Employment
			SS unconditional forecasts			
$h = 1$	2.480	2.170	0.443	0.454	0.271	1.099
$h = 4$	2.857	2.123	1.474	1.353	1.001	1.997
$h = 8$	2.883	2.205	2.522	2.061	1.806	2.197
$h = 12$	2.676	2.234	3.148	2.244	2.170	1.979
			SS conditional forecasts 2-step estimation			
$h = 1$	0.949**	0.978	-	0.903**	1.016	0.942*
$h = 4$	0.879***	0.993	-	0.778*	0.899*	0.809**
$h = 8$	0.894*	0.975	-	0.583*	0.842*	0.829**
$h = 12$	1.036	0.945**	-	0.456*	0.895*	1.014
			SS conditional forecasts joint estimation			
$h = 1$	0.940**	0.985	-	0.894**	1.000	0.929**
$h = 4$	0.867***	0.993	-	0.770*	0.883*	0.795**
$h = 8$	0.892*	0.966	-	0.599*	0.825*	0.819**
$h = 12$	1.025	0.939**	-	0.476*	0.874*	0.992
			SS-SV unconditional forecasts			
$h = 1$	2.559	2.150	0.375	0.443	0.283	1.113
$h = 4$	2.809	2.117	1.363	1.190	1.020	1.985
$h = 8$	2.772	2.187	2.489	1.809	1.827	2.182
$h = 12$	2.607	2.248	3.223	2.067	2.185	1.948
			SS-SV conditional forecasts 2-step estimation			
$h = 1$	1.077	1.060	-	0.926*	1.093	1.134
$h = 4$	0.839**	1.004	-	0.779**	0.982	0.820
$h = 8$	0.847**	0.975	-	0.604*	0.821*	0.762*
$h = 12$	0.925***	0.947	-	0.518*	0.792*	0.855
			SS-SV conditional forecasts joint estimation			
$h = 1$	1.017	1.039	-	0.900**	1.092	1.055
$h = 4$	0.842**	0.999	-	0.724**	0.977	0.848
$h = 8$	0.843**	0.954	-	0.561*	0.808*	0.809
$h = 12$	0.919***	0.939**	-	0.486*	0.778*	0.885

Note. Panels with unconditional forecasts show the raw RMSEs. Conditional forecasts for each specification show the RMSE ratios with respect to the unconditional forecasts of the same specification. Stars refer to the p-values of the Diebold and Mariano (1995) one-sided test with finite sample adjustment of Harvey, Leybourne, and Newbold (1997) with respect to the unconditional forecasts of the same specification. *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively. 2-step and joint estimation conditional forecasts refer to parameters' uncertainty accounting for the 'in-sample' data only and the 'in-sample' data extended with the conditional forecasts respectively. The evaluation period is 1997-2016.

In the following table, the impact of the steady-state prior is evaluated. For the cases of homoscedastic and heteroscedastic specifications, firstly, the relative performance versus the benchmark is evaluated for unconditional forecasts and then for conditional forecasts, as discussed previously, by conditioning the forecasts on the actual path of the FFR.

Overall, for the 10-year bond yield, the results are consistent across all specifications in favour of the steady-state prior both for unconditional and conditional forecasts. For the remaining variables, the steady-state prior's unconditional forecasts of the FFR and the CPI outperform the benchmark, whereas, for the conditional forecasts, this pattern is reversed. On the contrary, for the GDP and unemployment, whereas for unconditional forecasts, the steady-state prior is outperformed by the benchmark for the conditional forecasts, the opposite holds, especially with the inclusion of stochastic volatility. For the case of unconditional forecasts of unemployment, Wright (2013) and Tallman and Zaman (2019) have concluded similar results,

i.e. the inclusion of survey-based long-term expectations does not improve forecast accuracy, and it is rather interesting that the steady-state prior combined with stochastic volatility outperforms the benchmark BVAR.

Table 1.4: Steady-state versus Minnesota prior for conditional and unconditional forecasts evaluated with RMSEs

	GDP	CPI	FFR	10-Y	Unemployment	Employment
			SS unconditional forecasts			
$h = 1$	1.012	0.967*	0.984	0.988	1.005	1.004
$h = 4$	1.048	0.972*	0.961	0.994	1.032	1.058
$h = 8$	1.031	0.973**	1.010	0.992	1.057	1.091
$h = 12$	1.023	0.981**	1.037	0.984	1.063	1.056
			SS conditional forecasts 2-step estimation			
$h = 1$	0.963**	0.946*	-	0.894**	0.997	1.002
$h = 4$	0.999	0.998	-	0.890*	0.967	1.062
$h = 8$	1.014	1.016	-	0.837	0.994	1.164
$h = 12$	1.053	1.015	-	0.823	1.027	1.155
			SS conditional forecasts joint estimation			
$h = 1$	0.954**	0.952*	-	0.884**	0.982	0.988
$h = 4$	0.984	0.999	-	0.880*	0.949*	1.043
$h = 8$	1.011	1.007	-	0.860	0.974	1.151
$h = 12$	1.041	1.009	-	0.858	1.004	1.130
			SS-SV unconditional forecasts			
$h = 1$	1.044	0.958*	0.832***	0.964	1.051	1.017
$h = 4$	1.031	0.969	0.889***	0.874*	1.052	1.052
$h = 8$	0.991	0.965**	0.997	0.871	1.069	1.083
$h = 12$	0.997	0.987	1.062	0.907	1.070	1.040
			SS-SV conditional forecasts 2-step estimation			
$h = 1$	1.127	1.015	-	0.894**	1.123	1.222
$h = 4$	0.937**	1.007	-	0.784*	1.076	1.069
$h = 8$	0.923*	1.008	-	0.760	0.980	1.063
$h = 12$	0.916*	1.025	-	0.861*	0.915**	0.959
			SS-SV conditional forecasts joint estimation			
$h = 1$	1.065	0.996	-	0.869**	1.121	1.136
$h = 4$	0.941**	1.002	-	0.729*	1.071	1.105
$h = 8$	0.919*	0.986	-	0.706	0.965*	1.128
$h = 12$	0.910*	1.015	-	0.807	0.899**	0.992

Note. Panels show the RMSE ratios of SS over the Minnesota prior (benchmark) for the same specification. Stars refer to the p-values of the Diebold and Mariano (1995) one-sided test with finite sample adjustment of Harvey, Leybourne, and Newbold (1997). *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively. 2-step and joint estimation conditional forecasts refer to parameters' uncertainty accounting for the 'in-sample' data only and the 'in-sample' data extended with the conditional forecasts respectively. The evaluation period is 1997-2016.

1.6.2.2 Density forecasts

Focusing on the density forecasts evaluated with the CRPS, it is obvious that the conditional forecasts are significantly better compared to their unconditional equivalents for all variables except for the CPI. Furthermore, in the majority of the cases, the joint-estimation approach yields more accurate results, especially for forecast horizons extending beyond the one-year, albeit this is often not reflected in significance.

Table 1.5: Steady-state prior conditional versus unconditional forecasts evaluated with CRPS

	GDP	CPI	FFR	10-Y	Unemployment	Employment
			SS unconditional forecasts			
$h = 1$	1.364	1.082	0.253	0.253	0.137	0.537
$h = 4$	1.587	1.073	0.833	0.734	0.493	1.073
$h = 8$	1.591	1.132	1.481	1.079	0.988	1.219
$h = 12$	1.436	1.147	1.923	1.237	1.256	1.025
			SS conditional forecasts 2-step estimation			
$h = 1$	0.945**	0.986	-	0.910**	0.976	0.920**
$h = 4$	0.868***	1.018	-	0.803*	0.843**	0.779***
$h = 8$	0.883**	1.008	-	0.592**	0.818**	0.808*
$h = 12$	1.034	0.976	-	0.444*	0.885*	1.058
			SS conditional forecasts joint estimation			
$h = 1$	0.936**	0.987	-	0.901**	0.966	0.902***
$h = 4$	0.856***	1.016	-	0.793*	0.820**	0.763***
$h = 8$	0.882*	0.999	-	0.604**	0.797*	0.798*
$h = 12$	1.019	0.967	-	0.451*	0.858*	1.029
			SS-SV unconditional forecasts			
$h = 1$	1.412	1.066	0.192	0.249	0.140	0.581
$h = 4$	1.567	1.105	0.775	0.631	0.509	1.062
$h = 8$	1.543	1.153	1.421	0.921	0.982	1.226
$h = 12$	1.456	1.189	1.847	1.082	1.253	1.133
			SS-SV conditional forecasts 2-step estimation			
$h = 1$	1.067	1.051	-	0.942*	1.116	1.107
$h = 4$	0.853**	0.969	-	0.807**	0.941	0.861
$h = 8$	0.851***	0.975	-	0.618**	0.795**	0.774*
$h = 12$	0.914***	0.957	-	0.519**	0.770*	0.838**
			SS-SV conditional forecasts joint estimation			
$h = 1$	1.021	1.017	-	0.912***	1.108	1.022
$h = 4$	0.851**	0.969	-	0.764**	0.936*	0.876
$h = 8$	0.839**	0.962	-	0.591**	0.779**	0.815*
$h = 12$	0.900***	0.948	-	0.501**	0.746**	0.839**

Note. Panels with unconditional forecasts show the raw CRPS. Conditional forecasts for each specification show the CRPS ratios with respect to the unconditional forecasts of the same specification. Stars refer to the p-values of the Diebold and Mariano (1995) one-sided test with finite sample adjustment of Harvey, Leybourne, and Newbold (1997) with respect to the unconditional forecasts of the same specification. *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively. 2-step and joint estimation conditional forecasts refer to parameters' uncertainty accounting for the 'in-sample' data only and the 'in-sample' data extended with the conditional forecasts respectively. The evaluation period is 1997-2016.

Regarding the comparison of unconditional and conditional density forecasts of the steady-state prior versus the benchmark, the results are presented in the following table. For the GDP, the unconditional forecasts are in favour of the benchmark; however, the inclusion of stochastic volatility further improves the steady-state prior's performance. In the case of the CPI, results are mixed, and only for the homoscedastic VAR unconditional forecasts, steady-state prior consistently outperforms the benchmark across all horizons.

For the FFR, the inclusion of stochastic volatility helps to improve the steady-state prior's unconditional forecast performance dramatically, while for 10-year bond yield, except for the homoscedastic VAR unconditional forecasts, results are in favour of the steady-state. Finally, for unemployment, while for unconditional density forecasts, the accuracy is almost on par, for conditional forecasts, results are in favour of the steady-state and further improve, allowing for heteroscedasticity.

Table 1.6: Steady-state versus Minnesota prior for conditional and unconditional forecasts evaluated with CRPS

	GDP	**CPI**	**FFR**	**10-Y**	**Unemployment**	**Employment**
			SS unconditional forecasts			
$h = 1$	1.010	0.968	1.004	0.989	0.985	1.001
$h = 4$	1.049	0.974	0.964	0.998	1.013	1.076
$h = 8$	1.031	0.969*	1.022	1.004	1.037	1.102
$h = 12$	1.020	0.983	1.062	0.996	1.040	1.050
			SS conditional forecasts 2-step estimation			
$h = 1$	0.961**	0.955	-	0.895**	0.946	0.991
$h = 4$	0.990	1.016	-	0.914	0.921*	1.046
$h = 8$	1.004	1.044	-	0.853*	0.968	1.146
$h = 12$	1.046	1.051	-	0.835*	1.003	1.172
			SS conditional forecasts joint estimation			
$h = 1$	0.952**	0.956*	-	0.886**	0.937*	0.971
$h = 4$	0.976	1.013	-	0.903	0.896**	1.024
$h = 8$	1.002	1.034	-	0.871*	0.943*	1.132
$h = 12$	1.031	1.042	-	0.850*	0.973	1.140
			SS-SV unconditional forecasts			
$h = 1$	1.045	0.955*	0.759***	0.972	1.003	1.083
$h = 4$	1.035	1.004	0.897***	0.858*	1.045	1.066
$h = 8$	0.999	0.987	0.981	0.856	1.030	1.109
$h = 12$	1.034	1.019	1.019	0.872	1.037	1.161
			SS-SV conditional forecasts 2-step estimation			
$h = 1$	1.124	1.003	-	0.911**	1.102	1.291
$h = 4$	0.960	0.995	-	0.790*	1.061	1.145
$h = 8$	0.938*	1.029	-	0.760*	0.935**	1.104
$h = 12$	0.938	1.068	-	0.855	0.871***	1.027
			SS-SV conditional forecasts joint estimation			
$h = 1$	1.075	0.970	-	0.882***	1.094	1.192
$h = 4$	0.958	0.996	-	0.748**	1.056	1.165
$h = 8$	0.925*	1.015	-	0.726*	0.916***	1.163
$h = 12$	0.922	1.058	-	0.825	0.844***	1.028

Note. Panels show the CRPS ratios of SS over the benchmark (Minnesota prior). Stars refer to the p-values of the Diebold and Mariano (1995) one-sided test with finite sample adjustment of Harvey, Leybourne, and Newbold (1997). *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively. 2-step and joint estimation conditional forecasts refer to parameters' uncertainty accounting for the 'in-sample' data only and the 'in-sample' data extended with the conditional forecasts respectively. The evaluation period is 1997-2016.

1.6.2.3 Time variation of results

This section explores whether the results presented previously over the full sample remain stable over time. Due to structural breaks in the time series, the impact of the financial crisis and the changes in the survey-based long-run expectations that feed into the models' forecasts via the steady-state prior, it is reasonable to assume that there is variation in steady-state prior's relative performance in different sub-samples with respect to the benchmark. Following the standard approach in the literature, the differences in the cumulative squared errors and the CRPS between the steady-state prior specifications and the benchmark are plotted over time.

Conditional and unconditional forecasts for GDP, which are produced using the steady-state prior, outperform the benchmark up to the financial crisis. After the financial crisis, for homoscedastic VARs, the benchmark outperforms the steady-state prior, while allowing for heteroscedasticity does help to reverse this pattern materially and in favour of the steady-state prior.

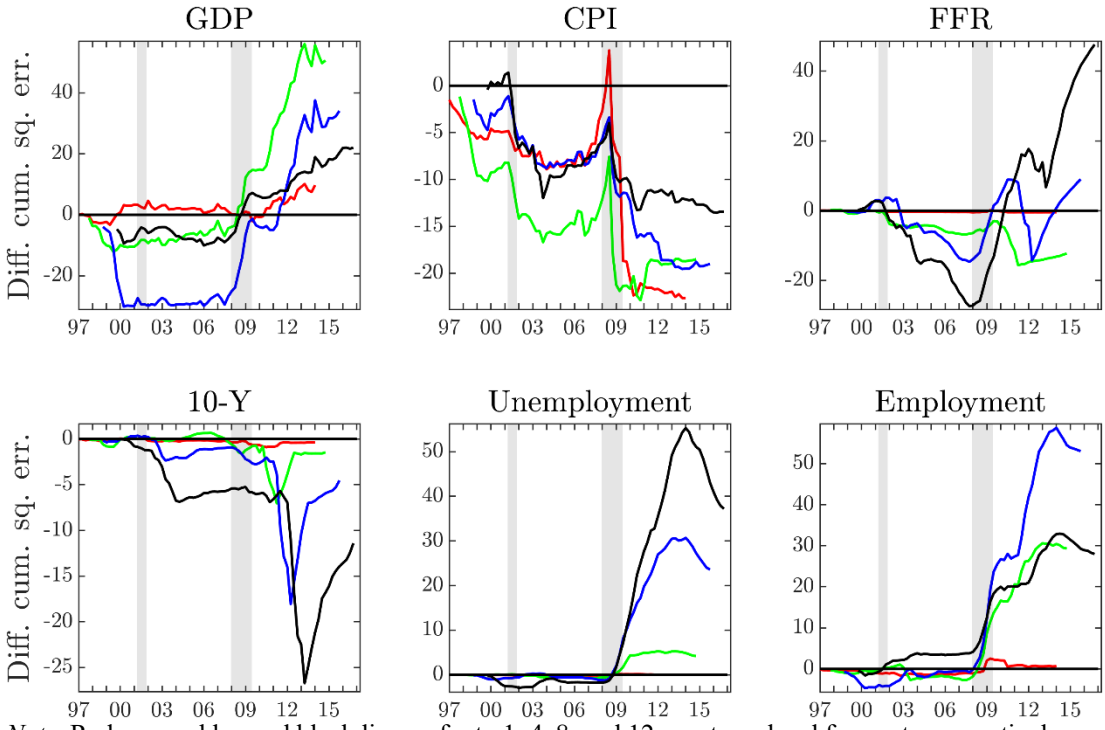
For CPI, the steady-state prior's unconditional forecasts are more accurate over the full sample. Conditional forecasts outperform the benchmark only until the financial crisis, whereas after the financial crisis, they deteriorate.

Regarding the FFR, quite surprisingly, the steady-state prior's unconditional forecasts deteriorate after the financial crisis despite the fact that long-run expectations anchor in a timely manner to a significantly lower level. However, the inclusion of stochastic volatility ameliorates this effect. A possible explanation over this period could be that under the zero lower bound, the steady-state prior behaves erratically over this period. A tighter prior would be required to capture the shift at the endpoints in the short-term interest rates.

Turning to the 10-year bond yield, the results are robust in favour of the steady-state prior for the majority of the cases as well as over the full sample both for unconditional and conditional forecasts evaluated with the point as well as the density forecasts.

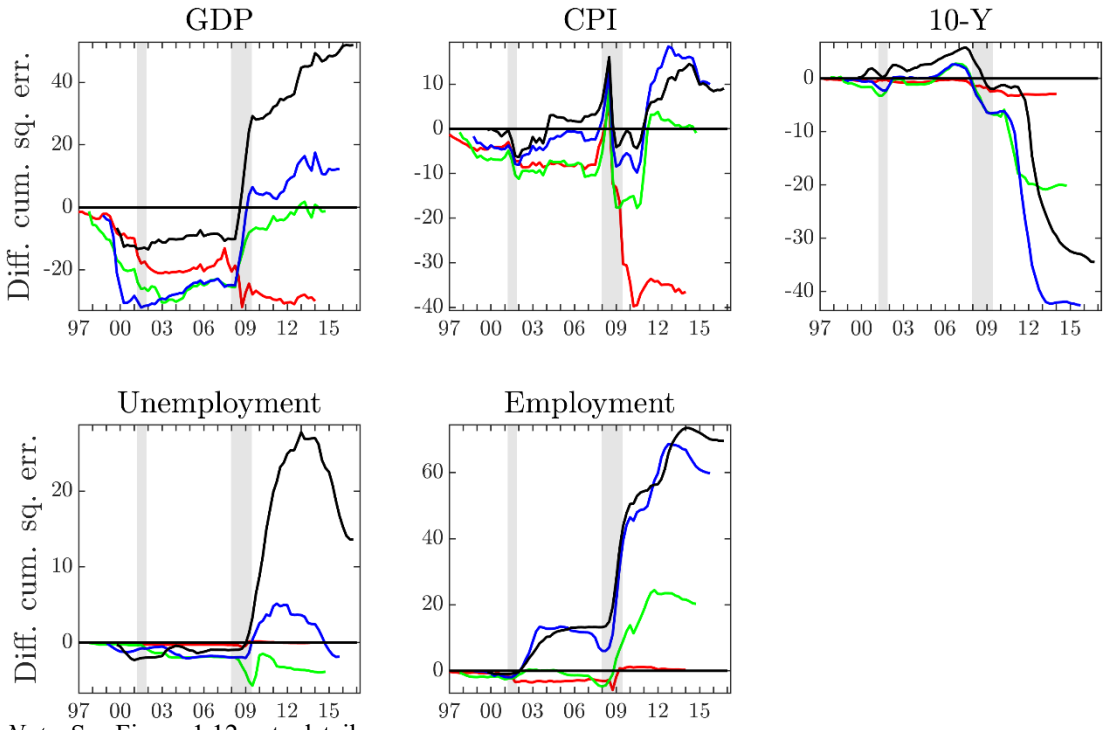
In the cases of unemployment and employment, for the unconditional forecasts, the steady-state prior and the benchmark behave roughly in a similar manner, while afterwards, results indicate the benchmark outperforms the steady-state prior, especially for point forecasts. Nonetheless, the conditional density forecasts from both the homoscedastic and the heteroscedastic BVARs improve the steady-state prior's accuracy after the financial crisis in the long-term conditional forecasts.

Figure 1.12: Difference of the cumulative squared errors of the steady-state prior minus the benchmark for unconditional forecasts



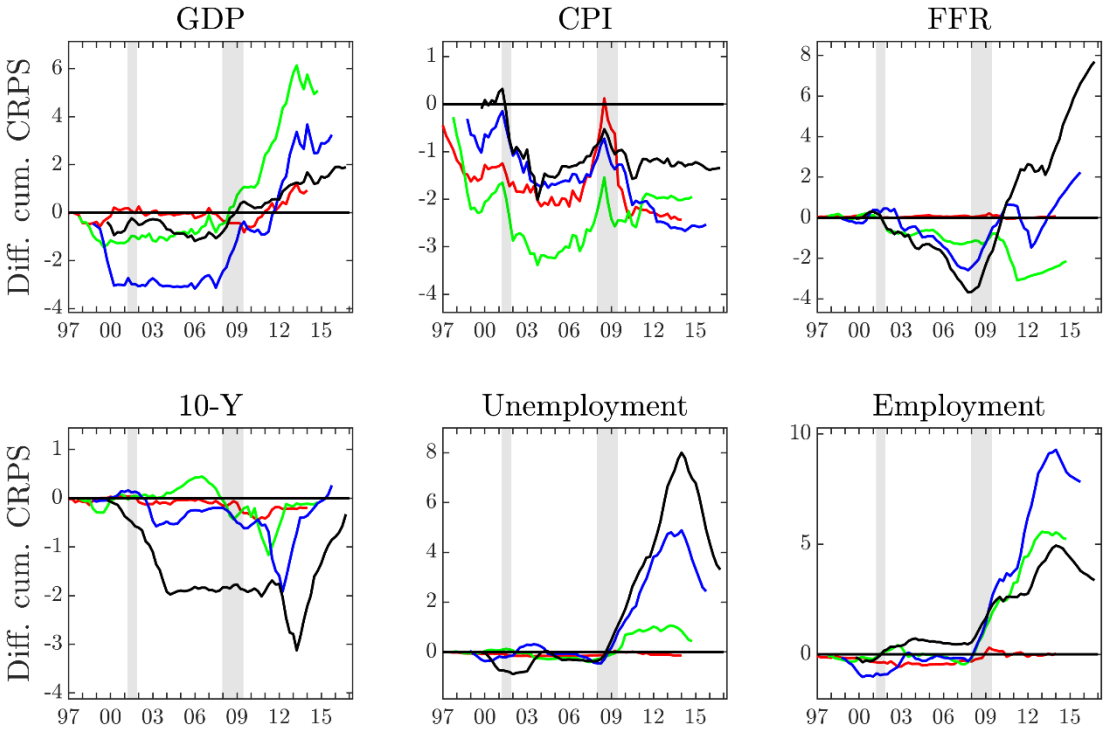
Note. Red, green, blue and black lines refer to 1, 4, 8, and 12-quarters ahead forecasts, respectively

Figure 1.13: Difference of the cumulative squared errors of the steady-state prior minus the benchmark for conditional forecasts



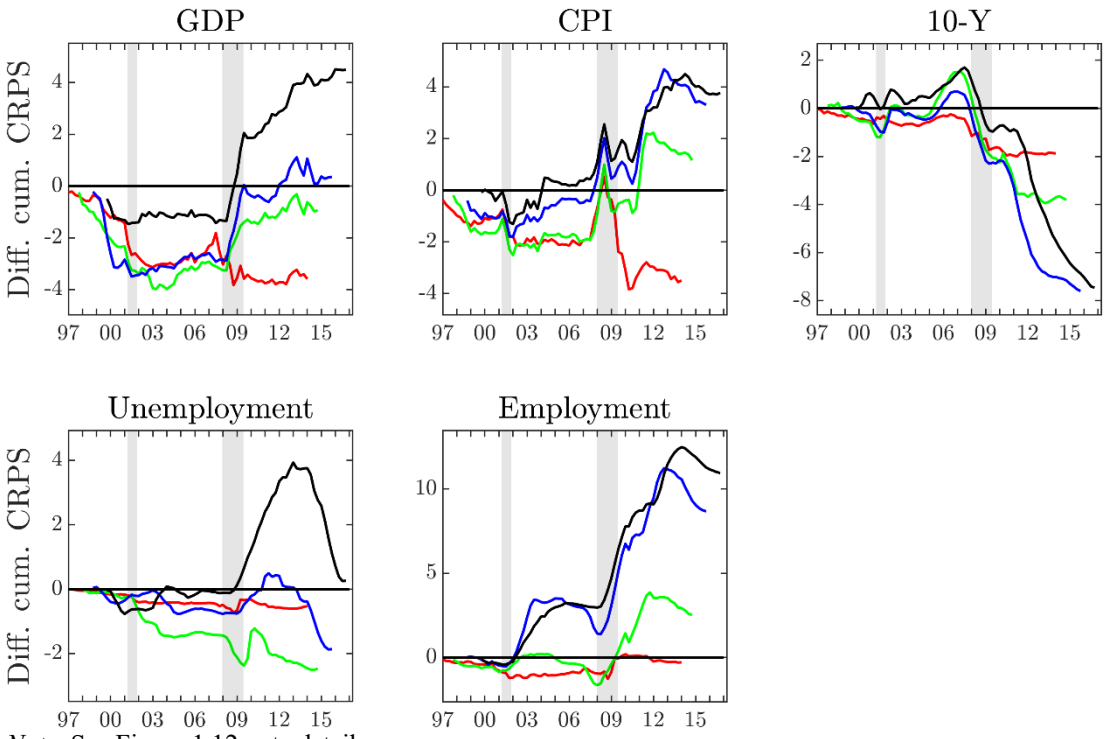
Note. See Figure 1.12 note details.

Figure 1.14: Difference of the cumulative CRPS of the steady-state prior minus the benchmark for unconditional forecasts



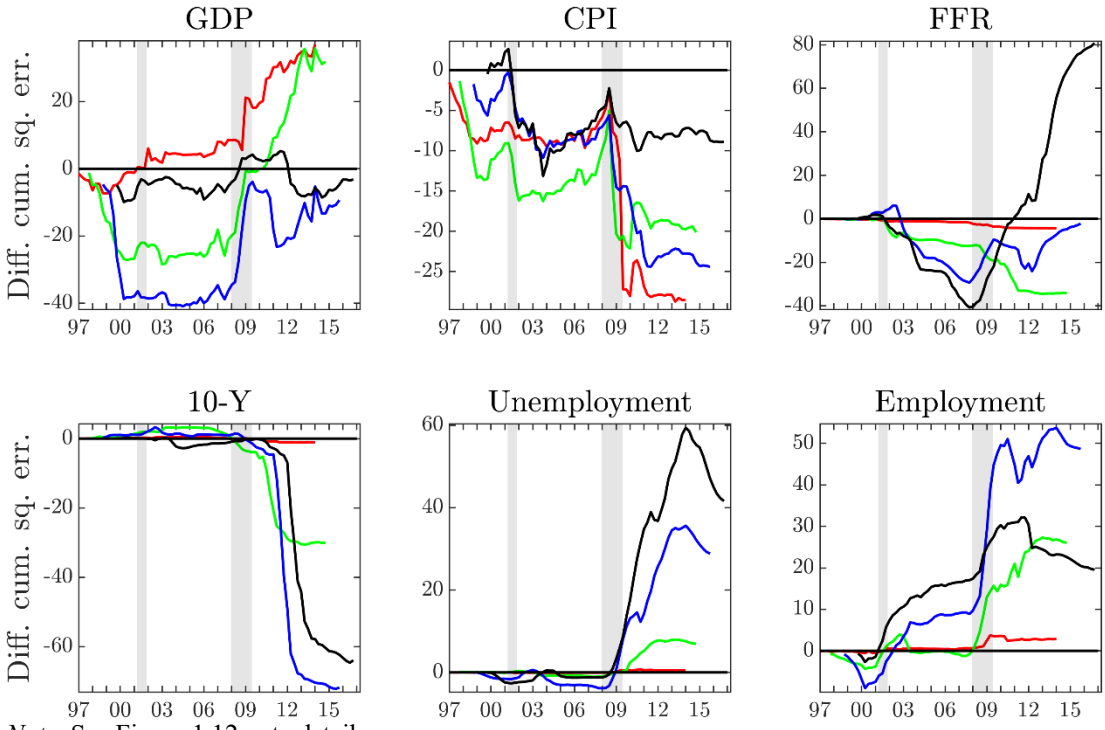
Note. See Figure 1.12 note details.

Figure 1.15: Difference of the cumulative CRPS of the steady-state prior minus the benchmark for conditional forecasts



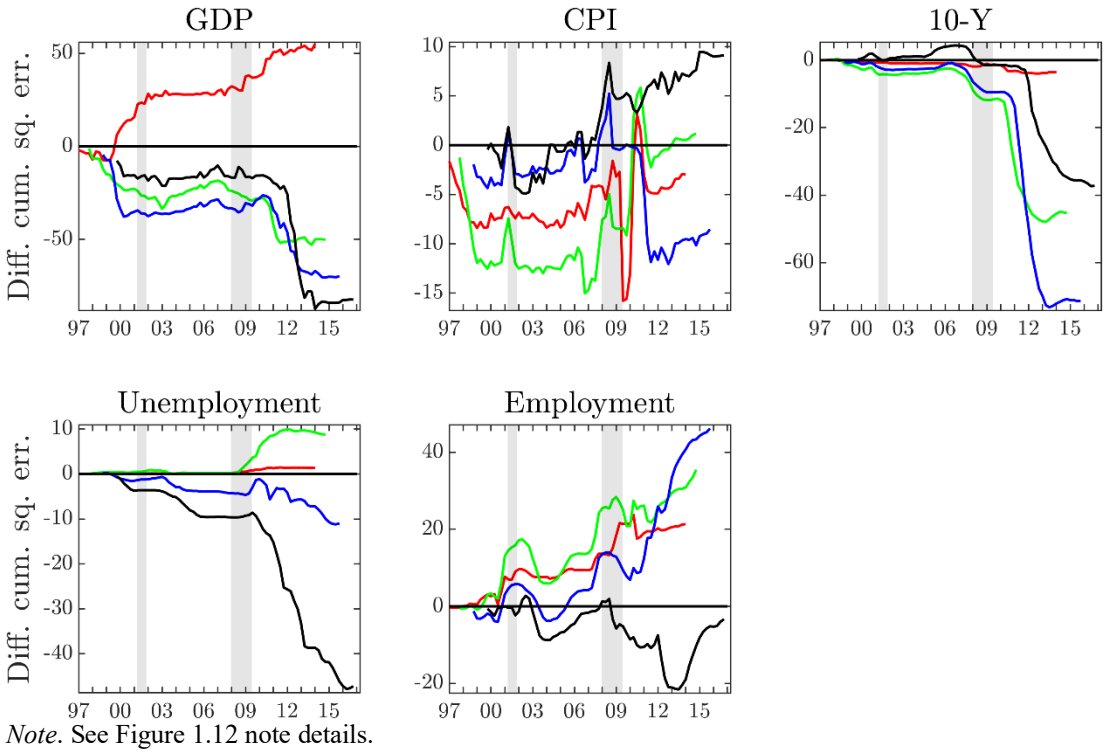
Note. See Figure 1.12 note details.

Figure 1.16: Difference of the cumulative squared errors of the SS-SV prior minus the benchmark for unconditional forecasts



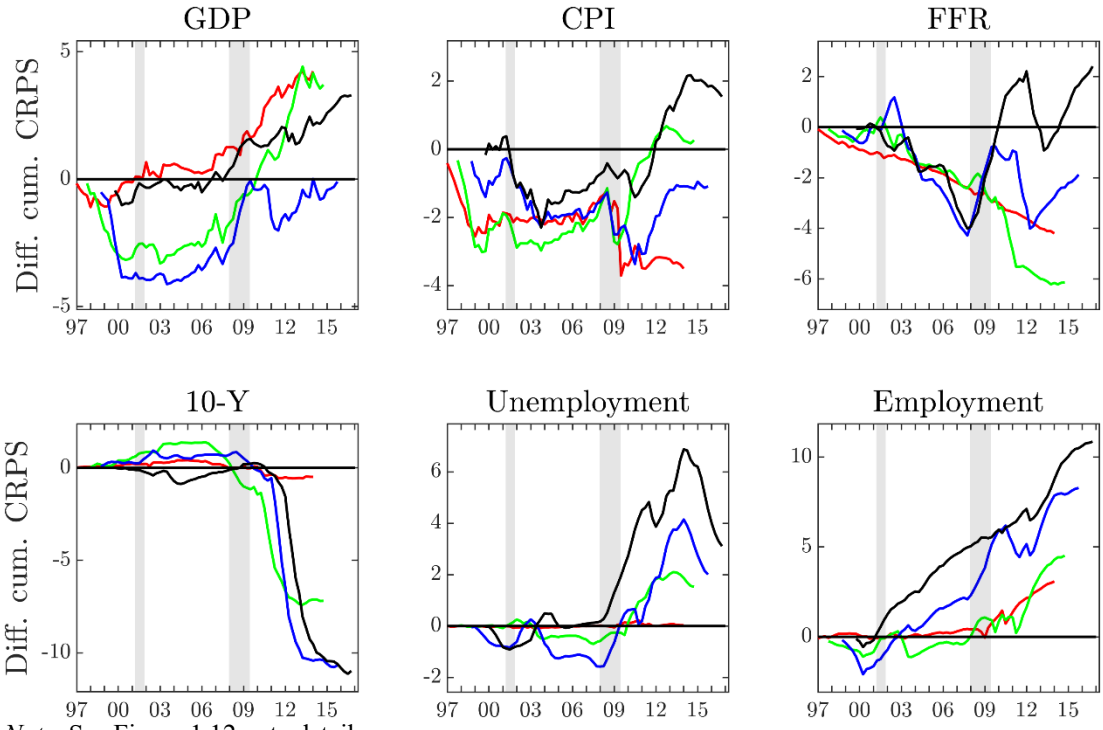
Note. See Figure 1.12 note details.

Figure 1.17: Difference of the cumulative squared errors of the SS-SV prior minus the benchmark for conditional forecasts with joint estimation



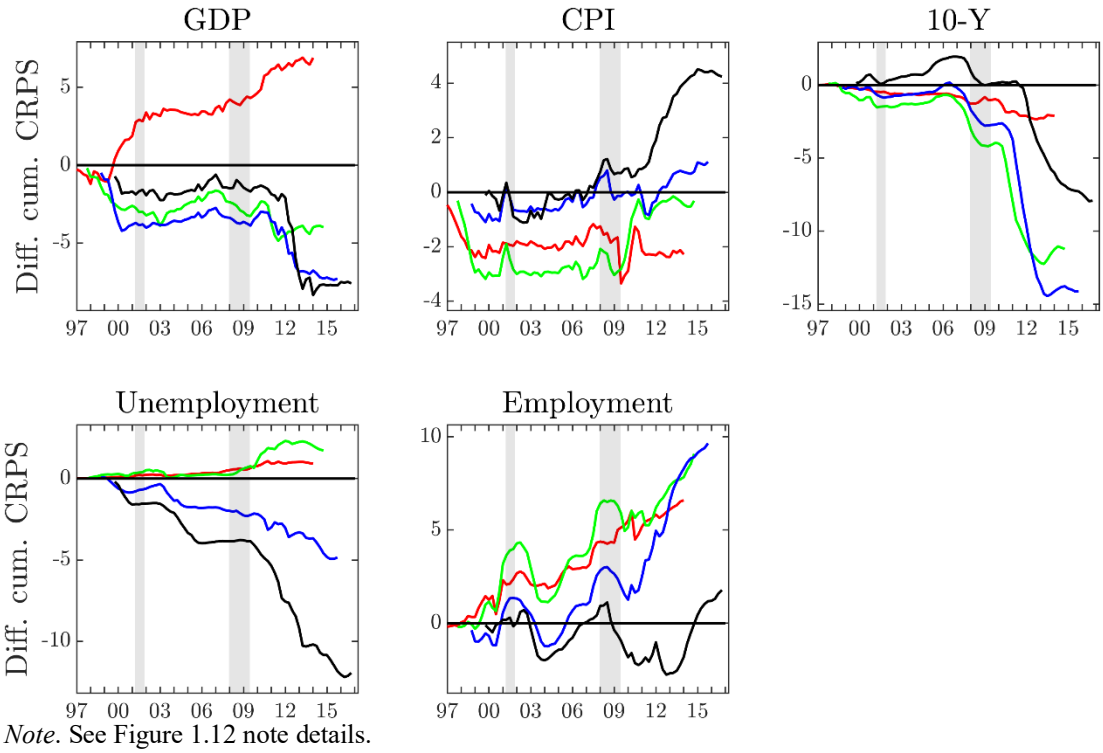
Note. See Figure 1.12 note details.

Figure 1.18: Difference of the cumulative CRPS of the SS-SV prior minus the benchmark for unconditional forecasts



Note. See Figure 1.12 note details.

Figure 1.19: Difference of the cumulative CRPS of the SS-SV prior minus the benchmark for conditional forecasts with joint estimation



Note. See Figure 1.12 note details.

1.6.3 Models' reality check

This section is akin to Stock and Watson (2012), Bańbura, Giannone, and Lenza (2015) and Giannone, Lenza, and Reichlin (2019). In particular, in their case, they first estimate the model's parameters with the sample ending before the great financial crisis and subsequently estimate the forecasts of an extended set of macroeconomic variables during and after the crisis conditional on the actual path of key macroeconomic variables such as economic activity, inflation and policy rates. I differentiate from this setting in two dimensions. Firstly, given that the steady-state prior should uses SPF long-term expectations at each vintage, I rather re-estimate the model's parameters at each vintage and estimate the conditional forecasts for a horizon up to 12-quarters ahead.

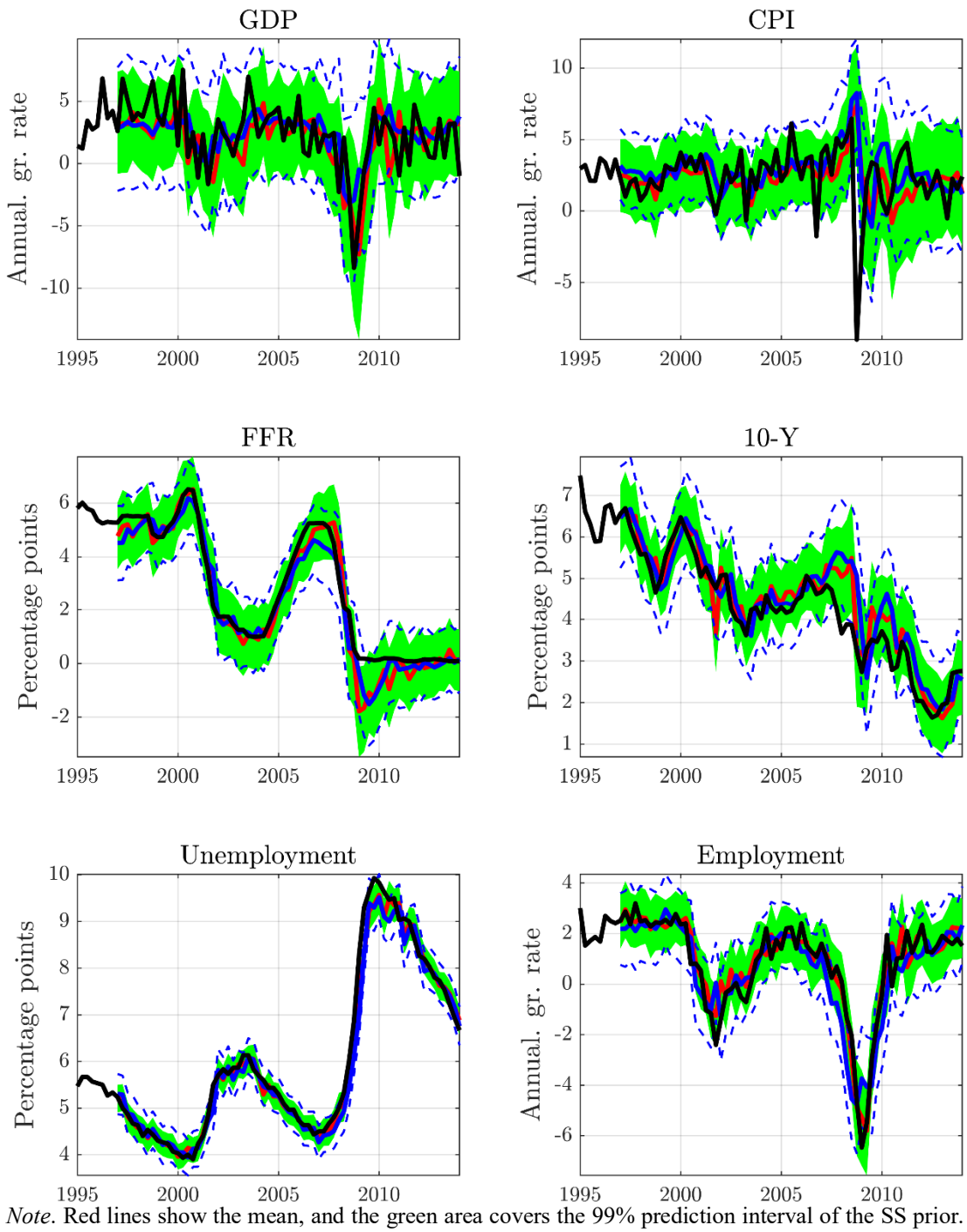
Secondly, rather than conditioning only on the actual path of a limited set of key macroeconomic variables, for each vintage, I estimate the conditional forecasts of each variable conditional on the future actual path of the remaining five variables of the VAR. For instance, the conditional forecasts of GDP are estimated assuming the actual paths of CPI, FFR, 10-year bond yield, unemployment and employment over the next three years. This exercise is performed rotationally for all variables, and subsequently, the prediction intervals are plotted against the realisations of the variables. In this manner, it is relatively easy to judge whether the models would be able to capture the statistical properties and behaviour of a variable under the assumption of a perfect foresight of the remaining variables, especially during the 2008 financial crisis excessive volatility.

The following two figures show the 1-quarter ahead and the 12-quarters ahead 99% prediction intervals of the conditional forecasts, respectively, constructed as described previously. In order to compare the performance of the steady-state prior (median forecast shown with the red line, the green area covers the 99% prediction interval), the respective results of the benchmark VAR are further shown (median forecast shown with the blue solid line, blue dash lines cover the 99% prediction interval).

Regarding the 1-quarter ahead conditional forecasts, overall, it is clear that both models' point forecasts track pretty closely the actual realisations of all variables, whereas the prediction intervals of the benchmark model are slightly wider for the majority of the variables. The only exception is that of CPI during the financial crisis, where there is a failure of the conditional forecasts to adapt abruptly to the evolving economic environment in order to fully reflect reality, and not even the 99% prediction intervals contain the realisation. This should not be interpreted as a complete failure of the specific models but rather as a misspecification where the inclusion of variables at a higher frequency, further assisted by nowcasts of the financial conditions, would have allowed the mixed-frequency VARs to update their conditional forecasts more quickly and in a greater magnitude.

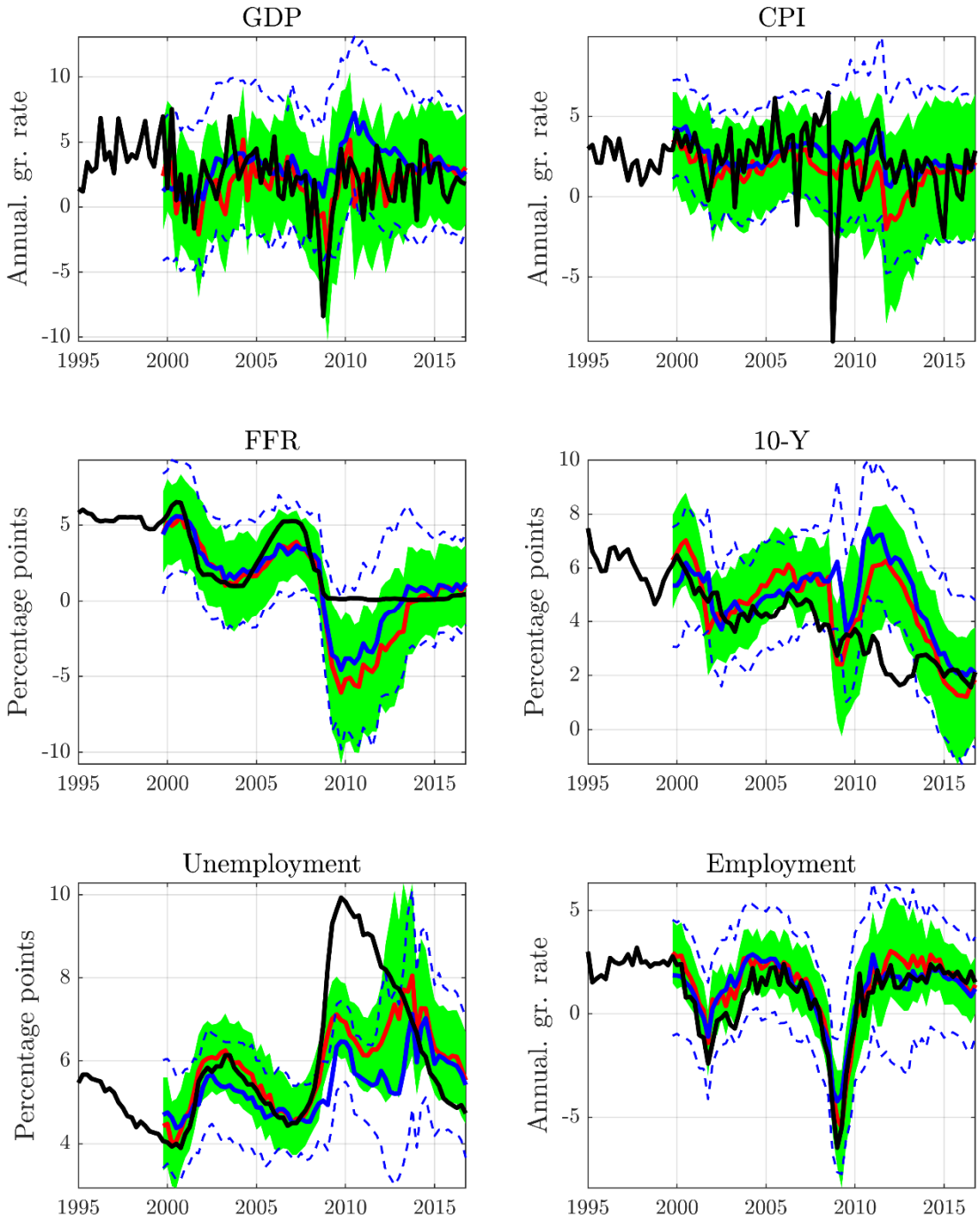
Turning to the 12-quarters ahead conditional forecasts in which the effect of the steady-state prior would be more accentuated and focusing on the point forecasts, there are indeed some material differences between the benchmark and the steady-state prior VAR. Employment is an exception for which no survey information is available, and both models behave roughly in a similar manner. In general, considerable differences are visible during the period following the financial crisis (2010-2014) and what stands out is that in the cases of GDP, 10-year bond yield, and unemployment, the steady-state prior outperforms the benchmark, while for CPI and FFR, the benchmark's point forecasts are closer to the realisation.

Figure 1.20: 1-quarter ahead conditional forecasts



Note. Red lines show the mean, and the green area covers the 99% prediction interval of the SS prior. Blue solid lines show the mean, and dash lines cover the 99% prediction interval of the Minnesota prior. Black lines show the realisation as available at the last vintage.

Figure 1.21: 12-quarters ahead conditional forecasts



Note. See Figure 1.20 note details.

1.7 Conclusion

Empirical evidence in the literature has found that the inclusion of information over the steady-state of a VAR through the use of survey-based long-run expectations improves forecast accuracy. So far, this evidence has been tested only for unconditional forecasts, while the key contribution of this essay is to assess the extent to which these findings hold in a conditional forecast setting as well.

In line with the previous results, the steady-state prior improves the accuracy of long-term yields consistently and further improves GDP and unemployment forecasts with the inclusion of stochastic volatility. However, under the models' reality check assuming that the future paths of the variables in the VAR, with the exception of one, are known ex-ante, survey expectations of unemployment improve conditional forecasts accuracy in the period following the financial crisis, whereas for the CPI and the FFR they do not.

These results could be further refined by relaxing the fixed tightness of the steady-state prior across all variables and instead allowing for a time-varying variable-dependent tightness with respect to optimising an out-of-sample forecast evaluation metric. I leave for further research and, subject to availability, the possible use of long-term expectations of the Blue Chip Economic Indicators, which are available bi-annually for a larger set of variables since 1985. Also left for future research is examining the performance of the steady-state prior for conditional forecasts conditioned not on the actual paths (realisations) of the variables but rather on real-time assumptions.

‘Remember that all models are wrong;
the practical question is how wrong do they have to be to not be useful.’

George Box & Norman Draper (1987)
Empirical Model-Building and Response Surfaces, p. 74

Essay 2

2. A hybrid time-varying logistic smooth transition autoregressive model

2.1 Introduction

Considering the linear versus non-linear division of time series, over the recent years, there has been a surge in models belonging to the latter for obvious reasons. Structural breaks, regime changes and non-homogeneous behaviour of time series during different phases of the business cycle, especially in light of the 2008 financial crisis, have motivated further research in this broad class of models.

With regard to macroeconomic and financial non-linear time series, the most prominent models[28] fall under the categories of time-varying parameter models appearing very early (Andel, 1976; Nicholls & Quinn, 1980), Markov switch regime autoregressive models (Hamilton, 1989), threshold autoregressive model (TAR) (Lim & Tong, 1980; Tong, 1990) and smooth transition autoregressive (STAR) models (Teräsvirta, 1994). The first category of time-varying parameters (TVP) models became extremely popular in macroeconomics after the introduction of TVP-VARs by Cogley and Sargent (2005) and Primiceri (2005). Empirical evidence with respect to forecasting accuracy (D'Agostino, Gambetti, & Giannone, 2013; Koop & Korobilis, 2013; Bekiros, 2014) has shown that they behave more favourably in terms of forecasting compared to their linear counterparts, especially during the financial crisis.

However, the results of the forecasting accuracy of the TAR and STAR subclass of non-linear models, amongst others, are somewhat mixed (Marcellino, 2002; Marcellino, 2004). Ferrara, Marcellino, and Mogliani (2015), in light of the 2008 financial crisis, re-examine non-linear models' accuracy in the frequentist domain under a direct forecasting approach and conclude that overall they do not present any systematic gains, although during specific evaluation periods and for specific variables where there is evidence of regime change, they do behave favourably. Barnett, Mumtaz, and Theodoridis (2014) using Bayesian techniques, conclude

[28] For a detailed literature review refer to Teräsvirta (2006) and Tsay and Chen (2018).

that overall, TVP-VAR outperforms (S)T-VAR models, ST-VARs are more accurate than T-VARs, and for the case of GDP, ST-VARs behave favourably.

This essay revisits the latter class of STAR models by partially relaxing the assumption of the constant parameters aiming to improve their performance by adopting some of the merits belonging to the literature on the TVP models. There have been some interesting publications trying to reconcile time variation in the parameters and different regimes. Building on the assumption that it is not mutually exclusive for the time series to present non-linearities and structural breaks, Telatar and Hasanov (2009) and Omay, Shahbaz, and Hasanov (2020) in the frequentist domain propose unit root tests on STAR time series allowing for time trends.

Lundbergh, Teräsvirta, and Van Dijk (2003) propose a time-varying STAR model whose regime parameters are allowed to change gradually with time in a deterministic fashion, while Galvão (2006) allows the regime-dependent parameters to be subject also to a structural break. Anderson and Low (2006) propose a STAR model with its autoregressive coefficients following a random walk process and find evidence of time variation in regime-specific parameters.

Dueker, Owyang, and Sola (2010) suggest a 3-variable ST-VAR using a time-varying threshold to model the natural rate of unemployment. Zhu and Chen (2017) and Zhu, Chen, and Lin (2019) estimate a time-varying threshold Taylor rule to model the non-linearity and asymmetry of the US policy reaction function where unemployment plays the role of the transition variable. Faria and Santos (2018) suggest a STAR model where all of its parameters are allowed to be time-varying, and its estimation is done via Taylor series expansion and Kalman filtering techniques. Finally, recently Yang, Lee, and Chen (2021) propose a TAR model with a time-varying threshold approximated with Fourier series and estimate a non-linear Taylor rule.

It has to be noted that the use of non-linear time series has gained great popularity in other fields as well. Time series of sunspot data (Koop & Potter, 2003), Canadian lynx (Haggan & Ozaki, 1981; Geweke & Terui, 1993), hydrological time series (Tong, Thanoon, & Gudmundsson, 1985), paleoclimatic data (Livingston Jr & Nur, 2019), electricity prices (Qu, Chen, Niu, & Li, 2016; Chen & Bunn, 2010) and seismic data (Khan & Mittnik, 2018) amongst others have been investigated extensively with the class of TAR and STAR models and have been found in many cases to exhibit an advantageous behaviour.

In general, the relevant literature on (S)TAR models with respect to model parameters estimation has been dominated by frequentist techniques, while Bayesian estimation has been introduced rather later. Geweke and Terui (1993) and Chen and Lee (1995) are the first to propose Bayesian estimation of TAR models, and Lopes and Salazar (2006a) deal with Bayesian STAR models where model uncertainty with respect to the number of lags is addressed by employing a reversible jump MCMC step. Notable extensions of Bayesian STAR models to multivariate time series with a focus on structural analysis rather than on forecasting accuracy include Gefang and Strachan (2010), Auerbach and Gorodnichenko (2012), investigating the impact of fiscal shocks on output under expansionary and recessionary regimes and Galvão and Owyang (2018) who propose a factor augmented ST-VAR to identify the variation of financial stress regimes' impact on the macroeconomy.

This essay will focus on the univariate setting, while multivariate time series can be investigated once the good properties of the suggested univariate models have been established.

The essay proceeds as follows. Section 2.2 reviews the basic TAR and STAR models and proposes a hybrid time-varying logistic STAR model with two variants, a homoscedastic and a heteroscedastic with a stochastic volatility specification. Section 2.3 discusses the estimation of the marginal likelihood and Deviance Information Criterion (DIC) for the existing TAR and STAR models, as well as the proposed models. Section 2.4 provides a Monte Carlo exercise of the proposed model with a known data generating process. Section 2.5 describes the empirical application with US macroeconomic data and reports the in-sample and out-of-sample results. Finally, Section 2.6 concludes.

2.2 (S)TAR models

This section builds upon the existing TAR and STAR models and, by relaxing some of their assumptions, presents the proposed models, as well as a concrete way for their estimation using Bayesian techniques. More details for the posterior estimation of the TAR and LSTAR models can be found in the section of the proposed hybrid time-varying logistic smooth autoregressive (TV-LSTAR) model, which nests the previously mentioned two models by selecting only the relevant MCMC blocks corresponding to the TAR and LSTAR models.

2.2.1 The TAR model

The generic TAR model of order K, entails K in total regimes and their corresponding K sets of autoregressive coefficients and innovations variances (Tsay & Chen, 2018). More compactly, it can be written as:

$$y_t = \beta_{k,c} + \sum_{p=1}^{P_k} \beta_{k,p} y_{t-p} + e_{1kt}, \quad e_{k,t} \sim iid\ \mathcal{N}\left(0, \sigma_k^2\right), \quad c_{k-1} \le z_t < c_K \tag{2.1}$$

The TAR model falls under the k-th regime when the transition variable z_t falls within $c_{k-1} \le z_t < c_K$, where to account for all the K regimes, their respective thresholds $\{c_i \mid i = 0,\cdots,K\}$ should be ordered as follows:

$$-\infty = c_0 < c_1 < \cdots < c_K = \infty \tag{2.2}$$

For the specific case where the d lagged value y_{t-d} of time series y_t is used as the transition variable[29] $z_t = y_{t-d}$, the TAR model[30] corresponds to the self-exciting threshold autoregressive (SETAR) model. For the purpose of this essay, the focus will be exclusively on two regime models, and thus the SETAR(P) model with P lags[31] is now written explicitly for each regime as follows:

[29] See more in Section 2.2.2 regarding the transition variable.
[30] Chen (1998) proposes a two-regime generalised TAR models where exogenous variables are added additionally to the lagged values of y_t in each regime.
[31] Campbell (2004) allows for uncertainty over the lags number P by using the reversible jump MCMC algorithm of Green (1995).

$$y_t = \begin{cases} \beta_{1,c} + \sum_{p=1}^{P} \beta_{1,p} y_{t-p} + e_{1,t}, & e_{1,t} \sim iid\ \mathcal{N}\left(0, \sigma_1^2\right), \quad y_{t-d} \le c \\ \beta_{2,c} + \sum_{p=1}^{P} \beta_{2,p} y_{t-p} + e_{2,t}, & e_{2,t} \sim iid\ \mathcal{N}\left(0, \sigma_2^2\right), \quad y_{t-d} > c \end{cases} \tag{2.3}$$

The model's estimation can be done easily along the lines of Chen and Lee (1995) by separating the sample into two subsamples, T_1 and T_2, respectively, given that the threshold c is known, and further estimating the autoregressive coefficients and the innovations variance by separate Bayesian linear regressions for each regime.

The model's likelihood is defined as:

$$\mathcal{L}\left(\beta_1, \sigma_1^2, \beta_2, \sigma_2^2, d, c \middle| Y\right) = \mathcal{L}_1\left(\beta_1, \sigma_1^2, d, c \middle| Y\right) \mathcal{L}_2\left(\beta_2, \sigma_2^2, d, c \middle| Y\right) \tag{2.4}$$

Where $\mathcal{L}_1$, $\mathcal{L}_2$ denote its respective regime's likelihood which is defined as :

$$\mathcal{L}_i = \left(2\pi\sigma_i^2\right)^{-T_i/2} \exp\left(-\frac{1}{2}\frac{\tilde{e}_i'\tilde{e}_i}{\sigma_i^2}\right), \quad i = 1,2 \tag{2.5}$$

Where $\tilde{e}_i$ is the vector of the residuals over time for each regime i = 1, 2.

With regard to the lagged value y_{t-d} that triggers the transition from one regime to the other, this is controlled by the delay parameter d. Under the assumption of d being unknown, the standard approach uses a uniform prior with discrete values between 1 and the maximum number of lags P. More details can be found in the relevant parts of Section 2.2.3.

Finally, by further relaxing the assumption that the value of threshold c is known ex-ante, Bayesian estimation requires the adoption of a prior for threshold c. However, since its conditional posterior distribution is not available in a closed-form solution, a Metropolis-Hastings step is required in order to approximate its posterior distribution. Usually, the threshold's c prior is modelled using a normal distribution, while in practice, it is more convenient to use a truncated distribution as in Barnett, Mumtaz, and Theodoridis (2014) to evaluate the posterior from the proposal distribution by excluding the lower (p_L) and higher and (p_H) percentiles, of the time series empirical distribution:

$$c \sim \mathcal{N}_{tr}\left(\underline{\mathrm{M}}_c, \underline{\Omega}_c, p_L, p_H\right) \tag{2.6}$$

For completeness, the MCMC algorithm is based on the following six blocks:

1. $p\left(\beta_1^{(r)} \middle| \left\{\sigma_1^2, \beta_2, \sigma_2^2, d, c\right\}^{(r-1)}, Y\right)$
2. $p\left(\sigma_1^{2(r)} \middle| \left\{\beta_1\right\}^{(r)}, \left\{\beta_2, \sigma_2^2, d, c\right\}^{(r-1)}, Y\right)$
3. $p\left(\beta_2^{(r)} \middle| \left\{\beta_1, \sigma_1^2\right\}^{(r)}, \left\{\sigma_2^2, d, c\right\}^{(r-1)}, Y\right)$
4. $p\left(\sigma_2^{2(r)} \middle| \left\{\beta_1, \sigma_1^2 \beta_2\right\}^{(r)}, \left\{d, c\right\}^{(r-1)}, Y\right)$
5. $p\left(d^{(r)} \middle| \left\{\beta_1, \sigma_1^2, \beta_2, \sigma_2^2\right\}^{(r)}, \left\{c\right\}^{(r-1)}, Y\right)$

6. $p\left(c^{(r)}\middle|\left\{\beta_1,\sigma_1^2,\beta_2,\sigma_2^2,d\right\}^{(r)},Y\right)$

The posterior draws for the first four blocks concerning the autoregressive coefficients and innovations variance for each of the two regimes follow standard formulas for Bayesian linear regressions.

The fifth block, concerning the posterior draws of the delay parameter *d,* follows a multinomial distribution with probabilities defined by the normalised weights using the model's likelihood. More details can be found in the relevant parts of Section 2.3.

The sixth block concerning the estimation of the threshold *c* is estimated using a random walk Metropolis step and is performed in the following way: At iteration (*r*) of the MCMC algorithm, a candidate value of c^* is drawn as:

$$c^* = c^{(r-1)} + \xi \ , \ \xi \sim iid \ \mathcal{N}\left(0,\Delta_c\right) \tag{2.7}$$

The probability of accepting c^* is estimated as:

$$\alpha = \min\left(\frac{p\left(y\middle|c^*,\theta^{(r)}\right)}{p\left(y\middle|c^{(r-1)},\theta^{(r)}\right)}\cdot\frac{p_{\mathcal{N}_{tr}}\left(c^*\middle|\underline{\mathrm{M}}_c,\underline{\Omega}_c,p_L,p_H\right)}{p_{\mathcal{N}_{tr}}\left(c^{(r-1)}\middle|\underline{\mathrm{M}}_c,\underline{\Omega}_c,p_L,p_H\right)},1\right) \tag{2.8}$$

Where $\theta^{(r)}=\left\{\beta_1,\sigma_1^2,\beta_2,\sigma_2^2,d\right\}^{(r)}$. Term α is compared with a draw u from a uniform distribution. Candidate draw c^* is accepted as $c^{(r)}$ if $u\le\alpha$ and otherwise $c^{(r)}=c^{(r-1)}$. Parameter Δ_c is the variance of the random walk Metropolis step and works as a tuning parameter to achieve an acceptance ratio between 10% and 50%.

2.2.2 The LSTAR model

A generic smooth transition autoregressive STAR(*P*) model with *P* lags can be written as follows:

$$y_t = \beta_{1,c} + \sum_{p=1}^{P}\beta_{1,p}y_{t-p} + G\left(z_t\right)\left(\beta_{0,c} + \sum_{p=1}^{P}\beta_{0,p}y_{t-p}\right) + e_t, \quad e_t \sim iid\ \mathcal{N}\left(0,\sigma^2\right) \tag{2.9}$$

or in a more compact notation as:

$$y_t = x_t\beta_1 + G\left(z_t\right)x_t\beta_0 + e_t \tag{2.10}$$

Where $x_t=\left[1,y_{t-1},\cdots,y_{t-P}\right]$.

In general, the transition function $G(z_t)$, which is triggered by the variable z_t, can take many forms. One of the most common approaches is to use the logistic function, which in its generic form of order *K* has the following form:

$$G(z_t,\gamma,\tilde{c},K)=\left(1+\exp\left(-\gamma\prod_{k=1}^{K}(z_t-c_k)\right)\right)^{-1},\gamma>0,\ \tilde{c}=[c_1,\cdots,c_K],c_k\in R \tag{2.11}$$

Depending on the order K and the threshold values $\tilde{c}$, the transition function can have many regimes. For example, for $K = 2$ the logistic function has a minimum at $(c_1 + c_2)/2$ and the parameters of the LSTAR model change symmetrically around this midpoint. For the specific case that $c_1 = c_2$, the transition function is actually the exponential function yielding the ESTAR model of Haggan and Ozaki (1981):

$$G(z_t,\gamma,c)=1-\exp\left(-\gamma(z_t-c)^2\right),\ \gamma>0,\ c\in R$$

As regards the transition variable z_t, in general, it can be any variable for which there is evidence that it triggers the transition from one regime to another via the transition function $G(z_t)$. Transition variable z_t can fall under three categories. Firstly, z_t can be an exogenous variable and potentially contemporaneous to y_t or a linear combination of stochastic variables which determine the regime that y_t falls into. Secondly, z_t can be set in a deterministic way according to time, for instance, in a linear way as $z_t = t$, which yields a deterministic time-varying parameter model. However, the third and most common approach consists of using a delayed value of the time series y_t itself as the transition variable, i.e. $z_t = y_{t-d}$, where for the delay parameter d, it holds $1 \le d \le P$.

The analysis below is restricted to the specific case of the LSTAR model where $K = 1$, and consequently vector $\tilde{c}$ in equation (2.11) collapses now to a scalar defined as c. The transition function is defined as:

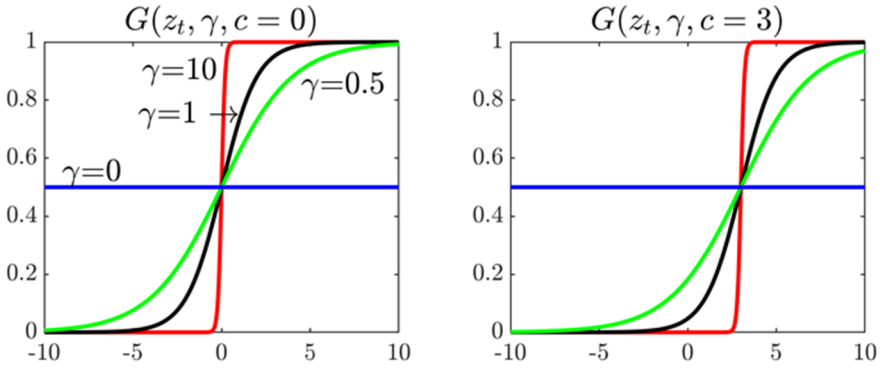
$$G(z_t,\gamma,c)=\left(1+\exp\left(-\gamma(y_{t-d}-c)\right)\right)^{-1} \tag{2.12}$$

For $\gamma = 0$, the transition function $G(y_{t-d},\gamma,c)=1/2$ such that the LSTAR model in equation (2.9) reduces to a linear AR model and an identification issue arises since the two regimes cannot be differentiated:

$$y_t=\beta_{1,c}+\sum_{p=1}^{P}\beta_{1,p}y_{t-p}+0.5\left(\beta_{0,c}+\sum_{p=1}^{P}\beta_{0,p}y_{t-p}\right)+e_t \tag{2.13}$$

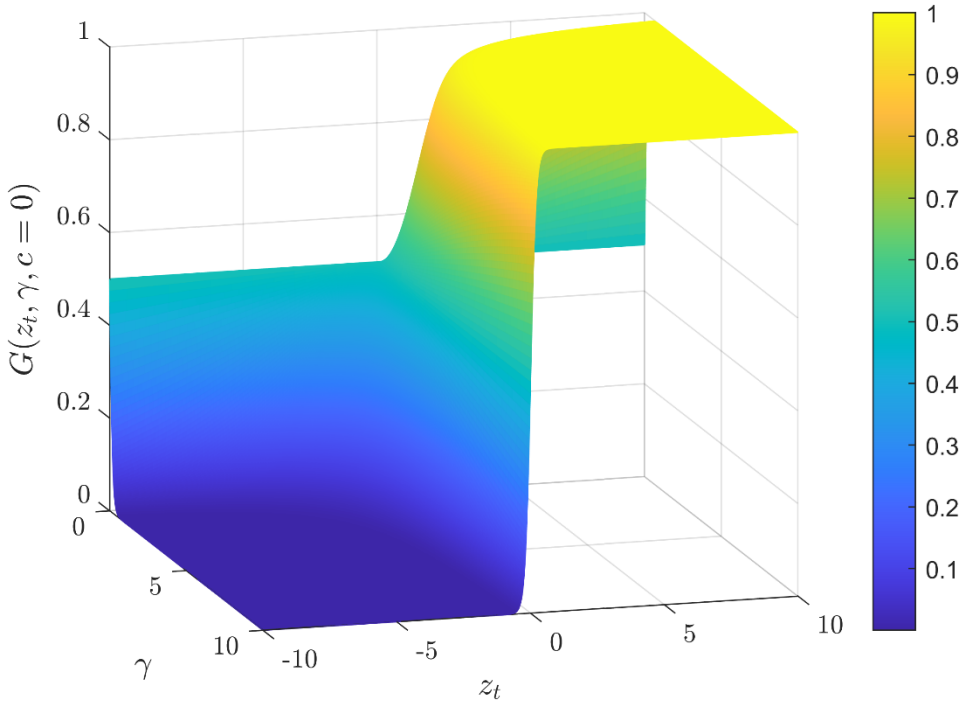
However, for large enough values of the shape parameter, $\gamma \to \infty$, the LSTAR model has a more abrupt transition from one regime to the other, yielding asymptotically a TAR model. See Figure 2.1 and Figure 2.2 below.

Figure 2.1: Parameterized logistic transition function



Note. Logistic transition function $G(z_t, \gamma, c) = \left(1 + \exp\left(-\gamma\left(y_{t-d} - c\right)\right)\right)^{-1}$ for different values of the shape parameter $\gamma = \{0, 0.5, 1, 10\}$ and threshold values $c = 0$ and $c = 3$ in the left and right plots, respectively.

Figure 2.2: Logistic transition function surface



Note. Logistic transition function $G(z_t, \gamma, c) = \left(1 + \exp\left(-\gamma\left(y_{t-d} - c\right)\right)\right)^{-1}$ surface for different values of shape parameter γ and transition variable z_t.

In order to make the effect of the logistic transition function on the model more obvious, using equation (2.10) and defining $\beta_2 = \beta_0 + \beta_1$, the LSTAR($P$) model with P lags[32] can be parameterized in its more common form in the literature:

$$y_t = \left(1 - G\left(y_{t-d}, \gamma, c\right)\right)\left(\beta_{1,c} + \sum_{p=1}^{P} \beta_{1,p} y_{t-p}\right) + G\left(y_{t-d}, \gamma, c\right)\left(\beta_{2,c} + \sum_{p=1}^{P} \beta_{2,p} y_{t-p}\right) + e_t \qquad (2.14)$$

As $G(y_{t-d}, \gamma, c) \rightarrow 0$, only the first regime defined by β_1 autoregressive coefficients holds for the LSTAR model, while for $G(y_{t-d}, \gamma, c) \rightarrow 1$, the second regime defined by β_2 autoregressive coefficients becomes active. Using this parametrization, as already discussed previously for $\gamma = 0$, that $G(y_{t-d}, \gamma, c) \rightarrow 1/2$, it becomes obvious from equation (2.14) why the two regimes cannot be differentiated since they become identical.

The LSTAR model's likelihood is given as follows:

$$\mathcal{L}\left(\beta_1, \beta_2, \sigma^2, d, \gamma, c \middle| Y\right) = \left(2\pi\sigma^2\right)^{-T/2} \exp\left(-\frac{1}{2}\frac{\tilde{e}'\tilde{e}}{\sigma^2}\right) \qquad (2.15)$$

In terms of the parameter priors and their respective posterior estimation, for parameters β_1, β_2, σ^2, and d, everything is similar to the case of the TAR model.

Regarding the autoregressive coefficients, β_1, β_2 prior, it has to be noted that although a conditionally conjugate prior has been used in the literature as in Lopes and Salazar (2006a), Lopes and Salazar (2006a, 2006b) have also proposed to use the Lubrano (2000) prior which takes the following form: $\beta_2 | \sigma^2, \gamma \sim \mathcal{N}(0, \sigma^2 e^{\gamma} I_{P+1})$ and $p(\beta_1, \gamma, \sigma^2, c) \sim (1+\gamma^2)^{-1} \sigma^2$. This conditional prior has the advantage of becoming informative for β_2 as $\gamma \rightarrow 0$ and addresses the identification problem that arises for the LSTAR model for $\gamma = 0$, where it collapses to a linear model. Throughout this essay, however, I use the same conditionally conjugate prior (see Section 2.2.3 for more details) for all the models to keep them comparable since for example Lubrano (2000) prior can not be applied to the TAR model and would need to be re-established for the proposed time-varying LSTAR models in Sections 2.2.3 and 2.2.4 that follow.

The estimation of the shape parameter γ combined with the threshold c is performed in a single random walk Metropolis-Hastings step as in Lopes and Salazar (2006a). Given the identification restriction of γ described previously, γ is modelled using a Gamma prior with a shape parameter $\underline{\alpha}_\gamma$ and a scale parameter $\underline{\beta}_\gamma$:

$$\gamma \sim \mathcal{G}\left(\underline{\alpha}_\gamma, \underline{\beta}_\gamma\right) \qquad (2.16)$$

Consequently, the MCMC algorithm is based on the following five blocks:

1. $p\left(\beta_1^{(r)} \middle| \left\{\beta_2, \sigma^2, d, \gamma, c\right\}^{(r-1)}, Y\right)$
2. $p\left(\beta_2^{(r)} \middle| \left\{\beta_1\right\}^{(r)}, \left\{\sigma^2, d, \gamma, c\right\}^{(r-1)}, Y\right)$
3. $p\left(\sigma^{2(r)} \middle| \left\{\beta_1, \beta_2\right\}^{(r)}, \left\{d, \gamma, c\right\}^{(r-1)}, Y\right)$

[32] Lopes and Salazar (2006a) allow for uncertainty over the lags number P by using the reversible jump MCMC algorithm of Green (1995).

4. $p\left(d^{(r)}\middle|\left\{\beta_1,\beta_2,\sigma^2\right\}^{(r)},\{\gamma,c\}^{(r-1)},Y\right)$
5. $p\left(\{\gamma,c\}^{(r)}\middle|\left\{\beta_1,\beta_2,\sigma^2,d\right\}^{(r)},Y\right)$

The posterior draws of the first four blocks are discussed previously for the case of the TAR model. Turning to the fifth block concerning the joint estimation of the threshold c and the shape parameter γ, its estimation is done with a random walk Metropolis-Hasting step as for the case of the TAR model, but now it is extended to accommodate the estimation of the shape parameter γ as well. At iteration (r) of the MCMC algorithm, a candidate set of values $\{\gamma, c\}^*$ is drawn as:

$$\{\gamma,c\}^* = \{\gamma,c\}^{(r-1)} + \xi, \quad \xi \sim iid\ \mathcal{N}\left(0,\Delta_\xi\right) \tag{2.17}$$

Where $\Delta_\xi = \text{diag}(\Delta_\gamma, \Delta_c)$ is a diagonal matrix with tuning parameters Δ_γ and Δ_c for the parameters γ and c, respectively.

The probability of accepting the candidate set of values $\{\gamma, c\}^*$ is estimated as:

$$\alpha = \min\left(\frac{p\left(y\middle|\{\gamma,c\}^*,\theta^{(r)}\right)}{p\left(y\middle|\{\gamma,c\}^{(r-1)},\theta^{(r)}\right)}\cdot\frac{p_{\mathcal{G}}\left(\gamma^*\middle|\underline{\alpha}_\gamma,\underline{\beta}_\gamma\right)}{p_{\mathcal{G}}\left(\gamma^{(r-1)}\middle|\underline{\alpha}_\gamma,\underline{\beta}_\gamma\right)}\cdot\frac{p_{\mathcal{N}_{tr}}\left(c^*\middle|\underline{\mathrm{M}}_c,\underline{\Omega}_c,p_L,p_H\right)}{p_{\mathcal{N}_{tr}}\left(c^{(r-1)}\middle|\underline{\mathrm{M}}_c,\underline{\Omega}_c,p_L,p_H\right)},1\right) \tag{2.18}$$

Where $\theta^{(r)} = \left\{\beta_1,\beta_2,\sigma^2\right\}^{\{r\}}$.

2.2.3 A homoscedastic hybrid time-varying LSTAR model

Having as a starting point the paper of Lopes and Salazar (2006a) on Bayesian estimation of STAR models and the modified ST-VAR model[33] with a time-varying threshold by Dueker, Owyang, and Sola (2010), this essay proposes the modified LSTAR model with two time-varying parameters:

1. The transition function $G(y_{t-d}, \gamma, c)$ is allowed to be a time-varying function $G_t(y_{t-d}, \gamma, c_t)$ by allowing the constant in time threshold c to be time-varying as c_t. The time-varying threshold can be interpreted in an economically meaningful way. Dueker, Owyang, and Sola (2010) have used it to model the natural rate of unemployment. In the same rationale, Zhu and Chen (2017) and Zhu, Chen, and Lin (2019) estimate a forward and backward looking, respectively regime dependent non-linear Taylor rule where the unemployment rate plays the role of the transition variable z_t, and the natural rate of unemployment is captured by the time-varying threshold c_t. While there are not any other prominent examples in the literature of (S)TAR models using time-varying thresholds, the long-run trends of many macroeconomic time series could fall under the same explanation. For example, for the case of short term interest rates c_t. could be attributed to the natural rate of interest which has been found to follow a declining trend (Hamilton, Harris, Hatzius, & West, 2016; Holston, Laubach, &

[33] Dueker, Owyang, and Sola (2010) use the normal cdf as a transition function instead of the logistic.

Williams, 2017). Productivity growth (Jorgenson, Ho, & Stiroh, 2008), output growth (Antolin-Diaz, Drechsel, & Petrella, 2017), and inflation (Ascari & Sbordone, 2014) exhibit similar trends. Another way of perceiving the threshold and, consequently, its time-varying equivalent falls under the VAR reduced form/time series literature rather than the more structural macroeconomic models. Along these lines, the threshold can represent the steady state of a stationary time series around which it fluctuates. Consequently, in the same way that the steady-state of a (multivariate) time-series (Villani, 2009) has been extended recently to its time-varying equivalent (Bańbura & van Vlodrop, 2018; Louzis, 2019), it could be advocated for the case of the time-varying threshold. The assumption of two regimes in the case of the logistic STAR model which follows different dynamics and in that sense impose a non-linearity in the time series can be explained by the voluminous literature (Clements & Krolzig, 2003; Morley & Piger, 2012) suggesting the asymmetry of the business cycle under recessionary and expansionary regimes.

2. The constant terms of the two regimes $(1-G(y_{t-d},\gamma,c))\beta_{1,c}$ and $G(y_{t-d},\gamma,c)\beta_{2,c}$ are combined in a single time-varying parameter $\beta_{c,t}$ such that it does not depend any more on the two regimes, but rather is independent and time-varying. In such a way, $\beta_{c,t}$ can capture potential structural breaks of the time-series in terms of level shifts, whereas in contrast, the combined time-varying equivalent coefficient from the two regimes when each regime has its own constant coefficient is bounded between the values of the two respective regimes. Using time-varying constants (intercepts) in linear models has also been used, for example, in mixed-frequency VARs (Götz & Hauzenberger, 2021), mixed-frequency factor models (Antolin-Diaz, Drechsel, & Petrella, 2017) and GDP unobserved component models with mean growth being modelled as a driftless random walk (Iseringhausen & Vierke, 2019). All of the above papers use time-varying constants as a way to capture slow-moving trends while, at the same time, the remaining autoregressive or factor-loading coefficients remain constant in time. I follow the same approach, assuming that a non-linear time series which exhibits different behaviour above and below a (time-varying) threshold may also exhibit long-run trends. It becomes obvious that when a time series presents a slow-moving trend, the threshold which defines the two regimes should also be flexible to vary in time such that it allows the time series to fall in any of the two regimes as time progresses. Otherwise, for example for a time series with an upward trend, if the threshold has a value equal to the time series around the mid of the sample, it would result in the first regime being activated for the first half of the time series and the second regime being activated for the second half of the series.

As a result, the corresponding hybrid time-varying LSTAR model (TV-LSTAR) can be represented by a non-linear state-space model. Assuming that the law of motion of the time-varying parameters follows the standard approach in TVP macroeconomic models of a random walk or an autoregressive process, it follows that the non-linear state-space model representation of the hybrid time-varying LSTAR is given by the following observation and state equations:

$$\begin{aligned} y_t &= \beta_{c,t} + \left(1-G\left(y_{t-d},\gamma,c_t\right)\right)\left(\sum_{p=1}^{P}\beta_{1,p}y_{t-p}\right) \\ &+ G\left(y_{t-d},\gamma,c_t\right)\left(\sum_{p=1}^{P}\beta_{2,p}y_{t-p}\right)+e_t, \quad e_t \sim \mathcal{N}\left(0,R\right) \end{aligned} \tag{2.19}$$

$$\underbrace{\begin{bmatrix}\beta_{c,t}\\ c_t\end{bmatrix}}_{S_t}=\underbrace{\begin{bmatrix}\rho_{\beta c,0}\\ \rho_{c,0}\end{bmatrix}}_{C}+\underbrace{\begin{bmatrix}\rho_{\beta c,1} & 0\\ 0 & \rho_{c,1}\end{bmatrix}}_{F}\underbrace{\begin{bmatrix}\beta_{c,t-1}\\ c_{t-1}\end{bmatrix}}_{S_{t-1}}+\underbrace{\begin{bmatrix}v_{\beta c,t}\\ v_{c,t}\end{bmatrix}}_{v_t},\quad v_t \sim \mathcal{N}(0,Q) \tag{2.20}$$

Where $G(y_{t-d},\gamma,c_t)=\left(1+\exp\left(-\gamma\left(y_{t-d}-c_t\right)\right)\right)^{-1}$ is the logistic transition function, as already discussed previously.

The unknown state vector of the system $S_t=\left[\beta_{c,t},c_t\right]'$ plays the role of the latent variables for the hybrid time-varying LSTAR that need to be estimated. For the rest of the essay, *S* denotes the stacked states S_t over the full sample. i.e. $S=\left\lfloor S_t' \right\rfloor_{t=1}^{T}$.

In order to estimate the model in equations (2.19) and (2.20), the standard Kalman filtering and smoothing techniques like Carter and Kohn (1994) algorithm which apply to linear state-space models with additive Gaussian noise and can be part of a Gibbs sampling routine is not valid anymore. The latent state variables of non-linear state-space models with known parameters, however, can be estimated using suitable algorithms. While the literature in this area is rather voluminous[34], the most popular approaches include the extended Kalman filter (EKF), the unscented Kalman filter (UKF) and the particle filtering (PF) or sequential Monte Carlo (SMC).

For non-linear state-space models with unknown parameters, there are two main avenues to proceed. The first one estimates latent states with a Metropolis-Hastings within Gibbs step separately for each time period *t*, where the candidates are drawn from a suitably tailored proposal density such that a suitable acceptance ratio is achieved (Geweke & Tanizaki, 2001). The second approach (Andrieu, Doucet, & Holenstein, 2010) uses particle filtering in MCMC algorithms and comprises two main approaches: particle Metropolis-Hastings which has been used, for example extensively in the literature of DSGE models[35] and particle Gibbs algorithm, which has been implemented for time series models for example in Nonejad (2016) and Mumtaz (2018).

In the relevant literature on time-varying threshold (S)TAR models (Dueker, Owyang, & Sola, 2010; Zhu & Chen, 2017; Zhu, Chen, & Lin, 2019), the time-varying threshold is estimated via an independent Metropolis-Hasting algorithm as discussed previously. In particular, in the ST-VAR of Dueker, Owyang, and Sola (2010), the time-varying threshold is estimated by drawing candidates from the proposal density of the smoothed state estimates after running the UKF[36], whereas Zhu and Chen (2017); Zhu, Chen, and Lin (2019) use suitable tailored distributions after expressing the time-varying threshold TAR in a state-space representation.

Instead, I choose to proceed to the estimation of the latent state variables and the model's parameters using an improved version of the Particle Gibbs algorithm named as Particle Gibbs

[34] For a detailed textbook treatment of state-space models refer to and Durbin and Koopman (2012); Särkkä (2013), and Chopin and Papaspiliopoulos (2020).

[35] See for example Fernández-Villaverde and Rubio-Ramírez (2005) and Herbst and Schorfheide (2015) for a textbook treatment.

[36] UKF (Julier, Uhlmann, & Durrant-Whyte, 1995; Julier & Uhlmann, 2004; Wan & Van Der Merwe, 2001) is based on the unscented transform which uses a set of sigma points to capture the mean and covariance of a random variable which are subsequently propagated via a non-linear function. The propagated sigma points and are used afterwards to estimate the mean and covariance of the non-linearly transformed random variable.

with Ancestor Sampling (Lindsten, Jordan, & Schon, 2014). The exact implementation details of this algorithm are presented in Appendix B.1.1 Particle Gibbs with ancestor sampling.

In a fully Bayesian estimation setting, the model's parameters have the following priors.

The parameters of the two conditional means of the model, i.e. β_1, β_2 are conditionally conjugate:

$$\beta_i \sim \mathcal{N}\left(\underline{\mathrm{M}}_{\beta_i}, \underline{\Omega}_{\beta_i}\right), \quad i = 1,2 \tag{2.21}$$

Where $\beta_i = \left[\beta_{i,1}, \cdots, \beta_{i,P}\right]$ for $i = 1,2$. With respect to $\underline{\Omega}_{\beta_i}$, instead of using an identical prior's variance for all lags' coefficients (Chen & Lee, 1995; Lopes & Salazar, 2006a; Dueker, Owyang, & Sola, 2010; Zhu & Chen, 2017; Zhu, Chen, & Lin, 2019), i.e. $\underline{\Omega}_{\beta_i} = \underline{\sigma}_{\beta}^2 I_{P \times P}$, I rather allow for extra shrinkage for the more distant lags' coefficients $\beta_{i,p}$ following a Minnesota style prior (Litterman, 1986) adapted to the univariate case:

$$\underline{\Omega}_{\beta_i} = \frac{\lambda_1}{p^{\lambda_2}}, \quad i = 1,2, \quad p = 1, \cdots, P \tag{2.22}$$

The autoregressive coefficients of the time-varying constant $\beta_{c,t}$ ($j = 1$), i.e. $\rho_{\beta c} = \rho_1 = \left[\rho_{\beta c,0}, \rho_{\beta c,1}\right] = \left[C_1, F_{1,1}\right]$ and the threshold[37] c_t ($j = 2$), i.e. $\rho_c = \rho_2 = \left[\rho_{c,0}, \rho_{c,1}\right] = \left[C_2, F_{2,2}\right]$ follow a normal prior:

$$\rho_j \sim \mathcal{N}\left(\underline{\mathrm{M}}_{\rho_j}, \underline{\Omega}_{\rho_j}\right), \quad j = 1,2 \tag{2.23}$$

Using a normal prior for the time-varying threshold process is similar to Dueker, Owyang, and Sola (2010), while Zhu and Chen (2017); Zhu, Chen, and Lin (2019) use a uniform prior defined over $(-1,1)$ such that the autoregressive process does not exhibit any explosive roots.

The variance Q of noise v_t for the state equation (2.20) is assumed to be diagonal, with its diagonal elements following an inverse Gamma prior:

$$Q_{j,j} \sim \mathcal{IG}\left(\underline{S}_{Q_{j,j}}, \underline{d}_{Q_{j,j}}\right), \quad j = 1,2 \tag{2.24}$$

Delay parameter d in the generic case follows a discrete distribution defined as:

$$d \sim p\left(d = d_p\right) \tag{2.25}$$

For models with a small number of maximum lag P, it is reasonable to assume that the delay parameter d follows a discrete uniform distribution $d \sim \mathcal{U}(1, P)$. However, when the number of lags increases, as would be the case, for example, when estimating a model with a time series in monthly frequency, it could be more appropriate to assign larger probabilities to the closest lags and smaller to the more distant ones.

The variance R of the model's innovations e_t follow an inverse Gamma prior:

[37] I use index j to refer to the 2 state variables $\beta_{c,t}$ and c_t to differentiate with index i used to denote the set of autoregressive coefficients β_1 and β_2 above and below the time-varying threshold c_t.

$$R \sim \mathcal{IG}\left(\underline{S}_R, \underline{d}_R\right) \tag{2.26}$$

Where $\underline{S}_R$ is the prior scale matrix, and $\underline{d}_R$ are the degrees of freedom.

Finally, the shape parameter γ is modelled with a gamma prior with a shape parameter $\underline{\alpha}_\gamma$ and a scale parameter $\underline{\beta}_\gamma$ as already discussed previously:

$$\gamma \sim \mathcal{G}\left(\underline{\alpha}_\gamma, \underline{\beta}_\gamma\right) \tag{2.27}$$

Consequently, the iteration (r) from the particle Gibbs with ancestor sampling algorithm is based on the following eight blocks:

1. $p\left(\beta_1^{(r)} \middle| \{\beta_2, \rho, Q, d, R, \gamma, S\}^{(r-1)}, Y\right)$
2. $p\left(\beta_2^{(r)} \middle| \{\beta_1\}^{(r)}, \{\rho, Q, d, R, \gamma, S\}^{(r-1)}, Y\right)$
3. $p\left(\rho^{(r)} \middle| \{\beta_1, \beta_2\}^{(r)}, \{Q, d, R, \gamma, S\}^{(r-1)}, Y\right)$
4. $p\left(Q^{(r)} \middle| \{\beta_1, \beta_2, \rho\}^{(r)}, \{d, R, \gamma, S\}^{(r-1)}, Y\right)$
5. $p\left(d^{(r)} \middle| \{\beta_1, \beta_2, \rho, Q\}^{(r)}, \{R, \gamma, S\}^{(r-1)}, Y\right)$
6. $p\left(R^{(r)} \middle| \{\beta_1, \beta_2, \rho, Q, d\}^{(r)}, \{\gamma, S\}^{(r-1)}, Y\right)$
7. $p\left(\gamma^{(r)} \middle| \{\beta_1, \beta_2, \rho, Q, d\}^{(r)}, \{S\}^{(r-1)}, Y\right)$
8. $p\left(S^{(r)} \middle| \{\beta_1, \beta_2, \rho, Q, d, \gamma\}^{(r)}, Y\right)$

The first and the second block of the Gibbs sampler have a conditional posterior, which is derived easily along the lines of a Bayesian linear regression conditional on the rest of the parameters:

$$y_t - \beta_{c,t} - G\left(y_{t-d}, \gamma, c_t\right)\left(\sum_{p=1}^{P} \beta_{2,p} y_{t-p}\right) = \left(1 - G\left(y_{t-d}, \gamma, c_t\right)\right)\left(\sum_{p=1}^{P} \beta_{1,p} y_{t-p}\right) + e_t \tag{2.28}$$

Redefining the left-hand side of equation (2.28) as $\tilde{y}_t = y_t - \beta_{c,t} - G\left(y_{t-d}, \gamma, c_t\right)\left(\sum_{p=1}^{P} \beta_{2,p} y_{t-p}\right)$ and defining the product of the transition function with the lagged values of y_t as $\tilde{x}_t = \left(1 - G\left(y_{t-d}, \gamma, c_t\right)\right)\left[y_{t-1}, \cdots, y_{t-P}\right]$ then equation (2.28) transforms to the following standard Bayesian linear regression:

$$\tilde{y}_t = \tilde{x}_t \beta_1 + e_t \tag{2.29}$$

Where the posterior of β_1 is distributed as:

$$\beta_1 \sim \mathcal{N}\left(\overline{\mathrm{M}}_{\beta_1}, \overline{\Omega}_{\beta_1}\right) \tag{2.30}$$

Where $\overline{\Omega}_{\beta_1} = \left(\underline{\Omega}_{\beta_1}^{-1} + R^{-1}\left(\tilde{X}'\tilde{X}\right)\right)^{-1}$ and $\overline{\mathrm{M}}_{\beta_1} = \overline{\Omega}_{\beta_1}\left(\underline{\Omega}_{\beta_1}^{-1}\underline{\mathrm{M}}_{\beta_1} + R^{-1}\tilde{X}'\tilde{Y}\right)$ with $\tilde{Y} = \lfloor \tilde{y}_t \rfloor_{t=1}^{T}$ and $\tilde{X} = \lfloor \tilde{x}_t \rfloor_{t=1}^{T}$.

Similarly, the parameters β_2 of the second regime can be estimated by formulating the equivalent Bayesian linear regression conditional on the draw of β_1 and the remaining set of parameters.

The third block concerning the posterior of the autoregressive coefficients $\rho = [\rho_1, \rho_2]$ for the time-varying constant $\beta_{c,t}$ and threshold c_t is estimated as:

$$\rho_j \sim \mathcal{N}\left(\overline{\mathrm{M}}_{\rho_j}, \overline{\Omega}_{\rho_j}\right), \quad j = 1,2 \tag{2.31}$$

Where $\overline{\Omega}_{\rho_j} = \left(\underline{\Omega}_{\rho_j}^{-1} + Q_{j,j}^{-1}\left(\tilde{X}'\tilde{X}\right)\right)^{-1}$ and $\overline{\mathrm{M}}_{\rho_j} = \overline{\Omega}_{\rho_j}\left(\underline{\Omega}_{\rho_j}^{-1}\underline{\mathrm{M}}_{\rho_j} + Q_{j,j}^{-1}\tilde{X}'\tilde{Y}\right)$ with $\tilde{Y} = \lfloor S_{j,t} \rfloor_{t=1}^{T}$ and $\tilde{X} = \lfloor S_{j,t-1} \rfloor_{t=1}^{T}$.

The fourth block concerning the variance of the gaussian errors for the state equations has the following conditional posterior:

$$Q_{j,j} \sim \mathcal{IG}\left(\overline{S}_{Q_{j,j}}, \overline{d}_{Q_{j,j}}\right) \tag{2.32}$$

Where $\overline{S}_{Q_{j,j}} = \underline{S}_{Q_{j,j}} + \tilde{e}_j'\tilde{e}_j$ and $\overline{d}_{Q_{j,j}} = \underline{d}_{Q_{j,j}} + T$ with $\tilde{e}_j = \lfloor S_{j,t} - C_j - S_{j,t-1}F_{j,j} \rfloor_{t=1}^{T}$ denoting the residuals of the j-th column of the $T \times 2$ matrix S.

The fifth block of the conditional posterior of the delay parameter d follows a multinomial distribution with probability:

$$p\left(d \middle| \beta_1, \beta_2, \rho, Q, R, d, \gamma, S, Y\right) = \frac{\mathcal{L}\left(\beta_1, \beta_2, \rho, Q, R, d_p, \gamma, S \middle| Y\right) p\left(d = d_p\right)}{\sum_{p=1}^{P} \mathcal{L}\left(\beta_1, \beta_2, \rho, Q, R, d_p, \gamma, S \middle| Y\right) p\left(d = d_p\right)}, \quad p = 1, \cdots, P \tag{2.33}$$

Where $p\left(d = d_p\right)$ denotes the prior of the delay parameter d as defined in equation (2.25) with $d_p = 1, \cdots, P$. $\mathcal{L}\left(\beta_1, \beta_2, \rho, Q, R, d, \gamma, S \middle| Y\right)$ denotes the conditional likelihood[38] of the model:

$$\mathcal{L}\left(\beta_1, \beta_2, \rho, Q, R, d, \gamma, S \middle| Y\right) = \left(2\pi R\right)^{-T/2} \exp\left(-\frac{1}{2}\frac{\tilde{e}'\tilde{e}}{R}\right) \tag{2.34}$$

With $\tilde{e}$ denoting the estimated residuals defined over the full sample:

[38] Conditional likelihood refers to the fact that the likelihood is conditional with respect to the latent states of the non-linear state-space model, i.e. $\mathcal{L}(\theta, S \mid Y) = p(Y \mid \theta, S)$, where $\theta = \{\beta_1, \beta_2, \rho, Q, d, R, \gamma\}$.

$$\tilde{e} = \left[y_t - \beta_{c,t} - \left(1 - G\left(y_{t-d}, \gamma, c_t\right)\right)\left(\sum_{p=1}^{P} \beta_{1,p} y_{t-p}\right) - G\left(y_{t-d}, \gamma, c_t\right)\left(\sum_{p=1}^{P} \beta_{2,p} y_{t-p}\right) \right]_{t=1}^{T}$$

For the specific case where the delay parameter d follows a discrete uniform prior, i.e. $d \sim \mathcal{U}(1, P)$, the relevant terms $p\left(d = d_p\right)$ simplify from equation (2.33) and the posterior is given as the ratios of the likelihood for each delay parameter d_p over the sum of the likelihoods for all potential delay parameters d_p, for $p = 1, \cdots, P$.

The sixth block of the conditional posterior for the variance R is given by:

$$R \sim \mathcal{IG}\left(\overline{S}_R, \overline{d}_R\right) \tag{2.35}$$

Where $\overline{S}_R = \underline{S}_R + \tilde{e}'\tilde{e}$ and $\overline{d}_R = \underline{d}_R + T$.

The seventh block concerning the estimation of the shape parameter γ is estimated following a Metropolis-within-Gibbs step, as already discussed previously. At iteration (r) of the MCMC algorithm, a candidate value of γ^* is drawn as:

$$\gamma^* = \gamma^{(r-1)} + \xi \,,\ \xi \sim iid\ \mathcal{N}\left(0, \Delta_\gamma\right) \tag{2.36}$$

The probability of accepting γ^* is estimated as:

$$\alpha = \min\left(\frac{p\left(y \middle| \gamma^*, \theta\right)}{p\left(y \middle| \gamma^{(r-1)}, \theta\right)} \frac{p_{\mathcal{G}}\left(\gamma^* \middle| \underline{\alpha}_\gamma, \underline{\beta}_\gamma\right)}{p_{\mathcal{G}}\left(\gamma^{(r-1)} \middle| \underline{\alpha}_\gamma, \underline{\beta}_\gamma\right)} \right) \tag{2.37}$$

Where $\theta = \left\{ \left\{ \beta_1, \beta_2, Q, d, R \right\}^{(r)}, S^{(r-1)} \right\}$.

Finally, the eighth block concerning the posterior distribution of the states S is estimated using the Lindsten, Jordan, and Schon (2014) algorithm. More details are provided in Appendix B.1.1 Particle Gibbs with ancestor sampling.

2.2.4 A heteroscedastic hybrid time-varying LSTAR model with stochastic volatility

One popular approach for modelling heteroscedasticity for LSTAR models usually comes under the form of regime-dependent heteroscedasticity with variances σ_1^2 and σ_2^2 for each respective regime, as follows:

$$\sigma_t^2 = \left(1 - G\left(y_{t-d}, \gamma, c_t\right)\right)\sigma_1^2 + G\left(y_{t-d}, \gamma, c_t\right)\sigma_2^2 \tag{2.38}$$

This approach allows for a smooth transition in innovations variances when alternating between the two regimes and has been applied in the literature even for smooth transition multivariate models in Galvão and Owyang (2018) using a ST-FAVAR and Auerbach and Gorodnichenko (2012) and Caggiano, Castelnuovo, and Groshenny (2014) using ST-VARs.

Along these lines Gerlach and Chen (2008) substitute the regime-dependent variances σ_1^2 and σ_2^2 in equation (2.38) with two GARCH(q,r) processes, each one for each regime i = 1,2, i.e. yielding an ST-GARCH(q,r) process:

$$\begin{aligned}\sigma_t^2 &= \left(1-G\left(y_{t-d},\gamma,c_t\right)\right)\sigma_{1,t}^2 + G\left(y_{t-d},\gamma,c_t\right)\sigma_{2,t}^2 \\ \sigma_{i,t}^2 &= \alpha_{i,0} + \sum_{k=1}^{q}\alpha_{i,k}\sigma_{t-k}^2 + \sum_{l=1}^{r}\gamma_{i,l}e_{t-l}^2, \quad i=1,2\end{aligned} \tag{2.39}$$

In a similar context, Lopes and Salazar (2006b) form an ST stochastic volatility model[39] for the conditional variance $\sigma_t^2 = h_t$, where the natural logarithm of the stochastic volatility follows an AR(1) process:

$$\begin{aligned}\ln h_t &= \left(1-G\left(y_{t-d},\gamma,c_t\right)\right)\ln h_{1,t} + G\left(y_{t-d},\gamma,c_t\right)\ln h_{2,t} \\ \ln h_{i,t} &= \alpha_{i,0} + \alpha_{i,1}\ln h_{t-1} + \eta_t, \quad i=1,2\end{aligned} \tag{2.40}$$

On the other hand, Livingston and Nur (2020) model the conditional variance of the LSTAR model as a simple GARCH(q,r) process without applying an ST structure. Following the same approach, I use instead for the proposed TV-LSTAR-SV model, the more popular process for macroeconomic time series[40] of stochastic volatility, i.e. $R_t = h_t$, where the natural logarithm of h_t follows a random walk:

$$\ln h_t = \ln h_{t-1} + \eta_t,\ \eta_t \sim iid\ \mathcal{N}\left(0,\Phi\right) \tag{2.41}$$

Thus, the suggested TV-LSTAR-SV is specified with heteroscedasticity, which does not depend directly on any of the regimes but rather captures the statistical properties of the non-linear time series in a timely manner. Conditional on the remaining blocks of the Gibbs sampler, stochastic volatility can now replace the sixth block in Section 2.2.3 and be estimated using Kim, Shephard, and Chib (1998) algorithm. An additional block is required to estimate the posterior of the variance of the innovations of the stochastic volatility under the standard approach following an inverse Gamma prior. The MCMC algorithm is now comprised of the following nine blocks.

1. $p\left(\beta_1^{(r)}\middle|\left\{\beta_2,\rho,Q,d,h,\Phi,\gamma,S\right\}^{(r-1)},Y\right)$
2. $p\left(\beta_2^{(r)}\middle|\left\{\beta_1\right\}^{(r)},\left\{\rho,Q,d,h,\Phi,\gamma,S\right\}^{(r-1)},Y\right)$
3. $p\left(\rho^{(r)}\middle|\left\{\beta_1,\beta_2\right\}^{(r)},\left\{Q,d,h,\Phi,\gamma,S\right\}^{(r-1)},Y\right)$
4. $p\left(Q^{(r)}\middle|\left\{\beta_1,\beta_2,\rho\right\}^{(r)},\left\{d,h,\Phi,\gamma,S\right\}^{(r-1)},Y\right)$
5. $p\left(d^{(r)}\middle|\left\{\beta_1,\beta_2,\rho,Q\right\}^{(r)},\left\{h,\Phi,\gamma,S\right\}^{(r-1)},Y\right)$
6. $p\left(h^{(r)}\middle|\left\{\beta_1,\beta_2,\rho,Q,d\right\}^{(r)},\left\{\Phi,\gamma,S\right\}^{(r-1)},Y\right)$
7. $p\left(\Phi^{(r)}\middle|\left\{\beta_1,\beta_2,\rho,Q,d,h\right\}^{(r)},\left\{\gamma,S\right\}^{(r-1)},Y\right)$

[39] The respective ST-SV MCMC block is estimated using the Jacquier, Polson, and Rossi (1994) algorithm.
[40] Livingston and Nur (2020) use their LSTAR-GARCH model on climate time series.

8. $p\left(\gamma^{(r)}\middle|\{\beta_1,\beta_2,\rho,Q,d,h,\Phi\}^{(r)},\{S\}^{(r-1)},Y\right)$
9. $p\left(S^{(r)}\middle|\{\beta_1,\beta_2,\rho,Q,d,h,\Phi,\gamma\}^{(r)},Y\right)$

The MCMC blocks need to be adjusted accordingly so that they take into account the stochastic volatility. Regarding the first and the second block, these should be adjusted in the following way: The posterior distribution of the autoregressive coefficients of the first regime β_1 conditional on the second regime autoregressive coefficients β_2 and the remaining parameters as described by equation (2.28) is now given by:

$$\beta_1 \sim \mathcal{N}\left(\overline{\mathrm{M}}_{\beta_1},\overline{\Omega}_{\beta_1}\right) \tag{2.42}$$

Where $\overline{\Omega}_{\beta_1}=\left(\underline{\Omega}_{\beta_1}^{-1}+\sum_{t=1}^{T}R_t^{-1}\left(\tilde{X}_t'\tilde{X}_t\right)\right)^{-1}$ and $\overline{\mathrm{M}}_{\beta_1}=\overline{\Omega}_{\beta_1}\left(\underline{\Omega}_{\beta_1}^{-1}\underline{\mathrm{M}}_{\beta_1}+\sum_{t=1}^{T}\tilde{X}_t'\tilde{Y}_tR_t^{-1}\right)$

More details on the sixth block concerning the estimation of the stochastic volatility using Kim, Shephard, and Chib (1998) algorithm can be found in Appendix

B.1.3 Stochastic volatility estimation.

The remaining blocks, i.e. the fifth, the eighth and ninth, need to be adjusted accordingly so that they take into account the stochastic volatility R_t at each period t. As a result, in the fifth and eighth blocks requiring the estimation of the likelihood, the conditional likelihood is used:

$$\mathcal{L}\left(\beta_1,\beta_2,\rho,Q,d,h,\Phi,\gamma,S\middle|Y\right)=\prod_{t=1}^{T}\left(2\pi R_t\right)^{-1/2}\exp\left(-\frac{1}{2}\frac{e_t^2}{R_t}\right) \tag{2.43}$$

Finally, the seventh block draws the posterior of the innovations variance to the stochastic volatility as:

$$\Phi \sim \mathcal{IG}\left(\overline{S}_{\Phi},\overline{d}_{\Phi}\right) \tag{2.44}$$

Where $\overline{S}_{\Phi}=\underline{S}_{\Phi}+\sum_{t=1}^{T}\left(\Delta\ln h_t\right)^2$ and $\overline{d}_{\Phi}=\underline{d}_{\Phi}+T$.

2.2.5 Forecasting

Multistep-ahead forecasts for the entire class of the non-linear models presented previously that entail thresholds are iterated forecasts. For the entire forecast horizon, the models are allowed to alternate between the two different regimes as the threshold or transition function dictates, given the lagged values of the iterated forecasts. In other words, for example, the TAR model does not remain in the regime that fell during the last observation y_T, throughout the entire forecast horizon $h = 1,\cdots, H$, but at each h-step ahead forecast value $\hat{y}_{T+h|T}$ regime

one or two is activated depending on whether the *d* lagged forecasted value $\hat{y}_{T+h-d|T}$ is below or above the threshold value *c*. To formalise ideas at each MCMC iteration (*r*), the *h*-step ahead forecast value $\hat{y}_{T+h|T}^{(r)}$ for the TAR model is given as:

$$\hat{y}_{T+h|T}^{(r)} = \begin{cases} \beta_{1,c}^{(r)} + \sum_{p=1}^{P} \beta_{1,p}^{(r)} \hat{y}_{T+h-p|T} + e_{1,T+h}^{(r)}, \quad e_{1,T+h}^{(r)} \sim \mathcal{N}\left(0, \sigma_1^{2(r)}\right), \quad \hat{y}_{T+h-d^{(r)}|T} \leq c^{(r)} \\ \beta_{2,c}^{(r)} + \sum_{p=1}^{P} \beta_{2,p}^{(r)} \hat{y}_{T+h-p|T} + e_{2,T+h}^{(r)}, \quad e_{2,T+h}^{(r)} \sim \mathcal{N}\left(0, \sigma_2^{2(r)}\right), \quad \hat{y}_{T+h-d^{(r)}|T} > c^{(r)} \end{cases} \tag{2.45}$$

For the proposed time-varying parameter LSTAR models, additionally to allowing the model to alternate between the two regimes throughout the forecast horizon, the time-varying parameters are propagated out-of-sample throughout the forecast horizon following their respective laws of motion. To be more specific, the *h*-step ahead forecast $\hat{y}_{T+h|T}^{(r)}$ at MCMC iteration (*r*) for the TV-LSTAR-SV model is given as:

$$\begin{aligned} \hat{y}_{T+h|T}^{(r)} = \hat{\beta}_{c,T+h|T}^{(r)} + \left(1 - G\left(y_{T+h-d|T}^{(r)}, \gamma^{(r)}, \hat{c}_{T+h|T}^{(r)}\right)\right)\left(\sum_{p=1}^{P} \beta_{1,p}^{(r)} \hat{y}_{T+h-p|T}^{(r)}\right) + \\ G\left(y_{T+h-d|T}^{(r)}, \gamma^{(r)}, \hat{c}_{T+h|T}^{(r)}\right)\left(\sum_{p=1}^{P} \beta_{2,p}^{(r)} \hat{y}_{T+h-p|T}^{(r)}\right) + e_{T+h}^{(r)}, \quad e_{T+h}^{(r)} \sim \mathcal{N}\left(0, \hat{R}_{T+h|T}^{(r)}\right) \end{aligned} \tag{2.46}$$

Where the time-varying parameters $\beta_{c,t}$, c_t, and R_t have been propagated for period $T + h$ with their respective modelling process:

$$\begin{cases} \hat{\beta}_{c,T+h|T}^{(r)} = \rho_{\beta c,0}^{(r)} + \rho_{\beta c,1}^{(r)} \hat{\beta}_{c,T+h-1|T}^{(r)} + v_{\beta c,T+h}^{(r)}, \quad v_{\beta c,T+h}^{(r)} \sim \left(0, Q_{v_{\beta c}}^{(r)}\right) \\ \hat{c}_{T+h|T}^{(r)} = \rho_{c,0}^{(r)} + \rho_{c,1}^{(r)} \hat{c}_{T+h-1|T}^{(r)} + v_{c,T+h}^{(r)}, \quad v_{c,T+h}^{(r)} \sim \left(0, Q_{v_c}^{(r)}\right) \\ \hat{R}_{T+h|T}^{(r)} = \exp\left(\ln \hat{h}_{T+h|T}^{(r)}\right): \ln \hat{h}_{T+h|T}^{(r)} = \ln \hat{h}_{T+h-1|T}^{(r)} + \eta_{T+h}^{(r)}, \quad \eta_{T+h}^{(r)} \sim \left(0, \Phi^{(r)}\right) \end{cases} \tag{2.47}$$

Having obtained the forecasts $\hat{y}_{T+h|T}^{(r)}$ for all $r = 1, \cdots, R$ MCMC iterations, the *h*-step ahead forecast $\hat{y}_{T+h|T}$ is estimated simply as the mean:

$$\hat{y}_{T+h|T} = \frac{1}{R} \sum_{r=1}^{R} \hat{y}_{T+h|T}^{(r)}$$

2.3 Model comparison

Model comparison in a Bayesian framework can be performed by using Bayes factors and consequently estimating marginal likelihoods or by using information criteria. Marginal likelihood estimation for the proposed models, which contain latent states due to the time-

varying parameters[41], poses some challenges which have been well discussed in the literature. The following sections present in more detail the marginal likelihood and information criteria estimation for the proposed models.

2.3.1 Marginal likelihood estimation

For all the previously discussed models, their marginal likelihood can be approximated using the Gelfand and Dey (1994) method. In more detail, it can be shown that the marginal likelihood $p(Y|\mathcal{M})$ for a given model $\mathcal{M}$ can be approximated[42] as:

$$\hat{p}\left(Y\middle|\mathcal{M}\right)=\left[\frac{1}{N_R}\sum_{r=1}^{N_R}\frac{f\left(\theta^{(r)}\right)}{p\left(Y\middle|\theta^{(r)},\mathcal{M}\right)p\left(\theta^{(r)}\middle|\mathcal{M}\right)}\right]^{-1} \tag{2.48}$$

Where $p\left(Y\middle|\theta^{(r)},\mathcal{M}\right)$ refers to the posterior, $p\left(\theta^{(r)}\middle|\mathcal{M}\right)$ is the prior, and $f(\theta)$ is any pdf function defined over the parameters vector θ of length K, with the property $\int f(\theta)d\theta=1$. The summation in equation (2.48) refers to the retained draws N_R after the burn-in period.

In line with Geweke (1999) proposal of a truncated normal distribution, $f\left(\theta^{(r)}\right)$ is defined as:

$$f\left(\theta^{(r)}\right)=\tau^{-1}(2\pi)^{-0.5K}\left|V_\theta\right|^{-0.5}\exp\left(-0.5\left(\theta^{(r)}-\bar{\theta}\right)'V_\theta^{-1}\left(\theta^{(r)}-\bar{\theta}\right)\right)$$
$$\times\mathbb{I}\left\{\left(\theta^{(r)}-\bar{\theta}\right)'V_\theta^{-1}\left(\theta^{(r)}-\bar{\theta}\right)\le F^{-1}(\tau)\right\}$$

Where $\bar{\theta}$ is the mean of the vector of parameters over all the retained draws N_R, and V_θ is the variance-covariance of the parameters. $\mathbb{I}\{z\le\alpha\}$ is the indicator function taking the value of one if $z\le\alpha$ and zero otherwise. $F^{-1}(\tau)$ is the inverse cdf of a chi-square distribution $\mathcal{X}_K^2$ with K degrees of freedom and probability τ. Thus the role of the tuning parameter τ is to remove the extreme values from the parameters θ during the Monte Carlo integration.

Thus for the estimation of the marginal likelihood of each of the four models mentioned above, i.e. TAR, LSTAR, TV-LSTAR, TV-LSTAR-SV, the essential part breaks down to the estimation of the posterior at each retained draw which is used in the denominator term in equation (2.48). This can be easily estimated by using the relevant likelihood functions as presented in equations (1.1) and (2.5) for the TAR model and equation (2.15) for the LSTAR model and further estimating the pdf of each parameter prior at the respective draw of the MCMC algorithm. However, for the TV-LSTAR, TV-LSTAR-SV models, which include latent states following Chan and Grant (2015) and Chan and Grant (2016), rather than using

[41] Frühwirth-Schnatter and Wagner (2010) have proposed a stochastic model specification search algorithm to address model uncertainty regarding constant versus time-varying coefficients for linear state-space models. For an application in inflation's volatility and NAIRU time variation see Chan (2018).

[42] For a formal proof you can refer to Koop (2003), p.105 and Herbst and Schorfheide (2015), p. 93.

the conditional likelihoods presented in equations (2.34) and (2.43) for each model, respectively, the integrated likelihood[43] also referred as the observed-data likelihood $p(Y|\theta)$ is used, where the latent states have been integrated out.

An approximation of $p(Y|\theta)$ can be estimated as a by-product of the particle filtering algorithm[44]. In more detail, this is provided in Appendix B.1.2 Particle filtering – equation (B.3).

For the autoregressive coefficients following multivariate normal priors, the innovation errors following inverse gamma priors, the shape parameter following gamma prior, and the threshold following a truncated normal distribution, where applicable, the estimation of the posterior is straightforward. For the delay parameter, which follows a uniform prior, its contribution to the posterior is always $1/P$ at any value of the delay parameter.

Dropping index $\mathcal{M}$, and the superscript (r) denoting a retained draw of the parameters' vector θ at an MCMC iteration, the log posterior of each model is explicitly estimated as below:

1. TAR model:

$$\ln p\left(\theta\middle|Y\right)=\ln p\left(Y\middle|\theta\right)+\sum_{i=1}^{2}\ln p_{\mathcal{N}}\left(\beta_i\middle|\underline{\mathrm{M}}_{\beta i},\underline{\Omega}_{\beta i}\right)+\sum_{i=1}^{2}\ln p_{\mathcal{IG}}\left(\sigma_i^2\middle|\underline{S}_{\sigma_i^2},\underline{d}_{\sigma_i^2}\right)$$
$$+\ln p\left(d_p\right)+\ln p_{\mathcal{N}_{tr}}\left(c\middle|\underline{\mathrm{M}}_c,\underline{\Omega}_c,p_L,p_H\right)$$

Where $\theta=\left\{\beta^1,\sigma_1^2,\beta^2,\sigma_2^2,d,c\right\}$.

2. LSTAR model

$$\ln p\left(\theta\middle|Y\right)=\ln p\left(Y\middle|\theta\right)+\sum_{i=1}^{2}\ln p_{\mathcal{N}}\left(\beta_i\middle|\underline{\mathrm{M}}_{\beta i},\underline{\Omega}_{\beta i}\right)+\ln p_{\mathcal{IG}}\left(\sigma^2\middle|\underline{S}_{\sigma^2},\underline{d}_{\sigma^2}\right)$$
$$+\ln p\left(d_p\right)+\ln p_{\mathcal{N}_{tr}}\left(c\middle|\underline{\mathrm{M}}_c,\underline{\Omega}_c,p_L,p_H\right)+\ln p_{\mathcal{G}}\left(\gamma\middle|\underline{\alpha}_\gamma,\underline{\beta}_\gamma\right)$$

Where $\theta=\left\{\beta^1,\beta^2,\sigma^2,d,\gamma,c\right\}$.

3. TV-LSTAR model:

$$\ln \hat{p}\left(\theta\middle|Y\right)=\ln \hat{p}\left(Y\middle|\theta\right)+\sum_{i=1}^{2}\ln p_{\mathcal{N}}\left(\beta_i\middle|\underline{\mathrm{M}}_{\beta i},\underline{\Omega}_{\beta i}\right)+\ln p_{\mathcal{IG}}\left(R\middle|\underline{S}_R,\underline{d}_R\right)$$
$$+\ln p_{\mathcal{N}}\left(\rho\middle|\underline{\mathrm{M}}_\rho,\underline{\Omega}_\rho\right)+\sum_{j=1}^{2}\ln p_{\mathcal{IG}}\left(Q_{jj}\middle|\underline{S}_{Q_{jj}},\underline{d}_{Q_{jj}}\right)+\ln p\left(d_p\right)$$
$$+\ln p_{\mathcal{G}}\left(\gamma\middle|\underline{\alpha}_\gamma,\underline{\beta}_\gamma\right)$$

Where $\theta=\left\{\beta_1,\beta_2,\rho,Q,d,R,\gamma\right\}$.

[43] It holds $p(Y\mid\theta)=\int p(Y,S\mid\theta)\,dS=\int p(Y\mid \mathrm{S},\theta)\,p(S\mid\theta)dS$, where $p(Y,S\mid\theta)$ is the complete data likelihood and $p(Y\mid \mathrm{S},\theta)$ is the conditional likelihood.

[44] In the context of state-space models this likelihood is also referred as a marginal likelihood denoting that the latent states have been marginalised out. This marginal likelihood should not be confused with the model's total marginal likelihood used for the purpose of model comparison in a Bayesian setting.

4. TV-LSTAR-SV model:

$$\ln \hat{p}\left(\theta|Y\right) = \ln \hat{p}\left(Y|\theta\right) + \sum_{i=1}^{2} \ln p_{\mathcal{N}}\left(\beta_i|\underline{\mathrm{M}}_{\beta i}, \underline{\Omega}_{\beta i}\right) + \ln p_{\mathcal{N}}\left(\rho|\underline{\mathrm{M}}_{\rho}, \underline{\Omega}_{\rho}\right)$$
$$+ \sum_{j=1}^{2} \ln p_{\mathcal{IG}}\left(Q_{jj}|\underline{S}_{Q_{jj}}, \underline{d}_{Q_{jj}}\right) + \ln p\left(d_p\right) + \ln p_{\mathcal{IG}}\left(\Phi|\underline{S}_{\Phi}, \underline{d}_{\Phi}\right)$$
$$+ \ln p_{\mathcal{G}}\left(\gamma|\underline{\alpha}_{\gamma}, \underline{\beta}_{\gamma}\right)$$

Where $\theta = \{\beta_1, \beta_2, \rho, Q, d, \Phi, \gamma\}$.

2.3.2 Information criteria

Turning now to the estimation of the information criteria, for high dimensional models, the deviance information criterion (DIC) (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) is more robust and thus preferable to the classic Bayesian information criterion (BIC), (Schwarz, 1978) and the Akaike information criterion (AIC), Akaike (1974). Celeux, Forbes, Robert, and Titterington (2006) define DIC in more detail as:

$$DIC = -4E_{\theta}\left[\ln p\left(Y|\theta\right)\middle|Y\right] + 2\ln p\left(Y|\hat{\theta}\right)$$

Following Chan and Grant (2016), the first term is evaluated as the mean of the integrated likelihood at each of the MCMC retained draws of θ. For the second term, the quantity $\hat{\theta}$ is approximated as the draw (r) of parameters θ, which provides the highest value of the posterior $p\left(\theta|Y\right)$. With this value, $\hat{\theta}$ the term $\ln p\left(Y|\hat{\theta}\right)$ can be further easily computed. For the proposed TV-LSTAR and TV-LSTAR-SV models, the term $\ln p\left(Y|\hat{\theta}\right)$ is estimated using the standard particle filtering algorithm described in Appendix B.1.2 Particle filtering - equation (B.3).

2.4 Monte Carlo simulation

In order to evaluate to what extent the estimation procedure discussed analytically in Section 2.2.2 is efficient, the following Monte Carlo simulation is performed. A sample of $T = 500$ periods is created from the following data generating process (DGP), and afterwards, the true parameters and the latent states are compared to the estimated ones. The true DGP has the following values: $\beta_1 = [0.4, -0.2]$, $\beta_2 = [-0.5, 0.1]$, $d = 1$, $\gamma = 4$, $R = 0.01$, $Q = \text{diag}\ (0.01, 0.01)$. In order to satisfy that the logistic transition function $G\ (z_{t-d}, \gamma, c_t)$ alternates sufficient times between the two regimes throughout the sample; as a transition variable, the cosine function is used. The DGP is the following:

$$y_t = \beta_{c,t} + \left(1 - G\left(z_{t-1}, 4, c_t\right)\right)\left(0.4 y_{t-1} - 0.2 y_{t-2}\right)$$
$$+ G\left(z_{t-1}, 4, c_t\right)\left(-0.5 y_{t-1} + 0.1 y_{t-2}\right) + e_t, \quad e_t \sim \mathcal{N}\left(0, 0.01\right)$$

$$\begin{bmatrix} \beta_{c,t} \\ c_t \end{bmatrix} = \begin{bmatrix} \beta_{c,t-1} \\ c_{t-1} \end{bmatrix} + \begin{bmatrix} v_t^{\beta_c} \\ v_t^{c} \end{bmatrix}, \quad v_t \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}\right)$$

$$z_t = 5 \sin\left(0.05 t\right)$$

The priors for the respective parameters are set as follows. The mean of the prior distribution for coefficients β_1 and β_2 is set equal to zero for all elements, i.e. $\underline{\mathrm{M}}_{\beta_{1,2}} = 0$ and the hyperparameters λ_1 and λ_2 determining the coefficients' prior covariance $\underline{\Omega}_{\beta_{1,2}}$ are set equal to one and zero, respectively, resulting in the identity matrix. The autoregressive coefficients for the time-varying parameters do not follow a prior, but rather it is assumed a priori that both of the time-varying parameters, i.e. $\beta_{c,t}$ and c_t follow a random walk. The innovations variance R, as well as the time-varying parameters innovations variance $Q_{j,j}$ follow an inverse Gamma prior with a scale $\underline{S}_R = \underline{S}_{Q_{j,j}} = 0.01$ and degree of freedom $\underline{d}_R = \underline{d}_{Q_{j,j}} = 1$. The delay parameter d follows a discrete uniform distribution with probability $p(d=d_p) = 1/P = 1/2$. Finally, parameter γ, which follows a Gamma prior, has parameters $\underline{\alpha}_\gamma = 2$ and $\underline{\beta}_\gamma = 1$. In short, the Monte Carlo exercise prior specifications are summarised in the following table.

Table 2.1: Monte Carlo prior specifications

Parameter	Prior	
β_1, β_2	$\underline{\mathrm{M}}_{\beta_{1,2}} = 0$	$\underline{\Omega}_{\beta_{1,2}} = I_2$
R	$\underline{S}_R = 0.01$	$\underline{d}_R = 1$
Q	$\underline{S}_{Q_{j,j}} = 0.01$	$\underline{d}_{Q_{j,j}} = 1$
d	$p\left(d = d_P\right) = 1/2$	
γ	$\underline{\alpha}_\gamma = 2$	$\underline{\beta}_\gamma = 1$

Furthermore, the MCMC algorithm for each Monte Carlo replication is initialised with the following values:

Table 2.2: Particle MCMC initial values

β_1, β_2	AR(4) coefficients
R	AR(4) σ^2
Q	AR(4) σ^2
d	2
γ	10
$S_{1:T}$	0
S_0	0
Var(S_0)	0.01

The particle Gibbs sampler uses 10000 draws as a burn-in sample, and the subsequent 5000 retained draws are used for inference. The particle Gibbs with ancestor sampling uses $N = 200$ particles.

The following tables present the results for $N = 100$ Monte Carlo replications of the model's parameters estimation.

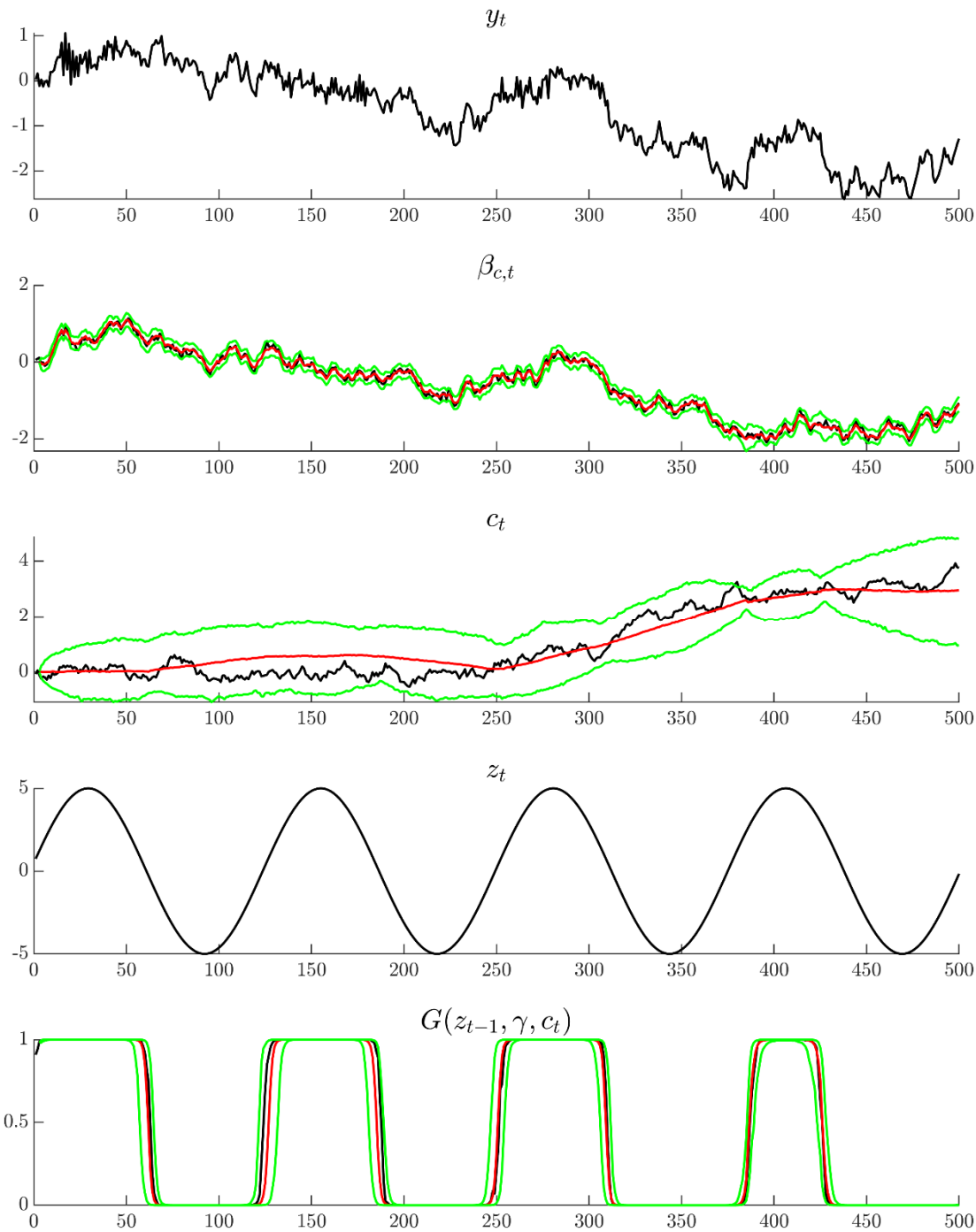
Table 2.3: Monte Carlo parameter estimates

Parameter	True	Mean	S.E	99% C.I.
β_1^1	0.4	0.421	0.010	[0.395, 0.447]
β_2^1	−0.2	−0.199	0.007	[−0.216, −0.182]
β_1^2	−0.5	−0.467	0.010	[−0.493, −0.441]
β_2^2	0.1	0.125	0.009	[0.102, 0.148]
R	0.01	0.011	0.000	[0.011, 0.012]
$Q_{1,1}$	0.01	0.010	0.000	[0.009, 0.011]
$Q_{2,2}$	0.01	0.011	0.001	[0.009, 0.013]
γ	4	3.798	0.100	[3.541, 4.055]
d	1	64%	−	−

Note. The delay parameter percentage shows the percentage of correctly specifying the true value of d over all Monte Carlo replications. Delay parameter d is correctly specified for each Monte Carlo replication when the majority of the retained draws have values equal to the true value of the delay parameter.

Figure 2.3 below shows a randomly selected realisation from the Monte Carlo replications using the true DGP mentioned above. The transition variable z_t shown in the fourth subpanel is periodic, and its impact on the transition function $G(z_{t-1}, \gamma, c_t)$ is shown in the last subpanel. In particular, the transition function $G(z_{t-1}, \gamma, c_t)$ fluctuates between zero and one, corresponding to the two respective regimes.

Figure 2.3: DGP and latent state estimates



Note. Black lines show the true values. Blue lines show the median estimates, and green lines cover the 99% credible intervals.

2.5 Empirical application

2.5.1 Data description

All model specifications are estimated with the following four variables at quarterly frequency[45]: real GDP growth, CPI inflation, Federal Funds Rates (FFR), and 10-year bond yield (10-Y). Variables are transformed into annualised quarter-on-quarter growth rates except for the FFR and 10-Y, which remain in percentage points. See Table 2.4 below.

Table 2.4: Variables

Variable	Transformation
Real GDP	$100\ ((y_t / y_{t-1})^4 - 1)$
CPI	$100\ ((y_t / y_{t-1})^4 - 1)$
FFR	-
10-year bond yield	-

Source: Archival Federal Reserve Economic Data of the Federal Reserve Bank of St. Louis.

The out-of-sample exercise is performed recursively using an expanding window since 1960Q1, with the first out-of-sample exercise performed for 1985Q1 and the last one for 2018Q4. For each of the out-of-sample exercises at period t, the last available data point is that of $t-1$. The data used are as they were available during the 2020Q1 vintage.

2.5.2 Priors and other estimation details

The two versions of the proposed TV-LSTAR model, with and without heteroscedasticity, are benchmarked versus an AR(4) model estimated by means of OLS. Furthermore, in order to gauge how the proposed models compare relevant to other standard non-linear time series models, the standard TAR and LSTAR models are included in the analysis as well.

All models, including the benchmark, are estimated with $P = 4$ lags. All Bayesian non-linear models (TAR, STAR, TV-LSTAR, TV-LSTAR-SV) have the following priors with respect to the autoregressive coefficients: $\lambda_1 = 1, \lambda_2 = 1$, whereas for the case of TAR and LSTAR models, the constants at each regime follow a diffuse prior. All model innovations follow a prior with a scale equal to the variance estimated by an AR(4) model, i.e. $\underline{S}_R = \sigma^2$ and 1 degree of freedom $\underline{d}_R = 1$. The shape coefficient γ prior has parameters $\underline{\alpha}_\gamma = 5$ and $\underline{\beta}_\gamma = 1$. The delay parameter d follows a discrete uniform distribution with probability $p(d=d_p) = 1/P = 1/4$.

[45] The variables whose original frequency is monthly are first transformed into quarterly frequency by calculating each quarter's respective average.

Furthermore, for the TAR and LSTAR models, the threshold follows a truncated normal prior at the 25th and 75th percentile of the empirical time series distribution. The threshold's prior variance is set equal to the variance of the time series itself.

With regard to the TV-LSTAR and the TV-LSTAR-SV models in more detail, the following priors are used. The innovations of $\beta_{c,t}$ and c_t follow a prior with scale $\underline{S}_{Q_{j,j}} = 0.1$ and $\underline{d}_{Q_{j,j}} = 1$. Regarding the time-varying parameter $\beta_{c,t}$, it is assumed a priori that it follows a random walk, and thus a prior over its autoregressive coefficients vector ρ_1 is not used. However, the autoregressive coefficients ρ_2 for the time-varying threshold c_t follow a prior with mean $\underline{\mathrm{M}}_{\rho_2} = [0,0]'$ and variance $\underline{\Omega}_{\rho_2} = diag(1,1)$. Any draws of ρ_2 with explosive roots are discarded. For the case of the TV-LSTAR-SV the innovations of the stochastic volatility follow a prior with $\underline{S}_{\Phi} = 0.0001$ and $\underline{d}_{\Phi} = 5$.

The TV-LSTAR and TV-LSTAR-SV models are initialised in the following manner. The autoregressive coefficients β_1 and β_2 are set equal to the mean of the posterior from the LSTAR estimates of the respective coefficients. In addition, the Particle Gibbs with Ancestor Sampling algorithm (Lindsten, Jordan, & Schon, 2014) is initialised with $\beta_{c,1:T}$ set equal to the mean of posterior draws (r) of the STAR model: $\left(1 - G_t^{(r)}\right)\beta_c^{1(r)} + G_t^{(r)}\beta_c^{2(r)}$ and $c_{1:T} = HP_{trend}$, where HP_{trend} denotes the trend of the Hodrick-Prescott filter (Hodrick & Prescott, 1997) applied to time series y_t. The initial values of the state-space model are set as follows. For the time-varying constant $\beta_{c,t}$, the relevant values are $S_{1,0} = \beta_{c,0} = \text{mean}(\beta_{c,1:T})$, and $P_{1,1,0} = 1$, whereas for the time-varying threshold c_t, the values are $S_{2,0} = c_0 = \text{mean}(Y)$, and $P_{2,2,0} = 1$.

Regarding the initialisation values with respect to the stochastic volatility for the TV-LSTAR-SV for the first iteration of the MCMC algorithm, these are R_t for $t = 1,\cdots,T$ is set equal to σ^2 estimated with AR(P), $\ln h_t$ for $t = 1,\cdots,T$ is set equal to $\ln(0.8\ diag(\text{var}(Y)))$ and for the Kim, Shephard, and Chib (1998) algorithm the initial conditions are set as diffuse: $\ln h_0 = 0$, $\text{var}(\ln h_0) = 10$ and $\Phi = 0.01$.

The particle Gibbs sampler uses a burn-in sample of 10000 draws for convergence, and the subsequent 5000 draws are retained for inference and forecasting purposes.

2.5.3 In-sample results

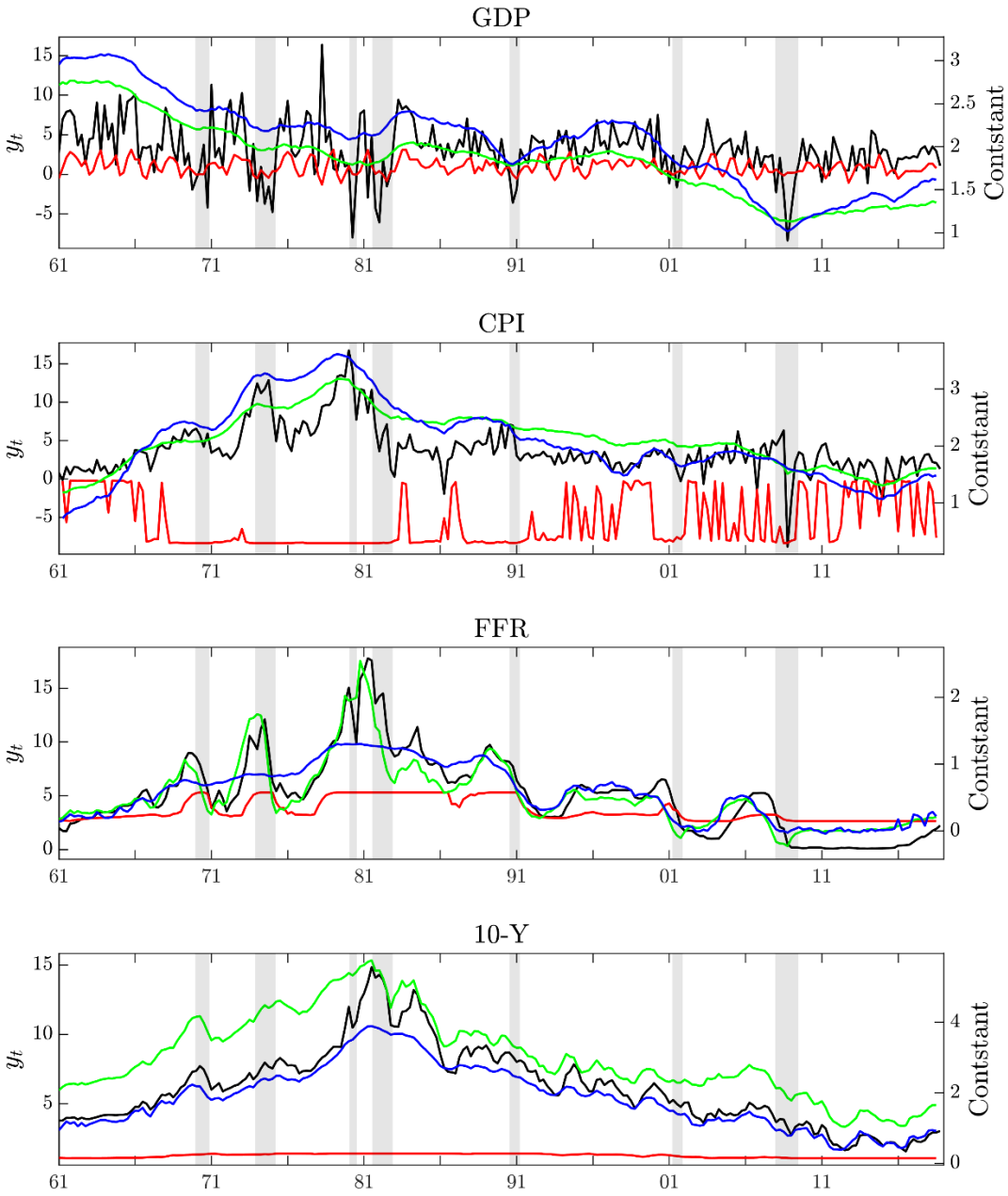
This section shows the in-sample results with respect to the estimated models. Firstly, the posterior estimates of the time-varying parameters and the stochastic volatility estimates of the proposed models are shown in order to understand how they capture the underlying dynamics. In continuation, the estimates of the log marginal likelihood and the deviance information criterion are presented.

2.5.3.1 Time-varying parameters – Transition function

Figure 2.4 below shows on the right axis the estimates of the time-varying constants for the TV-LSTAR and TV-LSTAR-SV models with green and blue lines, respectively, in comparison with the bounded combined constant of the LSTAR model in red for comparison purposes. The left axis refers to the time series itself, which is also shown for reference in the same sub-figure for reference purposes.

Overall, it is observed that the unbounded time-varying constants of the TV-LSTAR and TV-LSTAR-SV follow closely and proportionally the change in levels of the time series, especially for the case of CPI, FFR and the 10-year bond yield. However, for the case of the GDP, the time-varying constant overall present a declining trend from the sample's start in the 60's up until the 2008 Great Recession, where this trend reverses. What stands out is the difference between the TV-LSTAR and TV-LSTAR-SV time-varying constants for FFR in the periods in the 1970-75 period and the late 70s-early 80s period, where the time series excessive volatility in the TV-LSTAR model is attributed to the time-varying constant, whereas is in the TV-LSTAR-SV model is attributed to the stochastic volatility as it is shown later in Figure 2.10.

Figure 2.4: Autoregressive constant estimates



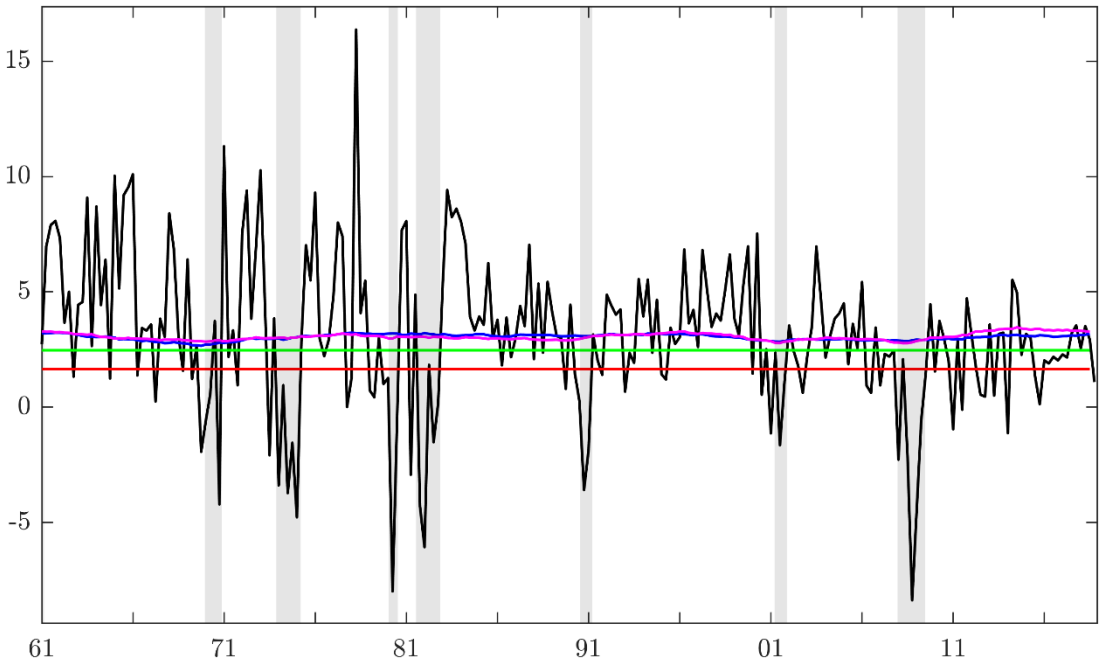
Note. Left axis: Black lines show the time series for each variable. Right axis: Red, green and blue lines show the (implied) constant posterior median $(1\text{-}G)\beta_{1,c}+G\beta_{2,c}$ for LSTAR, and $\beta_{c,t}$ for TV-LSTAR and TV-LSTAR-SV respectively for 2018Q4 vintage.

Figure 2.5 to Figure 2.8 below show the estimates of the time-varying thresholds. For GDP, there is no evidence of almost any time variation of the threshold, but this should be examined jointly with the aforementioned consistent time variation of the time-varying constant.

The time-varying threshold of the CPI behaves rather counter-intuitively during the 70s high inflation period, where both for the homoscedastic and heteroscedastic models, the time-varying threshold is below the constant threshold LSTAR. However, during the Great Recession (2007-2009), the local maximum of the threshold is estimated with a time delay during the lowest v-shaped recession, which can be explained by the fact that the delay parameter d in y_{t-d} estimates for the given variable have a mean value of 3 (with a maximum of 4 lags). After this local maximum, in the period following the Great Recession, the threshold both for the homoscedastic and heteroscedastic models presents a weak declining trend in line with the underlying CPI time series itself.

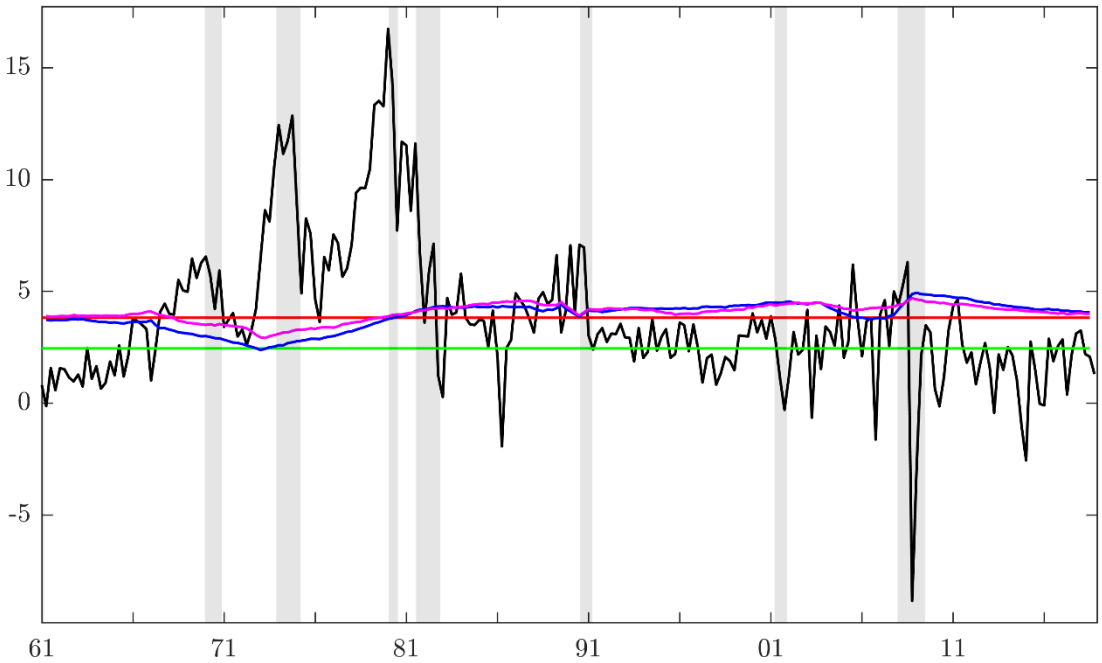
For the case of the FFR the time variation of the threshold is, as expected, more greatly pronounced, although with big differences between the homoscedastic and the heteroscedastic variants of the models. For example, during the early 80s period, when the FFR peaked with respect to the entire time series, the threshold of the homoscedastic model in blue is far above the heteroscedastic one in magenta. In that case, the excessive volatility of the time series captured in the stochastic volatility estimates in Figure 2.10 allows the threshold to move lower and rather let the increased innovations variance explain the behaviour of the time series. During the 90s, the homoscedastic model's threshold follows a declining trend which is consistent with the decreasing FFR rates. Finally, 10-year bond yield time-varying thresholds resemble the behaviour of the FFR during the late 70s-early 80s period. However, for the post-2000s period, contrary to FFR, the threshold both for the homoscedastic and heteroscedastic models, rather than following a declining trend, exhibit a strong mean reversion.

Figure 2.5: GDP threshold estimates



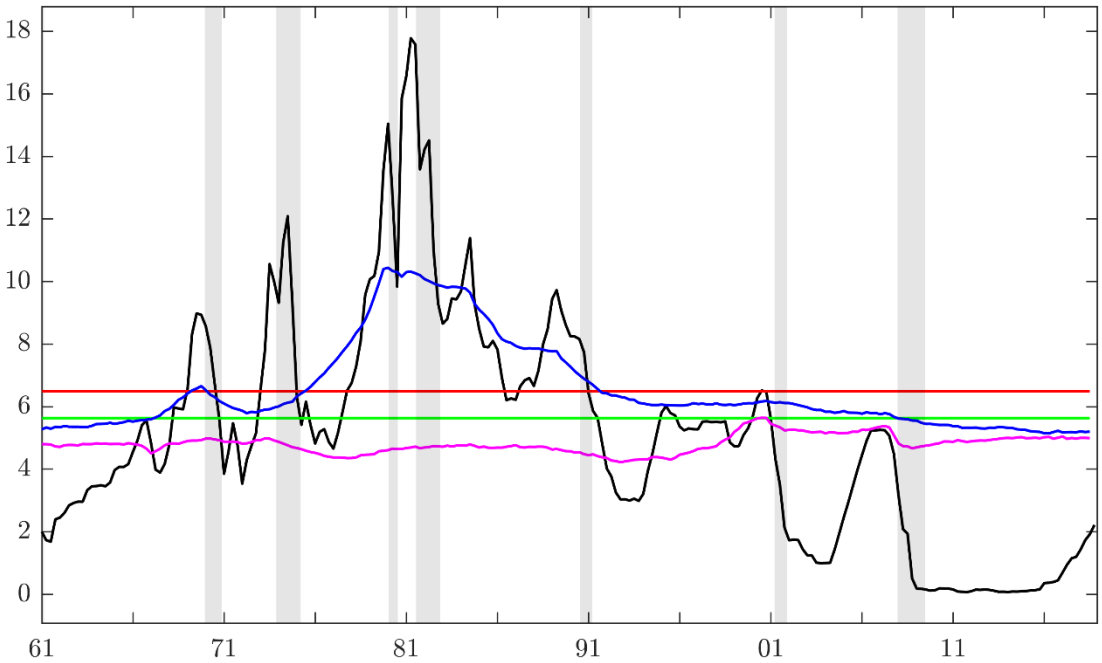
Note. Red, green, blue and magenta lines show the threshold posterior median for TAR, LSTAR, TV-LSTAR and TV-LSTAR-SV, respectively for 2018Q4 vintage. Black lines show the time series.

Figure 2.6: CPI threshold estimates



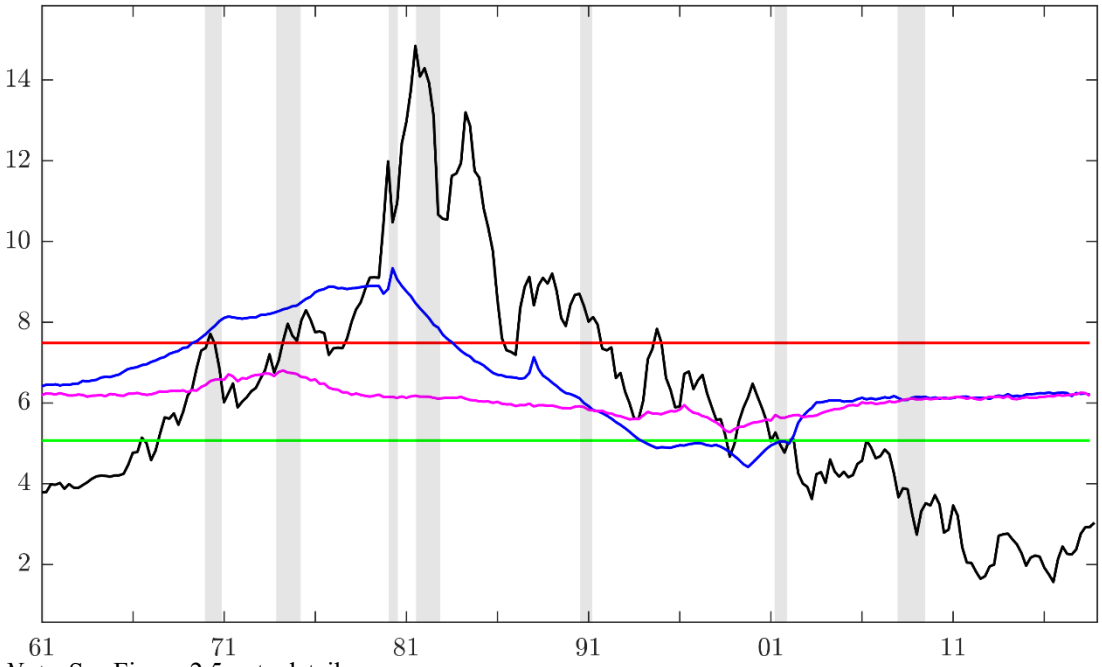
Note. See Figure 2.5 note details.

Figure 2.7: FFR threshold estimates



Note. See Figure 2.5 note details.

Figure 2.8: 10-year bond yield threshold estimates

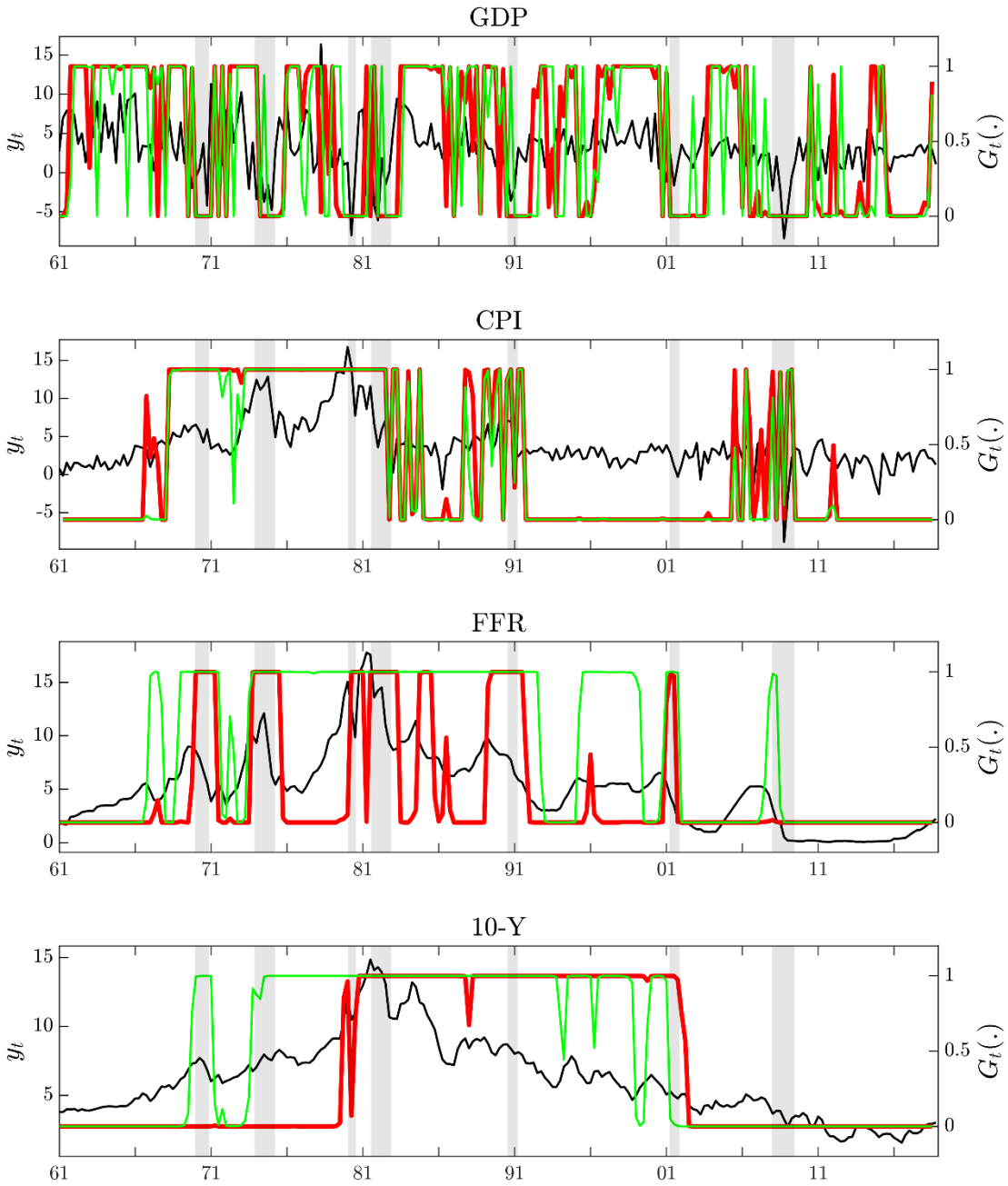


Note. See Figure 2.5 note details.

Figure 2.9 below shows the posterior median of the transition function for the homoscedastic and the heteroscedastic model with red and green lines, respectively. Overall, for GDP, the values are very close and in line with the estimates of the time-varying threshold, which almost coincide, as shown in Figure 2.5 above. For CPI, overall, both versions of the model fall in regime one during the 70s high inflation period and in regime two in the subsequent low inflation environment, albeit with short periods of very frequent alternations during 1985-1990 and 2005-2010.

For the FFR, as already explained previously, the accentuated role of stochastic volatility in defining the threshold is also reflected in the estimates of the transition function, especially during the 80s and the periods preceding the 2001 dot-com bubble crash and the 2008 Great Recession. The homoscedastic model's threshold, which follows the time series behaviour closely, results in more frequent alternations between the two regimes during the local maxima of the time times series in contrast to the heteroscedastic models, which follow fewer alternations. Finally, for the 10-year bond yield, both models broadly are capturing correctly two regimes for the elevated rates above the long run mean and below, respectively.

Figure 2.9: Transition function $G(y_{t-d}, \gamma, c_t)$ estimates

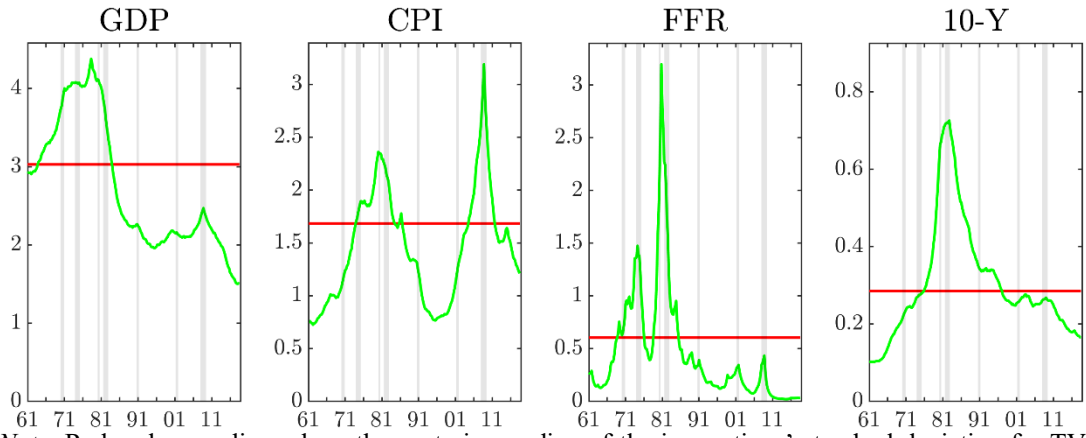


Note. Left axis: Black lines show the time series for each variable. Right axis: Red and green lines show the transition function posterior median for TV-LSTAR and TV-LSTAR-SV, respectively for 2018Q4 vintage.

2.5.3.2 Stochastic volatility estimates

The estimates of the stochastic volatility capture the changing dynamics of the proposed time series model innovations variances over different sub-periods. For example, the GDP stochastic volatility is increased during the period preceding the Great Moderation and then far decreased except for the Great Recession, where it peaked again, although at a level lower than the highs during the volatile 70s-80s. Similarly, CPI's stochastic volatility peaks around the late 70s - early 80s and captures the big shock of the 2008 Great Recession correctly. For the case of the FFR, stochastic volatility reaches excessive levels during the late 70s oil crisis, and afterwards, it increases again during the Great Recession. Finally, 10-year bond yield volatility estimates increase around the late 70s - early 80s and decrease subsequently in a progressive way during the Great Moderation.

Figure 2.10: Stochastic volatility estimates



Note. Red and green lines show the posterior median of the innovations' standard deviation for TV-LSTAR and TV-LSTAR-SV, respectively for 2018Q4 vintage.

2.5.3.3 Model comparison

In order to indicate how the proposed models (TV-LSTAR, TV-LSTAR-SV) compare against the other non-linear time series models (TAR, LSTAR) in-sample, as already discussed in Section 2.2.5, the marginal likelihood and the deviance information criterion are used. According to Table 2.5 below, the TV-LSTAR-SV has the largest marginal likelihood across all non-linear models and variables. Furthermore, comparing the LSTAR and TV-LSTAR models, which differ only in the time variation allowed for the constant and the threshold, it is evident that the proposed TV-LSTAR model presents the largest marginal likelihood.

Table 2.5: Log marginal likelihood

	GDP	**CPI**	**FFR**	**10-Y**
TAR	-604.93	-440.88	-200.50	-120.47
LSTAR	-601.29	-464.64	-287.53	-147.66
TV-LSTAR	-587.48	-441.57	-227.89	-112.22
TV-LSTAR-SV	-125.33	-283.77	-47.39	-15.82

Note. Log marginal likelihood for the 1960-2018 period (2018Q4 vintage).

Turning now to the deviance information criterion values in the following table, the same pattern holds as discussed previously, i.e. the TV-LSTAR-SV presents the lowest values among all models, while between the LSTAR and TV-LSTAR models, the latter outperforms the former.

Table 2.6: Deviance information criterion (DIC)

	GDP	CPI	FFR	10-Y
TAR	1173.46	902.59	416.34	252.07
LSTAR	1189.78	941.27	590.08	302.82
TV-LSTAR	1186.90	924.63	455.43	263.46
TV-LSTAR-SV	1119.16	877.55	222.05	228.45

Note. Deviance information criterion for the 1960-2018 period (2018Q4 vintage).

2.5.4 Out-of-sample results – Forecast evaluation metrics

Point forecasts for horizon h are evaluated according to the root mean squared error (RMSE):

$$\text{RMSE}_h = \left[\frac{1}{T-h-t_0-1} \sum_{t=t_0}^{T-h} \left(y_{t+h} - \hat{y}_{t+h|t} \right)^2 \right]^{1/2}$$

Where t_0 refers to 1985Q4 and T refers to 2018Q4.

Density forecasts are evaluated according to the continuous ranked probability score (CRPS). Following, Gneiting and Raftery (2007), the CPRS can be evaluated in a closed form as:

$$\text{CRPS}(F, y) = \frac{1}{2} E_F \left| \hat{y} - \hat{y}' \right| - E_F \left| y - \hat{y} \right|$$

Where $\hat{y}$ are independent draws from the predictive density with cumulative distribution F, and y is the outcome.

The exact formula[46] for the retained draws R of the MCMC algorithm at horizon h is given as:

$$\text{CRPS}_h = \frac{1}{T-h-t_0-1} \sum_{t=t_0}^{T-h} \left[\frac{1}{R} \sum_{r=1}^{R} \left| y_{t+h} - \hat{y}_{t+h|t}^{(r)} \right| - \frac{1}{2R^2} \sum_{r=1}^{R} \sum_{r'=1}^{R} \left| \hat{y}_{t+h|t}^{(r)} - \hat{y}_{t+h|t}^{(r')} \right| \right]$$

In order to provide a rough measure of whether the differences in the forecast accuracy are significant, the Diebold and Mariano (1995) test is employed. The hypothesis testing is always one-sided, i.e. rejection of the null of equal performance versus the benchmark model provides evidence of the proposed model outperforming the benchmark. The test applies in the same manner both for point forecasts (MSE) and density forecasts[47].

[46] More information can be found in Dieppe, Legrand, and Van Roye (2016).

[47] The results of Diebold-Mariano tests for density forecasts should be interpreted with care since this type of statistical testing applies to rolling estimation schemes and not to recursive re-estimation of the models under an expanding window (Amisano & Giacomini, 2007).

2.5.4.1 Point forecasts

The following Table 2.7 presents the out-of-sample results in terms of point forecast accuracy evaluated with RMSE ratios. What stands out clearly is that the proposed model, both for the homoscedastic and the heteroscedastic specification performs more favourably against the linear benchmark model and consistently across all variables (27 out of 32 cases in total), with few exceptions for the 10-year bond yield for forecasts up to 4-quarters ahead.

In more detail, for GDP, the TV-LSTAR and TV-LSTAR-SV models outperform both the linear and the other non-linear models of the existing literature. The inclusion of stochastic volatility does not help to improve the forecast accuracy any further. Given that as shown previously, the TV-LSTAR models' thresholds are very close to TAR and STAR models and do not exhibit any time variation, one could presumably argue that the gains in forecast accuracy can be attributed exclusively to the time-varying constant rather than the assumption of the non-linearity of the time series.

Turning to CPI, the TV-LSTAR and TV-LSTAR-SV models outperform the benchmark significantly in seven out of the eight cases in total but present only marginal improvements over the existing TSTAR and LSTAR models. In this case, stochastic volatility improves the results versus the homoscedastic TV-LSTAR model for forecasts up to 8-quarters ahead.

Regarding the FFR, TV-LSTAR and TV-LSTAR-SV models present the biggest gains at all horizons both with respect to the benchmark AR model, as well as against the TAR and LSTAR models. The superior forecasting performance is found to be significant for four out of eight combined comparisons of the TV-LSTAR and TV-LSTAR-SV models at all horizons. Stochastic volatility further improves the results though only for the 1-quarter ahead forecasts.

Finally, turning to the 10-year bond yield, the proposed models underperform the benchmark for forecasts up to 4-quarters ahead, albeit they present improvements over the existing TAR and LSTAR models. For horizons of 8 and 12-quarters ahead, both specifications of the proposed models outperform the linear benchmark, with the differences being significant for the 12-quarters ahead forecasts.

Table 2.7: Non-linear models versus benchmark – point forecast evaluation

	GDP	CPI	FFR	10-Y
			AR	
$h = 1$	2.161	2.081	0.424	0.391
$h = 4$	2.421	2.352	1.486	1.034
$h = 8$	2.515	2.381	2.485	1.401
$h = 12$	2.541	2.522	3.103	1.661
			TAR	
$h = 1$	1.022	0.953	0.911**	1.184
$h = 4$	1.004	0.971**	0.937**	1.207
$h = 8$	0.989	1.016	0.962*	1.135
$h = 12$	0.995	1.020	0.965*	1.073
			LSTAR	
$h = 1$	1.000	0.948*	0.982	1.185
$h = 4$	1.016	0.972*	0.994	1.253
$h = 8$	0.995	1.018	1.017	1.177
$h = 12$	1.009	1.050	1.013	1.113
			TV-LSTAR	
$h = 1$	0.988	0.949*	0.874***	1.037
$h = 4$	0.982	0.947**	0.857**	1.042
$h = 8$	0.966	0.925*	0.883	0.941
$h = 12$	0.985	0.905*	0.901	0.792*
			TV-LSTAR-SV	
$h = 1$	0.982	0.940*	0.813***	1.050
$h = 4$	0.984	0.941***	0.864*	1.035
$h = 8$	0.984	0.920*	0.909	0.945
$h = 12$	1.002	0.928	0.899	0.805**

Note. First sub-panel shows benchmark's raw RMSEs. Remaining sub-panels show RMSE ratios of the non-linear models with respect to the benchmark. Stars refer to the p-values of the Diebold and Mariano (1995) one-sided test with respect to the benchmark. *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively. The evaluation period is 1985-2018.

2.5.4.2 Density forecasts

Turning to the evaluation of the density forecasts using CRPS, it is evident that with few exceptions, all of the remaining results show the improved accuracy of the TV-LSTAR and TV-LSTAR-SV models against the linear benchmark in 25 out of 32 cases in total, whereas for 14 cases the differences are found to be significant.

Moreover, whereas for GDP, only the TV-LSTAR-SV presents gains consistently across all horizons, for the case of CPI, both specifications of the model outperform the linear benchmark and the existing TAR and LSTAR models, especially at the 12-quarters ahead forecasts. For the same variables, the inclusion of stochastic volatility does not improve density forecasts.

Similar to the point forecast results, the improvement in terms of density forecast accuracy becomes more noticeable for the case of the FFR and 10-year bond yield and especially with the inclusion of stochastic volatility, the CRPS ratio drops down to 0.60 at the 1-quarter ahead forecasts for FFR.

Table 2.8: Non-linear models versus benchmark – density forecast evaluation

	GDP	**CPI**	**FFR**	**10-Y**
			AR	
$h = 1$	1.290	1.008	0.299	0.222
$h = 4$	1.407	1.229	0.871	0.584
$h = 8$	1.446	1.336	1.426	0.806
$h = 12$	1.457	1.397	1.793	0.969
			TAR	
$h = 1$	1.008	0.935**	0.709***	1.093
$h = 4$	1.015	0.905***	0.887***	1.119
$h = 8$	1.008	0.910***	0.944***	1.079
$h = 12$	1.014	0.907***	0.954***	1.033
			LSTAR	
$h = 1$	1.004	0.944**	1.002	1.152
$h = 4$	1.020	0.950**	1.038	1.175
$h = 8$	1.005	0.972	1.062	1.149
$h = 12$	1.018	1.009	1.055	1.112
			TV-LSTAR	
$h = 1$	1.002	0.946*	0.846***	0.993
$h = 4$	1.005	0.923**	0.848***	1.018
$h = 8$	0.996	0.889**	0.907	0.943
$h = 12$	1.009	0.826**	0.934	0.797*
			TV-LSTAR-SV	
$h = 1$	0.969*	0.953	0.602***	1.031
$h = 4$	0.981	0.934**	0.854*	1.009
$h = 8$	0.988	0.908*	0.994	0.944
$h = 12$	0.995	0.871**	1.003	0.813**

Note. First sub-panel shows benchmark's raw CRPS. Remaining sub-panels show CRPS ratios of the non-linear models with respect to the benchmark. Stars refer to the p-values of the Diebold and Mariano (1995) one-sided test with respect to the benchmark. *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively. The evaluation period is 1985-2018.

2.5.4.3 Time variation of results

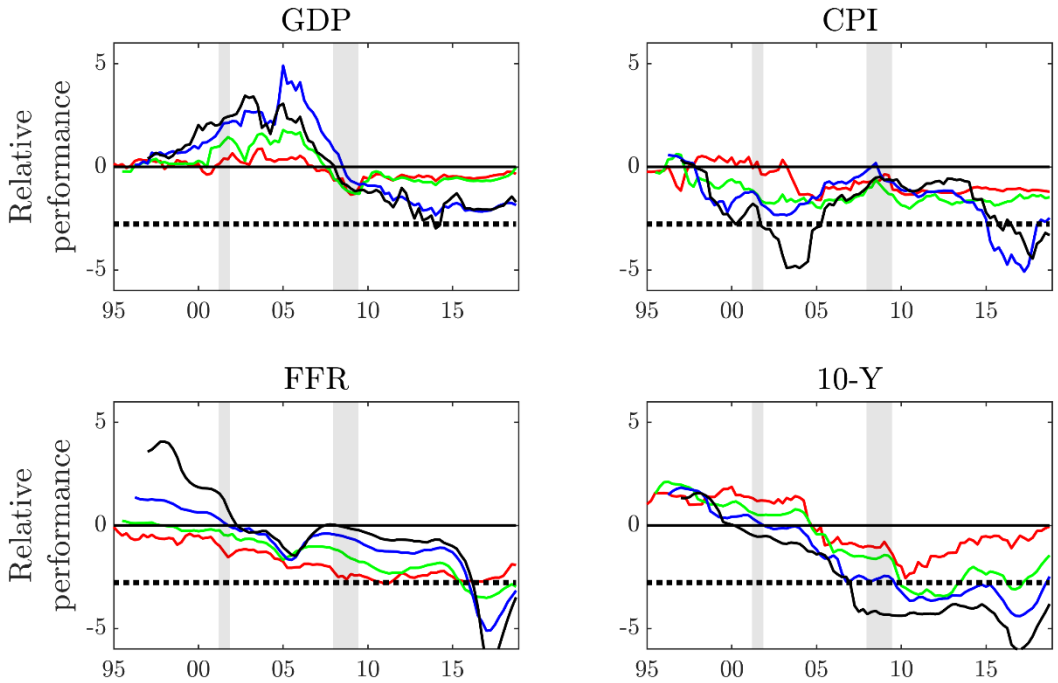
In order to evaluate the time variation of the results and to what extent the proposed models outperform linear models over different sub-periods, the Giacomini and Rossi (2010) fluctuation test is employed for the point forecasts in Figure 2.11 and Figure 2.12. The one-side fluctuation test is performed with a rolling window of 40 observations out of the total 136. Critical values can be found in Giacomini and Rossi (2010), Table 2. With respect to the density forecasts, the models are compared vis-a-vis the linear benchmark model by plotting the difference of the cumulative CRPS of the non-linear model minus the benchmark model in Figure 2.13 and Figure 2.14.

Regarding the point forecasts, the TV-LSTAR model outperforms the benchmark model in the post-Great Recession period and, in particular, for CPI, FFR, and 10-year bond yield, this difference becomes significant towards the end. For the period before the Great Recession, the forecast accuracy of the TV-LSTAR for CPI is very close to the benchmark AR model. However, the inclusion of stochastic volatility has the reverse effect on GDP and 10-year bond yields (TV-LSTAR-SV deteriorates further compared to the TV-LSTAR), whereas, for the FFR, the TV-LSTAR-SV continues to outperform the linear benchmark at the 8 and 12-quarters ahead forecasts especially after the end of the Great Recession.

Turning to the time variation of the density forecasts, the results follow a similar pattern. Overall the results improve after the Great Recession in favour of the TV-LSTAR and TV-LSTAR-SV models, except for the GDP in the TV-LSTAR model, where the gains result in equal performance with the AR model towards the end of the sample. The accuracy of the 8 and 12-quarters ahead forecasts remains equal to or better than the AR model consistently across all variables and the two variants of the proposed model, except for the TV-LSTAR GDP results.

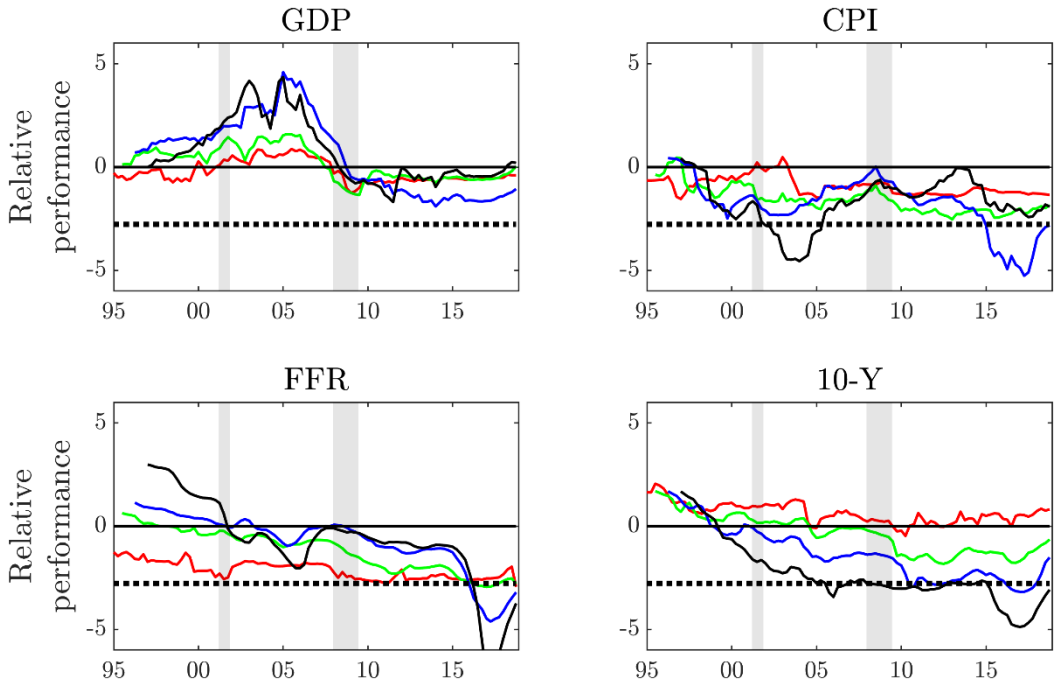
Due to the fact that in general, both for point and density forecasts, the results improve towards the end of the sample, which also happens to be after the Great Recession, someone could be sceptical on whether this behaviour can be rather attributed to the largest sample size that results in lower uncertainty for the estimated parameters or indeed it is an intrinsic ability of the models which can capture structural breaks. A similar analysis conducted employing a rolling window rather than the recursive one used for the current out-of-sample forecasting exercise could potentially shed light on this conundrum since the effect of the estimation sample could be controlled by using the same number of observations for the model estimation at each out-of-sample forecasting exercise.

Figure 2.11: Fluctuation test of TV-LSTAR minus the benchmark



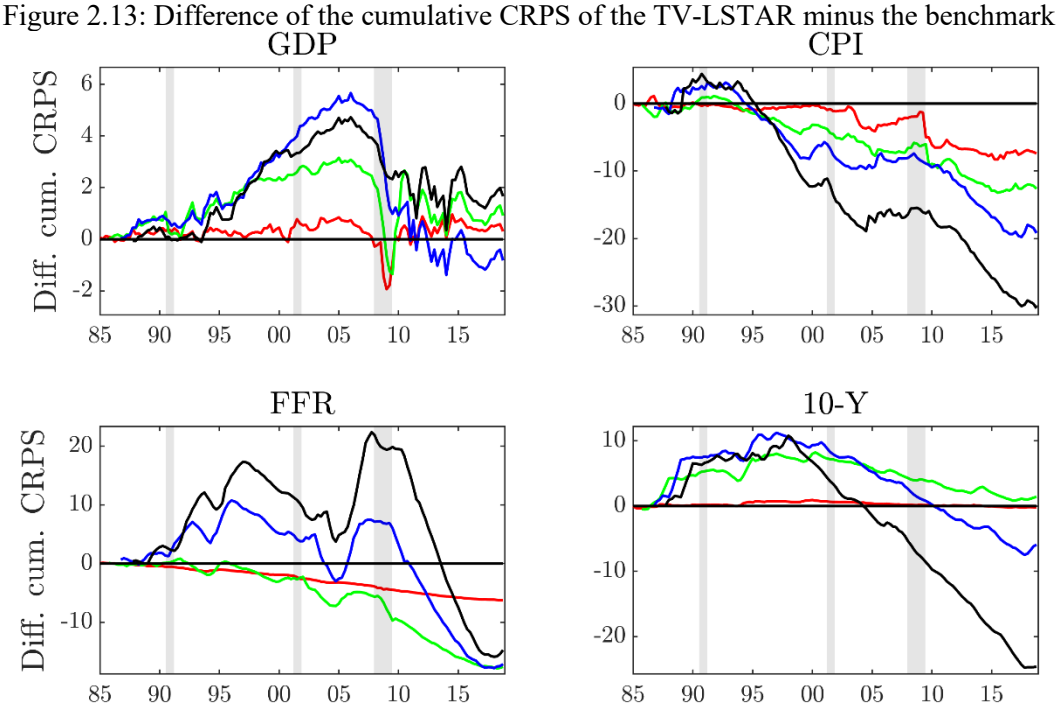
Note. Giacomini and Rossi (2010) one-sided fluctuation test. Red, green, blue and black lines refer to 1, 4, 8, and 12-quarters ahead forecasts, respectively. Dash lines show critical values at 5% significance level. Window size is 40 quarters.

Figure 2.12: Fluctuation test of TV-LSTAR-SV minus the benchmark



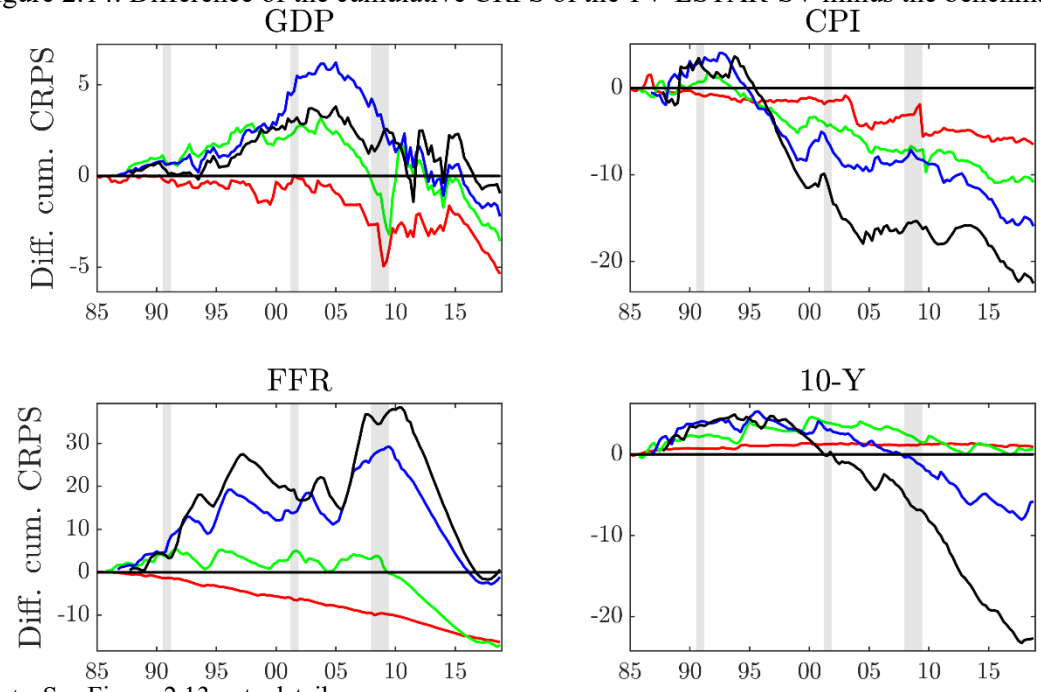
Note. See Figure 2.11 note details.

Figure 2.13: Difference of the cumulative CRPS of the TV-LSTAR minus the benchmark



Note. Red, green, blue and black lines refer to 1, 4, 8, and 12-quarters ahead forecasts, respectively.

Figure 2.14: Difference of the cumulative CRPS of the TV-LSTAR-SV minus the benchmark



Note. See Figure 2.13 note details.

2.6 Conclusion

The literature on non-linear time series models, non-linear in the sense of exhibiting different dynamics above and below a threshold, is not unanimous in terms of whether this specification improves or not forecasting accuracy. However, plenty of papers have shown that allowing for time variation in otherwise linear models can improve forecast accuracy substantially.

In this essay, by proposing an extension of the standard logistic smooth autoregressive model, which allows for time variation in both the constant and the threshold, I have tried to reconcile the aforementioned findings of the literature. Indeed, in the empirical application part, the proposed model applied to US data for GDP, CPI, FFR and 10-year bond yield shows clear advantages and can outperform not only the linear benchmark model but the existing non-linear models in many instances over the 1985-2018 out-of-sample evaluation period. Following the Great Recession, the time variation of the results shows stronger evidence in favour of the proposed model, and in many instances, the difference versus the linear benchmark model is found to be significant. The inclusion of stochastic volatility does not improve the forecasting performance except for the FFR at the 1-quarter ahead forecast horizon.

I leave for further research to investigate the effect of the sample size on the parameters' uncertainty and in relation to forecast accuracy, as well as to try to disentangle to what extent the improved forecasting accuracy is a result of the time variation of the constant or the threshold and how this contribution changes across different variables.

Finally, the extension of the model into the multivariate setting is another promising avenue for future research. Applying time-varying thresholds and constants in (S)TVARs can be investigated, for instance, by implementing a single transition function across all variables or instead using separate time-varying thresholds separately in each variable's transition function. In the former case, the selection of the transition variable from all the possible variables of a VAR via a suitable sampling algorithm could provide additional advantages for big VARs, where also a global-local shrinkage prior could be used on the VAR autoregressive coefficients at each regime.

‘It is difficult to predict, especially the future.’

Credited to
Niels Bohr (1973)

Essay 3

3. Wavelet-based single and mixed-frequency multivariate time series modelling and forecasting

3.1 Introduction

Time series econometrics typically examine the evolution of economic phenomena in the time domain. However, frequency characteristics of macroeconomic and financial time series are well acknowledged and have been studied extensively in the literature on business and financial cycle analysis by Strohsal, Proaño, and Wolters (2019), Beaudry, Galizia, and Portier (2020), and monetary policy by Assenmacher-Wesche and Gerlach (2007, 2008). In addition, the time series unobserved components modelling approach, i.e. trend-cycle extraction, for example, for GDP output gap estimation, relies on frequency-domain analysis via the use of band-pass filters – see Baxter and King (1999) and Christiano and Fitzgerald (2003).

Standard frequency-domain analysis is formally performed with Fourier analysis which via the power spectrum, decomposes a time series into sinusoidal components of different intensities and frequencies throughout the spectrum. Nonetheless, in the presence of time series irregularities such as structural breaks, regime swifts, or abrupt changes like big shocks or spikes, Fourier analysis, due to its global representation, cannot locate these phenomena anymore in the time dimension but only describes their frequency characteristics and intensity over the full sample.

Additionally, although the direct outcome of the Fourier transform, i.e. a complex-valued number, entails phase spectrum which incorporates time information, it is rather difficult to translate it in a meaningful way for economic quantities. Consequently, relying on the Fourier analysis power spectrum can only provide us with information about which frequencies are included in a time series and at which magnitude or intensity. This major drawback of frequency-domain analysis, i.e. the inability to localise spectral density in time, for example, does not allow to differentiate whether a frequency component appeared only during a sub-sample of the time series or continuously throughout the entire time series.

These shortcomings can be partially relieved via the short-time Fourier transform (Gabor, 1946), which applies Fourier transform on a window which slides over the entire time series and allows it to provide the changing frequency characteristics in the time dimension. However, the length of the sliding window establishes a priori a trade-off between the maximum time resolution and frequency resolution, which can be obtained and thus does not allow the representation of the entire spectrum contained in the time series at the finest possible time resolution. For instance, a small window (big time resolution-small frequency resolution) can localise in time the high-frequency components but will fail to capture lower frequency components. Reversely a large window (small time resolution-big frequency resolution) will capture a wider range of the spectrum but will fail to identify the various frequency components precisely in time.

Wavelet analysis, also known as time-frequency analysis, poses a perfect alternative to short-time Fourier analysis and allows the systematic feature extraction from time series without any compromise neither to the frequency nor the time resolution. The key idea is that sines and cosines used in the Fourier transform, which span the entire time axis from minus to plus infinity, are substituted from a short wavelet basis function, i.e. a short integrable function which can be shifted (translated) in time to allow time localisation and scaled (dilated) in frequency such that it captures time series various frequency characteristics. In contrast to the short-time Fourier transform, wavelet transform at a given point in time, i.e. after shifting the wavelet basis suitably, by scaling the wavelet basis appropriately, will capture the entire frequency spectrum from the highest to the lowest frequencies locally without losing any information in time or frequency.

The most common wavelet transforms[48] include: a) the Continuous Wavelet Transform (CWT), where the time series can be analysed over all possible time and frequency resolutions, b) the Discrete Wavelet Transform (DWT), which is a non-redundant (orthogonal) transformation that analyses time series into subcomponents of decreasing dyadic length where each one describes lower dyadic band-limited frequency characteristics, and c) the Maximal Overlap Discrete Wavelet Transform (MODWT) which contrary to the DWT is highly redundant (non-orthogonal) but preserves the time series length at all subcomponents, and thus it is particularly useful for forecasting applications.

Despite wavelet analysis widespread use in the field of signal processing, its use in economics has been rather limited, although steadily gaining interest. Following the seminal papers of Ramsey and Lampart (1998a, 1998b) investigating the money and income relationship under different time scales, Ramsey (2002) and Crowley (2007) were the first to provide a comprehensive presentation of wavelets analysis and the two main potential areas for applications in economics. The first allows for an exploratory and descriptive analysis of economic data, while the second utilises multiscale wavelet components for the purpose of forecasting. It has to be noted, though, that ever since wavelet analysis has been employed in modelling the new Keynesian Philips curve (NKPC) – see Martins and Verona (2020) and dynamic stochastic equilibrium models (DSGE) as well – see Sala (2015), Caraiani (2015), and Gallegati, Giri, and Palestrini (2019).

The use of wavelet analysis for exploratory and descriptive analysis poses an alternative way, for example, for GDP trend-cycle decomposition (Yogo, 2008) or understanding better the financial cycle across time and frequency (Verona, 2016). Extending the analysis beyond univariate times series is particularly useful to localise the co-movement and correlation of

[48] See Percival and Walden (2000) for a textbook treatment

two time series both in the time and frequency domain. Continuous wavelet transform and its extension to bivariate time series with wavelet coherence are powerful tools in that respect. Indicatively, Aguiar-Conraria and Soares (2014) review the potential uses of wavelet coherence, Rua (2010) investigate the synchronisation of the euro area countries' individual business cycles, and Tiwari (2013) examines the French inflation and output gap relationship over frequency and across time.

The second use of wavelet analysis which this study is about, is to use it for time series modelling and forecasting. While wavelet analysis has been employed for univariate time series and mainly in a frequentist setting, its use for multivariate time series in economics and in a Bayesian estimation framework is rather scarce. This gap is what this study attempts to fill with a systematic investigation. The four main wavelet-based univariate time series forecasting approaches are presented in Gallegati and Semmler (2014), Part 3.

The first approach consists of a two-step approach, i.e. initial wavelet-based denoising of the time series and subsequent ARIMA modelling – see indicatively Herwartz and Schlüter (2017) and Bruzda (2020). The second approach uses separate modelling for each wavelet component of the time series with ARIMA and/or GARCH modelling – see indicatively Zhang, Gençay, and Yazgan (2017) Uddin, Gençay, Bekiros, and Sahamkhadam (2019). The third approach originally proposed by Renaud, Starck, and Murtagh (2003) uses a single multiscale autoregressive model which uses lagged wavelet components to forecast a univariate time series.

The fourth wavelet-based forecasting approach employs wavelet analysis to model locally stationary time series via time-varying autoregressive models – see Fryzlewicz, Van Bellegem, and Von Sachs (2003) for univariate, and Sato, Morettin, Arantes, and Amaro Jr (2007) for multivariate time series. Additional approaches for the use of wavelet analysis extendable to multivariate time series, which are not investigated in this study, use factor models. For example, Rua (2011) uses a separate univariate factor augmented regression for each scale, and Rua (2017) employs a static factor model with principal components extracted from the multiscale wavelet components set.

The main contribution of this study is the extension of the first three methods in a multivariate setting using Bayesian estimation. Additionally, this study investigates the advantages of using wavelet analysis for mixed-frequency time series modelling, which is a particularly popular area in macroeconometrics, especially during the coronavirus pandemic recession (Huber, Koop, Onorante, Pfarrhofer, & Schreiner, 2020; Cimadomo, Giannone, Lenza, Monti, & Sokol, 2022), and the ever-growing number of high-frequency indicators. To the best of my knowledge, I have not found any similar approaches in the literature, and thus this study attempts to make a major contribution in that respect and establish some key findings.

The essay proceeds as follows. Section 3.2 provides a sound mathematical background and the properties of the wavelet analysis and transforms that will be used. Section 3.3 describes the three suggested wavelet-based approaches for single-frequency multivariate time series forecasting, and in continuation, Section 3.4 presents the findings of the empirical application with the proposed single-frequency time series models. Section 3.5 suggests a wavelet-based approach for mixed-frequency time series modelling, and Section 3.6 presents the relevant results. Finally, Section 3.7 concludes.

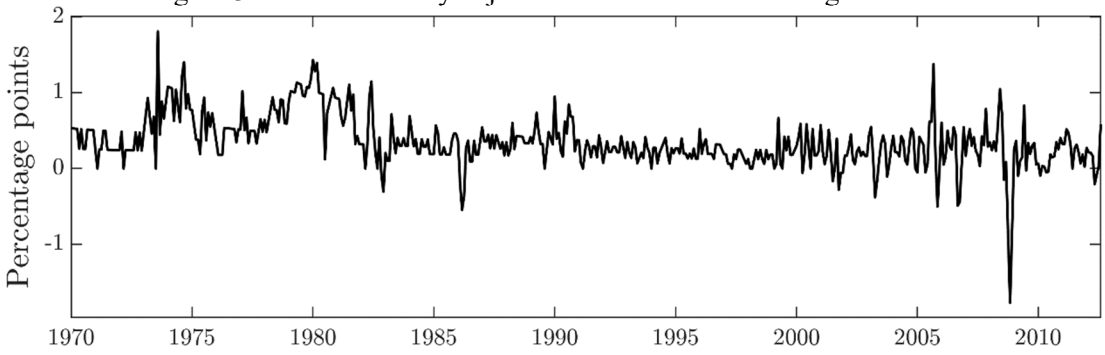
3.2 Wavelet analysis

This section presents the theoretical background of wavelet analysis. In particular, Section 3.2.1 provides the motivation for going beyond time-domain analysis and explains the key aspects of frequency analysis in terms of the Fourier transform and the short-time Fourier transform. In continuation, Section 3.2.2 provides the main advantages of simultaneous time-frequency analysis via the continuous wavelet transform. Finally, the subsequent Sections 3.2.3 and 3.2.4 focus on the discrete wavelet transform and the maximal overlap discrete wavelet transform, which are the two wavelet transforms used later for multivariate time series modelling and forecasting.

3.2.1 Introduction

In general, data points of economic indicators are originally observed usually on a regular basis and at a specific sampling frequency, e.g. monthly or quarterly frequency time series. This representation of a time series, i.e. how a series changes or fluctuates over time, occurs in the time domain, and its properties are examined in time series analysis which is the predominant form used in time series econometrics. For example, Figure 3.1 below shows the US seasonally adjusted CPI month-on-month growth rate.

Figure 3.1: US seasonally adjusted CPI month-on-month growth rate



However, this is not the only available representation. In the frequency-domain (or spectral analysis), a covariance stationary process $\{Y_t\}_{t=-\infty}^{\infty}$ can be viewed as the summation of weighted periodic (cyclical) underlying components in the form of sines and cosines at given cyclical frequencies ω in [0,π] – see Hamilton (1994), p.157:

$$Y_t = \mu + \int_o^{\pi} \left[\alpha(\omega)\cos(\omega t) + \delta(\omega)\sin(\omega t) \right] d\omega \tag{3.1}$$

Where α(ω), δ(ω) are zero mean random variables.

From the above representation, it becomes obvious that there exists a relationship between the time series Y_t and the cyclical frequencies ω. The inverse relationship between ω and Y_t is formally given by the Fourier transform FT(ω), which is defined as – see Mertins (1999), p.80:

$$\mathrm{FT}(\omega)=\sum_{t=-\infty}^{\infty} Y(t)e^{-i\omega t}\,,\quad \omega\in[-\pi,\pi],\quad \mathrm{FT}(\omega):\mathbb{R}\to\mathbb{C} \tag{3.2}$$

Where $i^2=-1$ is the imaginary number and ω corresponds to the cyclical frequency.

Fourier transform FT(ω) is a complex-valued representation of the time series that is difficult to understand its physical interpretation for economic quantities. Instead, we rather use the power spectral density function, which shows how much spectral power can be attributed to the underlying periodic components. Power spectral density $S(\omega)$ is related to the Fourier transform FT(ω) via the following formula – see Mertins (1999), p. 15::

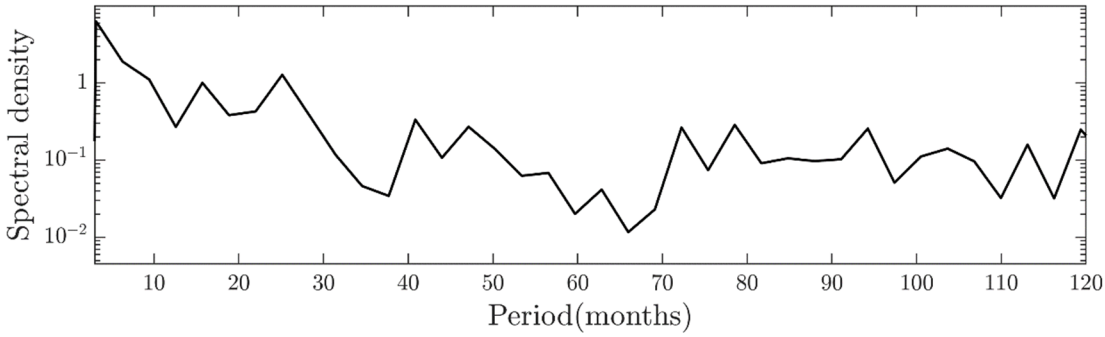
$$S(\omega)=\lim_{T\to\infty}\frac{1}{T}E\left[\left|\mathrm{FT}(\omega)\right|^2\right] \tag{3.3}$$

In more detail, power spectral density $S(\omega)$ can be viewed as the Fourier transform in equation (3.2) applied to the autocovariance function $\gamma(j)$ of a stationary process and is expressed as – see Hamilton (1994), p. 153:

$$S(\omega)=\frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma(j)e^{-i\omega j}\,,\quad \omega\in[-\pi,\pi],\quad S(\omega):\mathbb{R}\to\mathbb{R} \tag{3.4}$$

In practice, however, the population spectral density $S(\omega)$ given above is estimated using the equivalent sample spectral density (also known as periodogram), which uses consistent autocovariance estimators $\hat{\gamma}(j)$. Figure 3.2, which follows below, shows the periodogram for the case of the US seasonally adjusted month-on-month growth rate over 1970M1-2012M8, where the period shown on the x-axis is in months.

Figure 3.2: US CPI month-on-month growth rate spectral density periodogram



Note. Y-axis is logarithmic. CPI is seasonally adjusted.

In order to link the Fourier transform with the short-time Fourier transform and the continuous wavelet transform, which are introduced below, it is more convenient to define the equivalent continuous Fourier transform FT (ω) for a continuous-time signal $Y(t)$ – see Mertins (1999), p.26:

$$\mathrm{FT}(\omega)=\int_{-\infty}^{\infty} Y(t)e^{-i\omega t}\,dt\,,\quad \omega\in[-\infty,\infty],\quad \mathrm{FT}(\omega):\mathbb{R}\to\mathbb{C} \tag{3.5}$$

The frequency-domain analysis applies to the entire time series, and thus the frequency components cannot be localised during different subsamples of the time series, which might exhibit different dynamics. In order to overcome this drawback, short-time Fourier transform (STFT) can be applied to the time series, i.e. Fourier transform is applied to a sliding (potentially overlapping) window of the time series. Nonetheless, the time resolution (width) of the window affects the frequency bands that can be observed. A narrower window provides good time resolution but poor frequency resolution and vice-versa. To formalise ideas, for a continuous-time signal $Y(t)$, its continuous time STFT is defined as – see Mertins (1999), p. 197:

$$\mathrm{STFT}(\tau,\omega)=\int_{-\infty}^{\infty} Y(t)\,w(t-\tau)\,e^{-i\omega t}\,dt\,,\quad \omega\in[-\infty,\infty],\quad \mathrm{STFT}(\tau,\omega):\mathbb{R}^2\to\mathbb{C} \tag{3.6}$$

Where $w(\cdot)$ is a window function.

The relation between the FT and the STFT can be seen by comparing equations (3.5) and (3.6), where the STFT is based on the FT and uses the window function[49] $w(\cdot)$ additionally. The term $a_{\tau,\omega}(t)=w(t-\tau)\,\mathrm{e}^{-i\omega t}$ is also called the time-frequency atom – see Mallat (1999), p.67

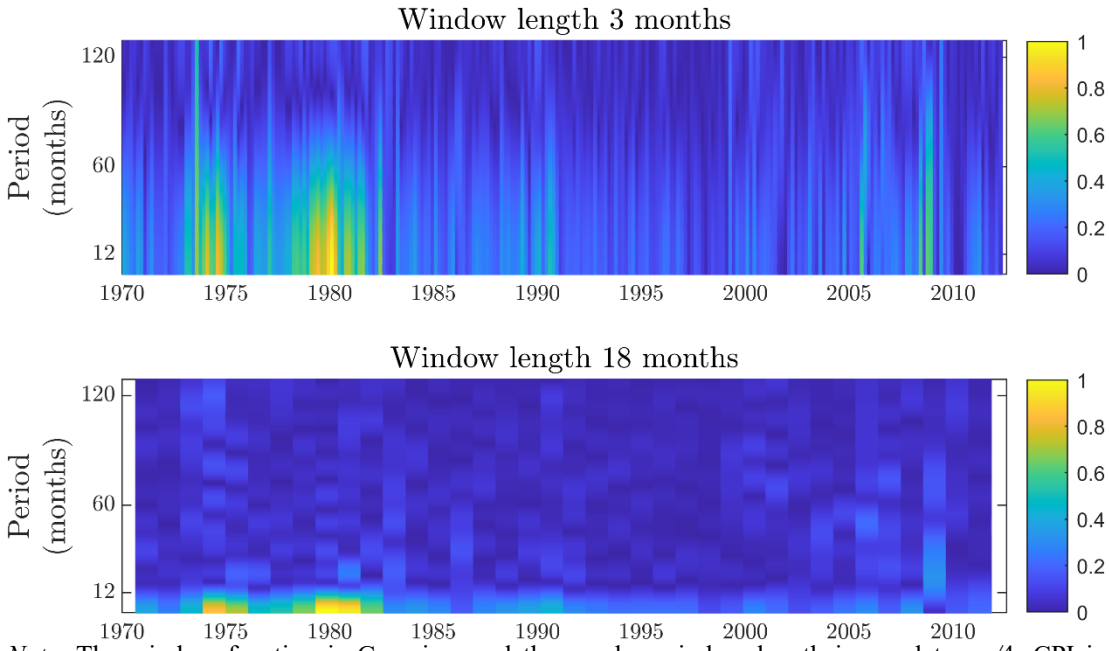
Analogous to the periodogram, which is used for the visual representation of Fourier transform's spectral density, the spectrogram is used to display the properties of the complex-valued STFT and is defined as the squared magnitude of the STFT – see Mertins (1999), p. 201:

$$S(\tau,\omega)=\left|\mathrm{STFT}(\tau,\omega)\right|^2 \tag{3.7}$$

For example, Figure 3.3 below shows the normalised spectrogram for US CPI using a window length of 3 and 18 months, respectively. The trade-off between time and frequency resolution becomes obvious. When a 3-month window length is used, a high time resolution can be achieved on the time axis, whereas on the y-axis representing the frequency, the resolution remains low. For example, the 1979-80 oil price shocks' effect on CPI and the 2008 global financial crisis shocks can be localised precisely on time; however, on the frequency axis, their magnitude changes rather smoothly. On the contrary, when the window length is set to 18 months, the time resolution decreases, and the aforementioned shocks cannot get localised precisely, but the frequency resolution increases. For instance, during the Great Moderation period 1985-2005, it is easier to distinguish different frequencies on the y-axis.

[49] For the special case that the window function takes a normalised Gaussian form, STFT translates to Gabor transform (Gabor, 1946).

Figure 3.3: US CPI month-on-month growth rate normalised spectrogram



Note. The window function is Gaussian, and the overlap window length is equal to $w_L/4$. CPI is seasonally adjusted.

These shortcomings of the short-time Fourier transform have been the motivation for the creation of the wavelet transform. In particular, the time resolution is not fixed anymore but varies, and consequently, the frequency resolution is adjusted to the time resolution accordingly without any loss of information in either of the two domains.

3.2.2 Continuous wavelet transform

In order to apply the Continuous Wavelet Transform (CWT), it is necessary to define the wavelet function. Wavelet is essentially a small wave in contrast to the infinite sines and cosines used in the Fourier transform defined over the entire time series length.

The wavelet function (mother wavelet) $\psi(t)$ captures the high-frequency components of the time series and satisfies the following two properties:

$$\int_{-\infty}^{\infty} \psi(t)\,dt = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} \psi^2(t)\,dt = 1 \tag{3.8}$$

Similarly, the scaling function (father wavelet) $\varphi(t)$ captures the low-frequency dynamics of the time series and satisfies:

$$\int_{-\infty}^{\infty} \varphi(t)\,dt = \sqrt{2} \quad \text{and} \quad \int_{-\infty}^{\infty} \varphi^2(t)\,dt = 1 \tag{3.9}$$

The wavelet function $\psi(t)$ has to be scaled (dilated) by factor α and shifted (translated) by factor b in order to apply to different frequency resolutions and points in time, respectively:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{3.10}$$

There exists a big class of wavelets that satisfy the above properties. The first proposed wavelet, known as the Haar wavelet (Haar, 1910), is defined by the following wavelet function:

$$\psi(t)=\begin{cases}-1/\sqrt{2}, & -1<t\le 0\\ 1/\sqrt{2}, & 0<t\le 1\\ 0, & \text{otherwise}\end{cases} \tag{3.11}$$

Equivalently, the Haar scaling function $\varphi(t)$ is defined as:

$$\varphi(t)=\begin{cases}1/\sqrt{2}, & -1<t\le 1\\ 0, & \text{otherwise}\end{cases} \tag{3.12}$$

The wavelet $\psi(t)$ and scaling $\varphi(t)$ functions defined in equations (3.11) and (3.12), respectively, are displayed in Figure 3.4 below.

Figure 3.4: Haar wavelet and scaling function

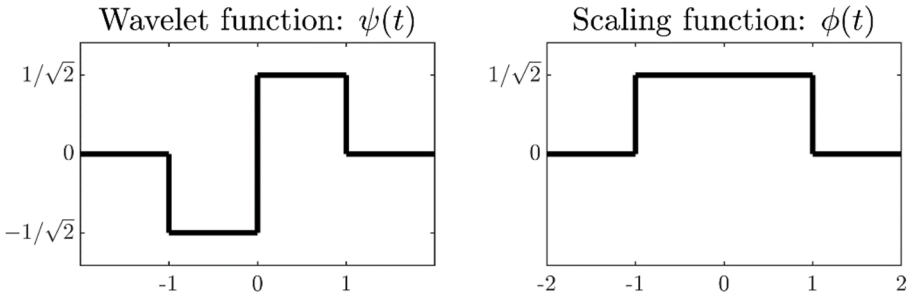
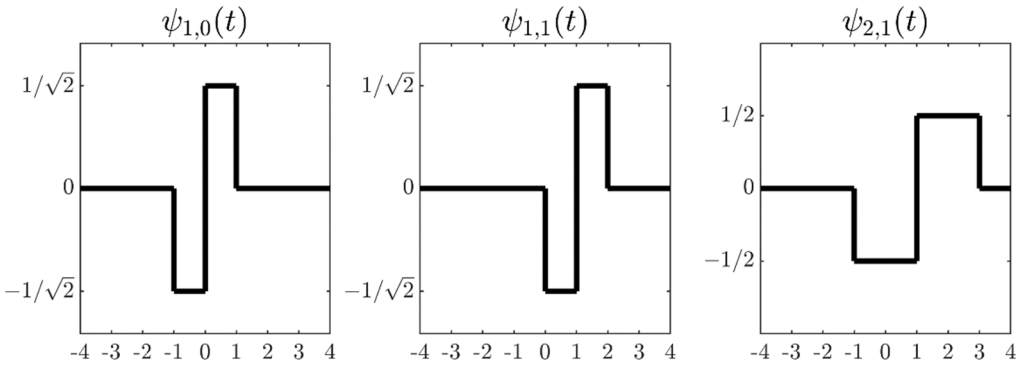


Figure 3.5 below shows the wavelet function $\psi_{\alpha,b}(t)$ as further explained in equation (3.10) when scaled and shifted by factors α and b respectively for the following 3 cases: i. $\alpha = 1$ and $b = 0$ ii. $\alpha = 1$ and $b = 1$, and iii. $\alpha = 2$ and $b = 1$.

Figure 3.5. Scaled and translated Haar wavelet function



Assuming a continuous-time signal $Y(t)$, its continuous wavelet transform, $CWT(\alpha,b)$ at scale α and time b, is defined as:

$$CWT(\alpha,b)=\int_{-\infty}^{\infty} Y(t)\,\psi_{\alpha,b}(t)\,dt \tag{3.13}$$

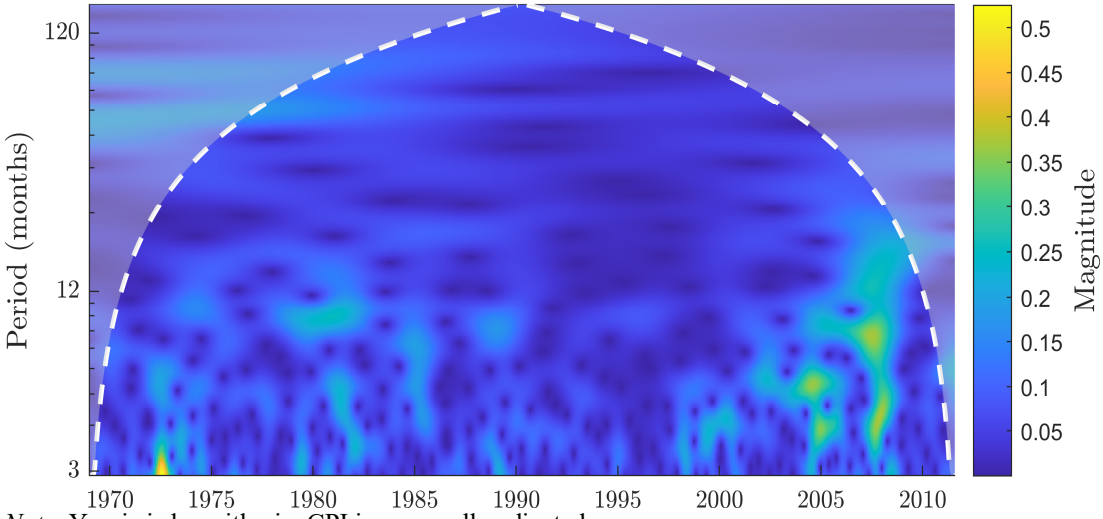
Where $\psi_{\alpha,b}(t)=\frac{1}{\sqrt{\alpha}}\psi\left(\frac{t-b}{\alpha}\right)$ as defined previously in equation (3.10), is the wavelet function scaled by factor α and translated at a given point b of the time series $Y(t)$.

In order to understand better how the CWT poses as an extension of the STFT defined previously, we should observe that the time-frequency atom, i.e. the term $w(t-\tau)\,e^{-i\omega t}$ in equation (3.6), is now substituted by the wavelet function $\psi_{\alpha,b}(t)$ in equation (3.13).

It is easier to understand the output of the CWT via visualisation; see for example, Verona (2016) for a time-frequency analysis of the US financial cycle. Figure 3.6, which follows, shows the CWT applied to the US CPI seasonally adjusted month-on-month growth rate. Y-axis in the logarithmic scale shows the period of the time-localised fluctuations, and the x-axis shows the time, i.e. the impact of changing the scaling factor α and b, respectively, in equation (3.11). Finally, $CWT(\alpha,b)$, i.e. the magnitude of the CWT, is shown via the colour. For example, the temporary shocks, which are visible in Figure 3.1, for August 1973.M8, around 2005 and during the 2008 financial crisis, can be localised in the yellow areas of Figure 3.6. The 1973.M8 shock had a period in the range of 1-2 months, while on the other hand, the 2005 and 2008 shocks, which lasted longer, are shown to have periods in the y-axis between 1 to 5 and 1 to 12 months, respectively.

The CWT can be extended beyond univariate time series resulting in the cross-wavelet transform and the wavelet coherence for bivariate time series. Wavelet coherence allows via a visualisation similar to Figure 3.6 to localise in the time-frequency plane areas with great power (correlation) between 2 time series, as well as the difference in phase between each other, i.e. whether one time series has a lead or lag relative to the other. Aguiar-Conraria and Soares (2014) provide an overview of the cross-wavelet transform and wavelet coherence and their relevance for economic analysis. Indicatively, articles investigating economic and/or financial time series interdependencies via the wavelet coherency include Rua (2010), Aguiar-Conraria and Soares (2011), Tiwari (2013), Trezzi (2013), Andrieș, Ihnatov, and Tiwari (2014), Gülerce and Ünal (2016), Hkiri, Hammoudeh, Aloui, and Shahbaz (2018), and Uddin, Bekiros, and Ahmed (2018).

Figure 3.6: US CPI month-on-month growth rate Haar CWT



Note. Y-axis is logarithmic. CPI is seasonally adjusted.

3.2.3 Discrete wavelet transform

The Discrete wavelet transform (DWT) (see Percival and Walden (2000), Ch. 4) is an orthogonal transformation of Y_t into discrete frequency bands, each one of which is called level j of decomposition or scale j. Assuming Y_t is of length T, which is $T = 2^J$, the maximum level of decomposition is $J = \log_2 T$. This is called the full DWT. However, if someone is not interested in the very low-frequency components (large scales), the partial DWT can be applied for $J_0 < J$, where in that case, T can be an integer multiple of 2^{J_0}.

The DWT[50] is given as follows:

$$\tilde{W} = \tilde{\mathcal{W}} Y \tag{3.14}$$

Where $\tilde{W}$ is $T \times 1$ vector and can be portioned $J+1$ subvectors $\tilde{W}_j$ of length $T_j = T/2^j$:

$$\tilde{W} = \left[\tilde{W}_1', \cdots, \tilde{W}_j', \cdots, \tilde{W}_J', \tilde{V}_J'\right]' \tag{3.15}$$

Obviously, for $j = J$, it follows that $\tilde{W}_J$ and $\tilde{V}_J$ are scalars.

Where $\tilde{\mathcal{W}}$ is the $T \times T$ square orthonormal[51] matrix basis of the transformation. $\tilde{\mathcal{W}}$ is partitioned into submatrices:

$$\tilde{\mathcal{W}} = \left[\tilde{\mathcal{W}}_1', \cdots, \tilde{\mathcal{W}}_j', \cdots, \tilde{\mathcal{W}}_J', \tilde{\mathcal{V}}_J'\right]' \tag{3.16}$$

[50] To keep the notation consistent and easily read in the subsequent Sections 3.2.2 and 3.2.4, the current section denotes DWT quantities with tilde and MODWT without, whereas in see Percival and Walden (2000) the opposite holds.

[51] Orthonormality property of the square matrix $\tilde{\mathcal{W}}$, implies $\tilde{\mathcal{W}}\tilde{\mathcal{W}}' = \tilde{\mathcal{W}}'\tilde{\mathcal{W}} = I_T$ and $\tilde{\mathcal{W}}' = \tilde{\mathcal{W}}^{-1}$.

Where the orthonormality property[52] is maintained for each of the submatrices $\tilde{\mathcal{W}}_j$, $\tilde{\mathcal{V}}_J$ of dimensions $T_j \times T$.

Using Parseval's theorem of energy preservation, the DWT follows the property[53]:

$$\|Y\|^2 = \left\|\tilde{W}\right\|^2 = \sum_{j=1}^{J}\left\|\tilde{W}_j\right\|^2 + \left\|\tilde{V}_J\right\|^2 \tag{3.17}$$

Using the orthonormality property of $\tilde{\mathcal{W}}$ is useful in order to define the inverse DWT, which reconstructs the original time series Y via the transpose or inverse of the basis of the transformation matrix applied to the $\tilde{W}$:

$$Y = \tilde{\mathcal{W}}'\tilde{W} = \tilde{\mathcal{W}}^{-1}\tilde{W} \tag{3.18}$$

So far, the DWT and the basis of the transformation $\tilde{\mathcal{W}}$ have been presented; however, the elements that constitute the matrix $\tilde{\mathcal{W}}$ have not been discussed. In practice, with the exception of the Haar transformation matrix $\tilde{\mathcal{W}}$ presented later in this section, it is more convenient to estimate the wavelet and scaling coefficients via the use of wavelet and scaling filters. Similar to the wavelet and scaling functions presented for the CWT in Section 3.2.2 and the properties they need to satisfy in equations (3.8) and (3.9), the DWT uses the equivalent discrete wavelet and scaling filters which satisfy the properties in discrete time. The wavelet filter $\tilde{h}_l$ of length L describes the high-frequency components of the time series and should satisfy the following two properties:

$$\sum_{l=0}^{L-1}\tilde{h}_l = 0 \quad \text{and} \quad \sum_{l=0}^{L-1}\tilde{h}_l^2 = 1 \tag{3.19}$$

Similarly, the scaling filter $\tilde{g}_l$ of length L as well captures the low-frequency dynamics of the time series and should satisfy:

$$\sum_{l=0}^{L-1}\tilde{g}_l = 0 \quad \text{and} \quad \sum_{l=0}^{L-1}\tilde{g}_l^2 = 1 \tag{3.20}$$

Now, following the so-called pyramid algorithm, which in practice, applies a wavelet and a scaling filter recursively to the output of a scaling filter from the previous step, the wavelet and scaling coefficients can be estimated recursively as follows (a visual representation of the pyramid algorithm is shown in Figure 3.9 in Section 3.2.4):

$$\tilde{W}_{j,t} = \sum_{l=0}^{L-1}\tilde{h}_l V_{j-1,(2t-1-l) \bmod T_{j-1}} \quad \text{and} \quad \tilde{V}_{j,t} = \sum_{l=0}^{L-1}\tilde{g}_l V_{j-1,(2t-1-l) \bmod T_{j-1}} \tag{3.21}$$

Where $t = 1,\cdots,T_j$ with $T_j = T/2^j$.

When the recursion initiates for $j = 1$, the original time series Y_t, $t = 1,\cdots, T$ is used:

$$\tilde{W}_{1,t} = \sum_{l=0}^{L-1}\tilde{h}_l Y_{(2t-1-l) \bmod T} \text{ and } \tilde{V}_{j,t} = \sum_{l=0}^{L-1}\tilde{g}_l Y_{(2t-1-l) \bmod T} \text{ for } t = 1,\cdots,T/2 \tag{3.22}$$

---

[52] Orthonormality property of $\tilde{\mathcal{W}}_j$ satisfies $\tilde{\mathcal{W}}_j\tilde{\mathcal{W}}_j' = I_{T_j}$ while $\tilde{\mathcal{W}}_j\tilde{\mathcal{W}}_j' \neq \tilde{\mathcal{W}}_j'\tilde{\mathcal{W}}_j$.

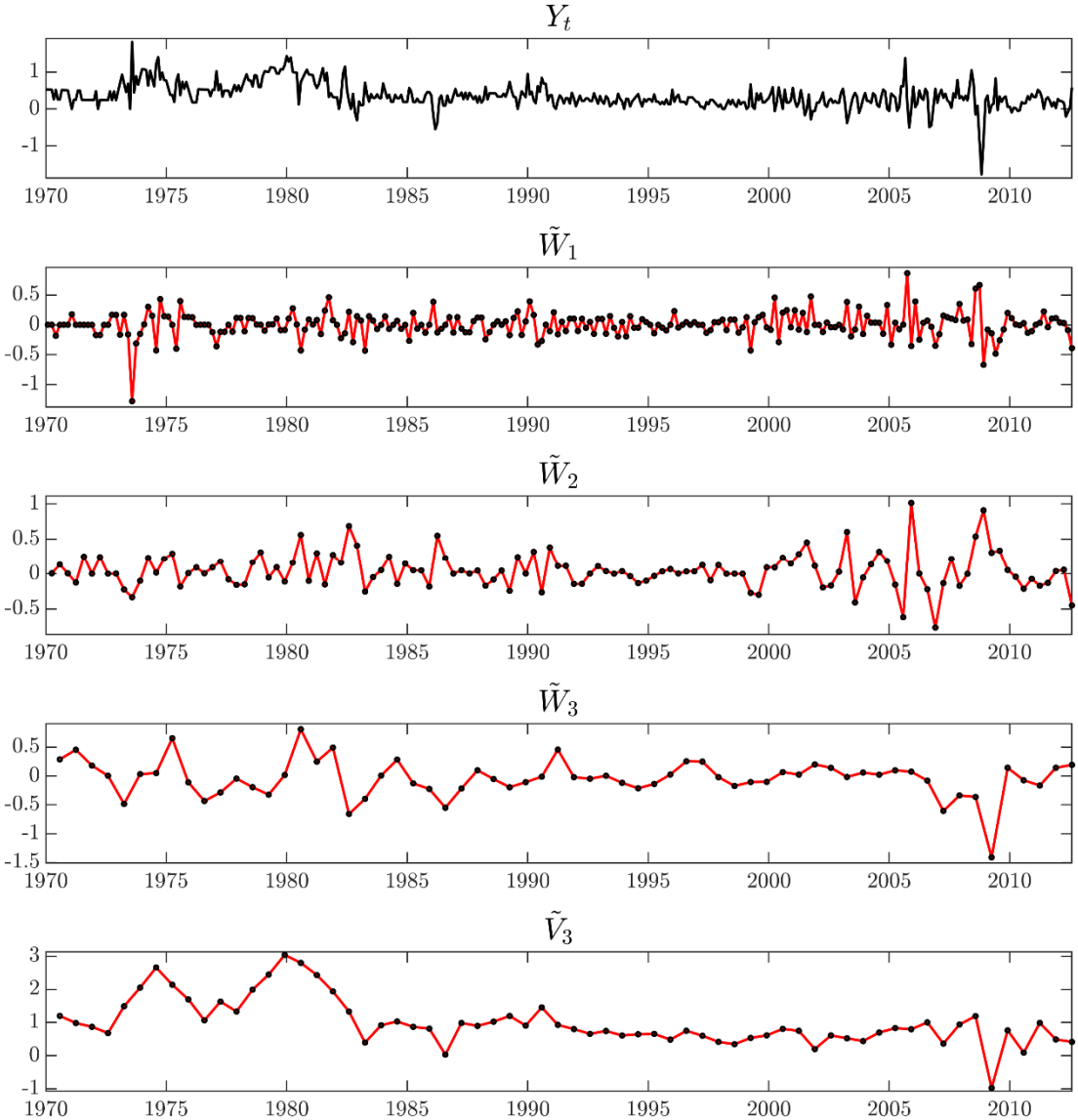
[53] Squared norm $\|\cdot\|^2$ for Y_t is defined as $\|Y\|^2 = \sum_{t=1}^{T} Y_t^2$.

By introducing j-level specific wavelet ($\tilde{h}_{j,l}$) and scaling filters ($\tilde{g}_{j,l}$) of length $L_j = (2^j - 1)(L - 1) + 1$, $j = 2,\cdots, J$ where L denotes the length of the filter for $j = 1$, it is possible to estimate the wavelet and scaling coefficients at any level directly from the time series Y_t, without the need of the recursion as follows:

$$\tilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} Y_{2^j(t+1)-1-l \bmod T} \quad \text{and} \quad \tilde{V}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} Y_{2^j(t+1)-1-l \bmod T} \tag{3.23}$$

Figure 3.7 below shows the $J_0 = 3$ partial Haar DWT of the US seasonally adjusted CPI month-on-month of growth rate for $T = 512$ observations. The full DWT would have $J = 9$ scales, i.e. $9 = \log_2(512)$.

Figure 3.7: US CPI month-on-month growth rate $J_0 = 3$ levels partial Haar DWT



Note. First subplot Y_t shows CPI month-on-month growth rate. Subsequent subplots show partial $J_0 = 3$ Haar DWT. Wavelet and scaling coefficients capture fluctuations with periods in months as follows: $\tilde{W}_1 : 2-4$, $\tilde{W}_2 : 4-8$, $\tilde{W}_3 : 8-16$ and $\tilde{V}_3 :> 16$. $\tilde{W}_1$, $\tilde{W}_2$, $\tilde{W}_3$, and $\tilde{V}_3$ have 256, 128, 64, and 64 observations, respectively. CPI is seasonally adjusted.

Using the orthonormality property of $\tilde{\mathcal{W}}$, the DWT multiresolution analysis (DWT-MRA) of time series Y_t can be formed as:

$$Y = \tilde{\mathcal{W}}'\tilde{W} = \sum_{j=1}^{J} \tilde{\mathcal{W}}_j' \tilde{W}_j + \tilde{\mathcal{V}}_J' \tilde{V}_J = \sum_{j=1}^{J} \tilde{D}_j + \tilde{S}_J$$

Where $\tilde{D}_j$ is called the detail coefficients at scale j and $\tilde{S}_J$ is the smooth (also called coarse or approximate) coefficients of Y_t at scale J. $\tilde{D}_j$ and $\tilde{S}_J$ have the same length T as Y_t. Using Parseval's theorem of energy preservation, the DWT-MRA follows the property:

$$\|Y\|^2 = \left\|\tilde{W}\right\|^2 = \sum_{j=1}^{J} \left\|\tilde{W}_j\right\|^2 + \left\|\tilde{V}_J\right\|^2 = \sum_{j=1}^{J} \left\|\tilde{D}_j\right\|^2 + \left\|\tilde{S}_J\right\|^2 \tag{3.24}$$

For the specific case of the Haar wavelet, the transform matrix $\tilde{\mathcal{W}}$ is given recursively (see Sundararajan (2016), Ch. 8) as:

$$\tilde{\mathcal{W}}_{\tilde{T}} = \begin{bmatrix} I_{\tilde{T}/2} \otimes \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \\ \tilde{\mathcal{W}}_{\tilde{T}/2} \otimes \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \end{bmatrix} \tag{3.25}$$

The recursion starts at $\tilde{T} = 4$ with $\tilde{T}/2 = 2$: $\tilde{\mathcal{W}}_2 = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$ and ends at $\tilde{T} = T$ resulting in total $J - 1 = \log_2(T) - 1$ repetitions and yields $\tilde{\mathcal{W}}_{\tilde{T}} = \tilde{\mathcal{W}}_T = \tilde{\mathcal{W}}$. Subscript $\tilde{T}$ should not be confused with subscripts $j = 1,\cdots, J$ for the submatrices in equation (3.16).

For example, for $T = 4$ and $J = 2$, the recursion would involve $\log_2(4) - 1 = 1$ repetitions. Using equation (3.25), the repetition would yield $\tilde{\mathcal{W}}_4$, i.e. the entire DWT transformation matrix of dimensions 4 × 4 as follows:

$$\tilde{\mathcal{W}}_4 = \begin{bmatrix} I_2 \otimes \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \\ \tilde{\mathcal{W}}_2 \otimes \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 & 1/2 \end{bmatrix}$$

Alternatively, one could use equations (3.21) and (3.23) with the Haar wavelet and scaling filters provided below:

$$\tilde{h}_0 = 1/\sqrt{2} \text{ and } \tilde{h}_1 = -1/\sqrt{2}, \quad \text{and} \quad \tilde{g}_0 = 1/\sqrt{2} \text{ and } \tilde{g}_1 = 1/\sqrt{2} \text{ for } j = 1$$

For $j \geq 2$, the Haar scaling filter $\tilde{g}_{j,l}$ is given as Percival and Walden (2000), p. 103:

$$\tilde{g}_{j,l} = \begin{cases} \dfrac{1}{2^{j/2}}, l = 0,\cdots, 2^j - 1 \\ 0, \quad \text{otherwise} \end{cases} \tag{3.26}$$

Regarding the Haar wavelet filter $\tilde{h}_{j,l}$, by application of the following generic formula for any wavelet Percival and Walden (2000), p. 103:

$$\tilde{h}_{j,l} = \sum_{k=0}^{L-1} \tilde{h}_k \tilde{g}_{j-1,l-2^{j-1}k}$$

$\tilde{h}_{j,l}$ is given as follows:

$$\tilde{h}_{j,l} = \begin{cases} \dfrac{1}{2^{j/2}}, & l = 0,\cdots,L_j/2-1 \\ -\dfrac{1}{2^{j/2}}, & l = L_j/2,\cdots,L_j \end{cases} \tag{3.27}$$

By substituting equations (3.26) and (3.27) in equation (3.23), one can show that for the full Haar DWT transform, wavelet $\tilde{W}_{j,k}$ and scaling $\tilde{V}_J$ coefficients are estimated as differences of moving averages of Y_t and a weighted average, respectively[54], given as:

$$\tilde{W}_{j,t} = \frac{1}{2^{j/2}}\left(\sum_{k=tc-d}^{tc} Y_k - \sum_{k=tc-2d-1}^{tc-d-1} Y_k \right), \quad j = 1,\cdots,J, \quad t = 1,\cdots,T/2^j$$

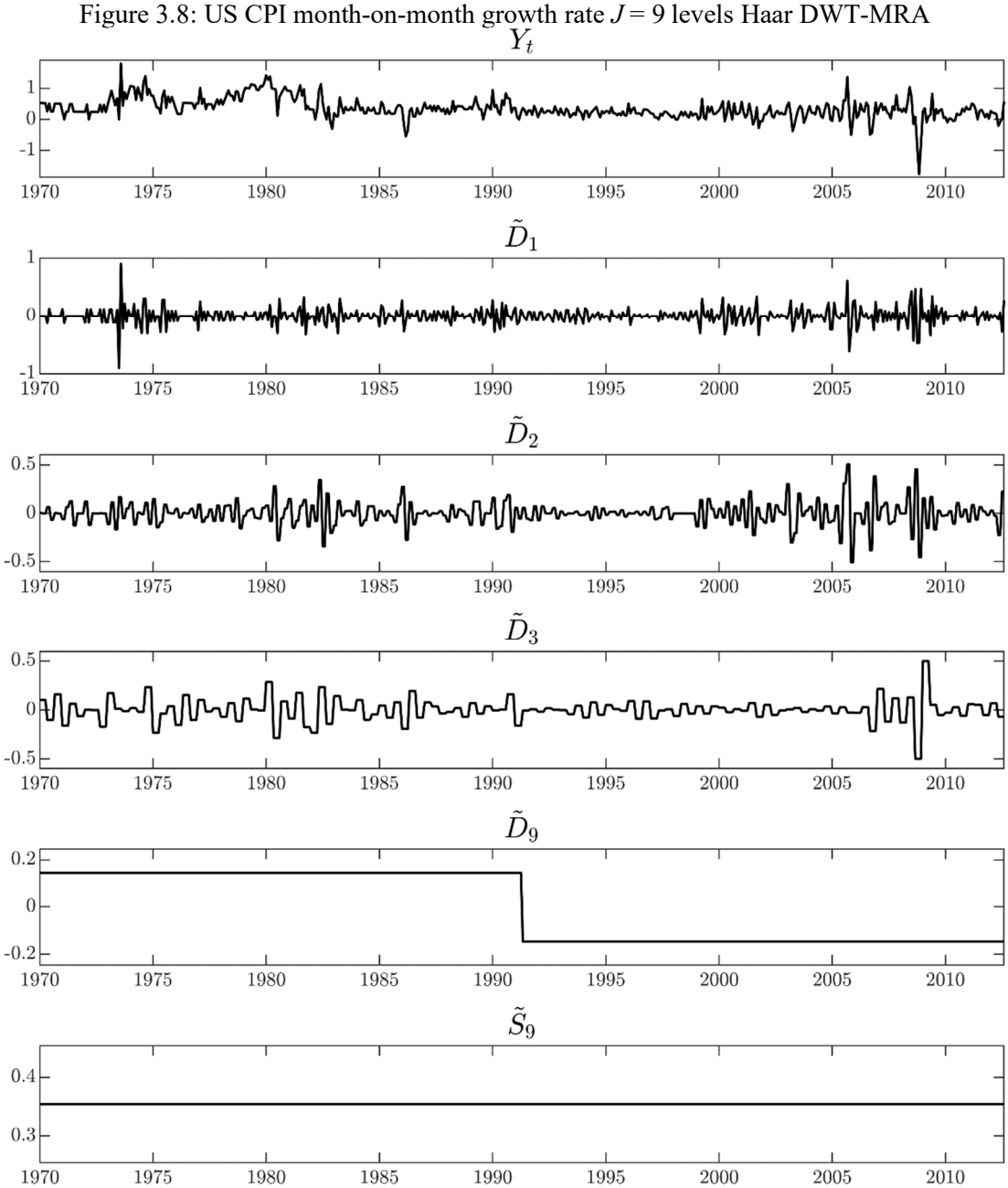
Where $c = 2^j$ and $d = 2^{j-1} - 1$.

The single scaling coefficient $\tilde{V}_J$ is given as $\tilde{V}_J = \frac{1}{2^{j/2}}\left(\sum_{k=1}^{T} Y_k \right)$.

Figure 3.8 below shows the Haar DWT-MRA of the US seasonally adjusted CPI month-on-month growth rate for $T = 512$ observations. Full DWT has $J = 9$ levels of decomposition ($9 = \log_2(512)$), i.e. detail coefficients at nine scales and the smooth coefficients $\tilde{S}_9$. For the purpose of the figure's visibility, only four out of the nine details coefficients ($\tilde{D}_1$, $\tilde{D}_2$, $\tilde{D}_3$, and $\tilde{D}_9$) are shown, as well as the smooth coefficients $\tilde{S}_9$.

[54] See Percival and Walden (2000), p. 58-59, for an example with $T = 16$.

Figure 3.8: US CPI month-on-month growth rate $J = 9$ levels Haar DWT-MRA



Note. First subplot Y_t shows CPI month-on-month growth rate. Subsequent subplots show the Haar DWT-MRA selected detail and smooth coefficients. CPI is seasonally adjusted.

It has to be noted that for practical applications, such as the DWT-based denoising, discussed later in Section 3.3.1, where the partial DWT should be used when the length T is not an integer multiple of 2^{J_0} the so-called 'padding' procedure is used. The original time series Y_t is 'padded' in a symmetric manner at the start and end of the time series with the mirrored values at both edges, such that a new time series $\tilde{Y}_t$ of appropriate length is constructed where the partial DWT can be applied.

3.2.4 Maximal overlap discrete wavelet transform

The Maximal overlap discrete wavelet transform (MODWT) is an extension of the DWT, and its main characteristic is that it is a highly redundant and nonorthogonal transformation – see Percival and Walden (2000), Ch. 5. Moreover, the MODWT wavelet W_j and scaling V_J coefficients are of length T (non decimated) in contrast to the DWT where they are of length $T / 2^j$ (decimated). In general, the MODWT exhibits the following advantages versus the DWT:

1. The length T of the time series Y_t is not anymore restricted to be equal to a power of 2 like in the full DWT, and thus it can be applied to a time series of arbitrary length.

2. The MODWT, in contrast to the DWT, is translation invariant[55], i.e. a circular shift in the time series for the MODWT will result in a similar shift in the W_j and V_j components, whereas for the DWT, this property does not hold. The same property of translation invariance holds for the MODWT-MRA[56], where the detail D_j and smooth S_J coefficients can be shifted in a similar way, whereas for the DWT-MRA, this property does not hold.

3. For the MODWT, the energy decomposition holds for the wavelet W_j and scaling V_J coefficients similar to the DWT in equation (3.17). However, for the MODWT-MRA, the energy decomposition for the details D_j and smooth S_J coefficients in contrast to the DWT-MRA in equation (3.24) does not hold.

4. The MODWT uses zero-phase filters in contrast to the DWT, where they are not used, which in practice means that the MODWT-MRA detail and smooth coefficients align with the original time series more meaningfully for practical applications – see an example in Figure 3.11.

The MODWT maximum number of decomposition levels J, contrary to the DWT, is not restricted anymore by the length Y_t and, in principle, can be any positive integer number. However, in such a case, i.e. $J > \log_2(T)$, the decomposition components at a level larger than J would offer information for periodic components whose period is larger than the entire length of the time series, and thus, this piece of information would not be an accurate reflection of the reality. In practice, the numbers of scales are set as $J \leq \log_2(T)$ at an optimal value for the specific application working on.

The MODWT of a time series Y_t of arbitrary length T yields the wavelets and scaling coefficients of the same length T:

$$W_j = \mathcal{W}_j Y \quad \text{and} \quad V_J = \mathcal{V}_J Y \tag{3.28}$$

Where $\mathcal{W}_j$ and $\mathcal{V}_J$ are of dimensions $T \times T$ and are defined as:

[55] Another property of translation invariance is that the energy of the corresponding MODWT components remain the same, whereas for DWT is not.
[56] The MODWT-MRA was originally proposed by Shensa (1992).

$$\mathcal{W}_j = \begin{bmatrix} h^p_{j,0} & h^p_{j,T-1} & h^p_{j,T-2} & \cdots & h^p_{j,2} & h^p_{j,1} \\ h^p_{j,1} & h^p_{j,0} & h^p_{j,T-1} & \cdots & h^p_{j,3} & h^p_{j,2} \\ h^p_{j,2} & h^p_{j,1} & h^p_{j,0} & \cdots & h^p_{j,4} & h^p_{j,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ h^p_{j,T-2} & h^p_{j,T-3} & h^p_{j,T-4} & \cdots & h^p_{j,0} & h^p_{j,T-1} \\ h^p_{j,T-1} & h^p_{j,T-2} & h^p_{j,T-3} & \cdots & h^p_{j,1} & h^p_{j,0} \end{bmatrix}$$

$$\mathcal{V}_J = \begin{bmatrix} g^p_{j,0} & g^p_{j,T-1} & g^p_{j,T-2} & \cdots & g^p_{j,2} & g^p_{j,1} \\ g^p_{j,1} & g^p_{j,0} & g^p_{j,T-1} & \cdots & g^p_{j,3} & g^p_{j,2} \\ g^p_{j,2} & g^p_{j,1} & g^p_{j,0} & \cdots & g^p_{j,4} & g^p_{j,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ g^p_{j,T-2} & g^p_{j,T-3} & g^p_{j,T-4} & \cdots & g^p_{j,0} & g^p_{j,T-1} \\ g^p_{j,T-1} & g^p_{j,T-2} & g^p_{j,T-3} & \cdots & g^p_{j,1} & g^p_{j,0} \end{bmatrix}$$

The elements of matrices $\mathcal{W}_j$ and $\mathcal{V}_J$ consist of the MODWT periodized filters $h^p_{j,l}$ and $g^p_{j,l}$ of length T, which are defined as:

$$h^p_{j,l} = \begin{cases} h_{j,l}, & 0 \le l \le L_j - 1 \\ 0, & L_j \le l \le T-1 \end{cases} \quad \text{and} \quad g^p_{j,l} = \begin{cases} g_{j,l}, & 0 \le l \le L_j - 1 \\ 0, & L_j \le l \le T-1 \end{cases}$$

The MODWT periodized wavelet $h^p_{j,l}$ and scaling $g^p_{j,l}$ filters are estimated using the MODWT wavelet and scaling filters, $h_{j,l}$ and $g_{j,l}$, respectively, at level j and of length $L_j = (2^j - 1)(L - 1) + 1, j = 2,\cdots, J$ where L denotes the length of the filter for $j = 1$. The MODWT wavelet and scaling filters, $h_{j,l}$ and $g_{j,l}$, in turn, can be defined using the corresponding DWT equivalent filters denoted with a tilde below:

$$h_{j,l} = \frac{\tilde{h}_{j,l}}{2^{j/2}} \quad \text{and} \quad g_{j,l} = \frac{\tilde{g}_{j,l}}{2^{j/2}} \tag{3.29}$$

Where $\tilde{h}_{j,l}$ and $\tilde{g}_{j,l}$ correspond to the DWT wavelet and scaling filters at level j with length L_j and $l = 0,\cdots, L_j - 1$ for $j = 2,\cdots, J$.

Thus, in more detail, the wavelet W_j and scaling V_j coefficients at time t can be defined equivalently as:

$$W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} Y_{(t-l) \bmod T} = \sum_{l=0}^{T-1} h^p_{j,l} Y_{(t-l) \bmod T} \tag{3.30}$$

$$V_{j,t} = \sum_{l=0}^{L_j-1} g_{j,l} Y_{(t-l) \bmod T} = \sum_{l=0}^{T-1} g^p_{j,l} Y_{(t-l) \bmod T} \tag{3.31}$$

For each of the equations (3.30) and (3.31), the two expressions of the wavelet $W_{j,t}$ and scaling $V_{j,t}$ coefficients are equivalent and show the relationship between using the MODWT filters ($h_{j,l}$, $g_{j,l}$) or the MODWT periodised filters ($h^p_{j,l}, g^p_{j,l}$) in order to estimate directly from Y_t the wavelet $W_{j,t}$ and scaling $V_{j,t}$ coefficients at a given scale j.

Using Parseval's theorem of energy preservation, the MODWT follows the property:

$$\|Y\|^2 = \|W\|^2 = \sum_{j=1}^{J} \|W_j\|^2 + \|V_J\|^2 \tag{3.32}$$

Closely related to the energy decomposition across the wavelet and scaling coefficients of equation (3.32), it follows that the sample variance of Y_t can be decomposed into the variance of each one of the wavelet and scaling coefficients. Denoting the sample variance of Y as $\hat{\sigma}_Y$ it follows:

$$\hat{\sigma}_Y = \frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \bar{Y}\right)^2 = \frac{1}{T}\sum_{t=1}^{T} Y_t^2 - \bar{Y}^2 = \frac{1}{T}\|Y\|^2 - \bar{Y}^2 \tag{3.33}$$

Where $\bar{Y}$ is the sample mean of the elements of Y_t, i.e. $\bar{Y} = 1/T\sum_{t=1}^{T} Y_t$.

Substituting equation (3.32) in equation (3.33) results in the variance decomposition across the wavelet and scaling coefficients:

$$\hat{\sigma}_Y = \frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \bar{Y}\right)^2 = \frac{1}{T}\sum_{j=1}^{J}\|W_j\|^2 + \frac{1}{T}\|V_J\|^2 - \bar{Y}^2 \tag{3.34}$$

Or equivalently $\hat{\sigma}_Y = \sum_{j=1}^{J}\hat{\sigma}_{W_j} + \hat{\sigma}_{V_J}$, given the fact that W_j are zero mean and that the mean of V_J is equal to the mean of Y_t, i.e. $\bar{V}_J = \bar{Y}$. This variance decomposition can be used to conduct inference and hypothesis testing, or alternatively, as shown in Section 4.1, for a descriptive analysis of the time series, used to explain which of the wavelet and scaling coefficients are the ones that explain more the variability and fluctuations of the underlying time series.

Turning now to the MODWT-MRA of a time series Y_t of arbitrary length T, the following additive decomposition holds:

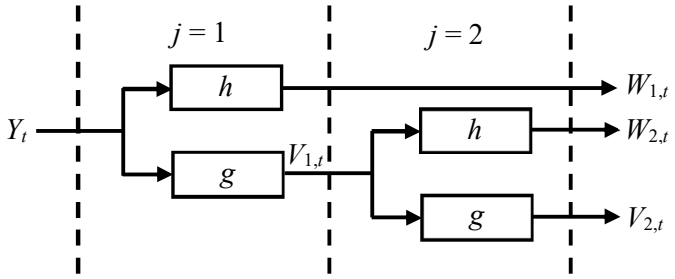
$$Y = \sum_{j=1}^{J}\mathcal{W}_j' W_j + \mathcal{V}_J' V_J = \sum_{j=1}^{J} D_j + S_J \tag{3.35}$$

However, the MODWT-MRA detail and smooth coefficients, as already mentioned in the introduction of this section, cannot be used for the energy decomposition of Y_t. In other words:

$$\|Y\|^2 = \|W\|^2 = \sum_{j=1}^{J}\|W_j\|^2 + \|V_J\|^2 \neq \sum_{j=1}^{J}\|D_j\|^2 + \|S_J\|^2 \tag{3.36}$$

An intuitive way to understand the use of the wavelet and scaling filters for the MODWT estimation is via Figure 3.9 below, showing a 2-level, $J = 2$ filter bank. Each level j, where $j = 1, 2$ of the filter bank comprises two branches which are the outcome of applying a high-pass and a low-pass filter, i.e. the wavelet h_l and scaling filter g_l, respectively. In other words, the low-pass and the high-pass filter separate the entire spectrum into two equal sub-bands. Consequently, the input time series of length T after passing via the high-pass and the low-pass filter is separated into frequency band limited time series W_j and V_j of the same length.

Figure 3.9: $J = 2$ level MODWT decomposition using a filter bank.



This property of the filters resulting in output signals of the same length is also called non-decimation. In contrast, in the DWT, after each filter, a downsampling step of two is applied, resulting in output signals of length $T / 2^j$ at each level j. In order to transform the DWT $\tilde{W}_j$ and $\tilde{V}_j$ into the MODWT equivalent W_j and V_j components, someone could apply an upsampling step of two, which in the time domain resembles of introducing zeros or 'holes' between subsequent values. This is the reason that the MODWT is also known in the literature as the 'à trous' wavelet transform (Dutilleux, 1989; Holschneider, Kronland-Martinet, Morlet, & Tchamitchian, 1989), meaning in French literally 'holes'. To illustrate this and the fact that the MODWT is highly redundant, the DWT wavelet $\tilde{W}_j$ and scaling $\tilde{V}_J$ coefficients for a time series of length 2^J can be expressed as – see Percival and Walden (2000), p. 203:

$$\tilde{W}_{j,t} = 2^{j/2} W_{j,2^j t-1} \quad \text{and} \quad \tilde{V}_{J,t} = 2^{J/2} V_{J,2^J t-1} \quad \text{for } t = 1, \cdots, T_j \tag{3.37}$$

Where $T_j = T / 2^j$ is the length of the DWT scaling and wavelet coefficients at scale j. The redundancy property of the MODWT becomes easily obvious since DWT does not use all of the MODWT wavelet and scaling coefficients at each period t but rather selected observations given as $2^j t - 1$ (subscript in equation (3.37)).

Using a filter bank, the MODWT components W_j and V_j can be estimated recursively using as input the output of the previous level $j - 1$ low-pass filter (for $j = 1$ the recursion begins using the original time series Y_t itself). Thus, equations (3.30) and (3.31) can be re-written as:

$$W_{j,t} = \sum_{l=0}^{L-1} h_l V_{j-1,\left(t-2^{j-1} l\right) \bmod T} \tag{3.38}$$

$$V_{j,t} = \sum_{l=0}^{L-1} g_l V_{j-1,\left(t-2^{j-1} l\right) \bmod T} \tag{3.39}$$

In other words, assuming the entire frequency spectrum when normalised ranges from 0 to 1, i.e. $f \in [0,1)$, then W_j is band limited $1/2^{j+1} \leq |f_j| < 1/2^j$ and V_J contains the frequencies $0 \leq |f_{J+1}| < 1/2^{J+1}$.

In order to get a better understanding of how different scales-levels of decomposition correspond to wavelet components describing fluctuation with the different periods, Table 3.1 below presents for the case of monthly, quarterly and annual time series the cycles-fluctuation periods described across different scales

Table 3.1: Scales and cycles-fluctuation periods for monthly, quarterly and annual time series

Scale	Monthly	Quarterly	Annual
$j = 1$	2-4M	2-4Q	2-4Y
$j = 2$	4-8M	4-8Q/1-2Y	4-8Y
$j = 3$	8-16M	8-16Q/2-4Y	8-16Y
$j = 4$	16-32M	16-32Q/4-8Y	16-32Y
⋯	>32M	>32Q/8Y	>32Y

It has to be noted that the Haar MODWT wavelet W_j and scaling V_J coefficients satisfy the additive decomposition as in the Haar MODWT-MRA; however, this does not imply $W_j = D_j$ and $V_J = S_J$. In more detail, it holds:

$$Y_t = \sum_{j=1}^{J} W_{j,t} + V_{J,t} = \sum_{j=1}^{J} \mathcal{W}_j' W_j + \mathcal{V}_J' V_J = \sum_{j=1}^{J} D_j + S_J \tag{3.40}$$

For the specific case of the Haar MODWT, a more intuitive approach can be used to explain this transform. The wavelet coefficients represent the difference of moving averages of the original time series, while the scaling coefficients are simply a moving average of the time series itself and thus capture the slow-moving trend component. To be more specific, at each level j, the wavelet and scaling coefficients (Renaud, Starck, & Murtagh, 2003) can be estimated recursively as:

$$\begin{aligned}
&j = 1: \;\; W_{1,t} = 0.5\left(Y_t - Y_{t-1}\right), \;\; V_{1,t} = 0.5\left(Y_t + Y_{t-1}\right) \\
&j = 2: \;\; W_{2,t} = 0.5\left(V_{1,t} - V_{1,t-2}\right), \;\; V_{2,t} = 0.5\left(V_{1,t} + V_{1,t-2}\right) \\
&\quad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
&j = J: \;\; W_{J,t} = 0.5\left(V_{J-1,t} - V_{J-1,t-2^{J-1}}\right), \;\; V_{J,t} = 0.5\left(V_{J-1,t} + V_{J-1,t-2^{J-1}}\right)
\end{aligned}$$

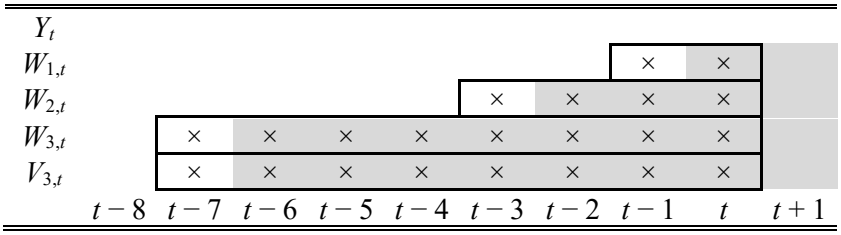
Using backward substitution, it follows that for $j \geq 1$, the wavelet and scaling coefficients can be expressed directly with respect to Y_t:

$$W_{j,t} = \frac{1}{2^j}\left(\sum_{i=0}^{2^{j-1}-1} Y_{t-i} - \sum_{i=2^{j-1}}^{2^j-1} Y_{t-i}\right) \quad \text{and} \quad V_{J,t} = \frac{1}{2^J}\sum_{i=0}^{2^J-1} Y_{t-i} \tag{3.41}$$

Where obviously, it holds that $V_{j-1,t} = W_{j,t} + V_{j,t}$ as is also documented graphically in the filter-bank approach in Figure 3.9. Using this estimation approach given by (3.41), it follows (Renaud, Starck, & Murtagh, 2003) that when a new observation is added to the time series, it is not necessary anymore to use the entire time series from the beginning to estimate the new observations of the wavelet and scaling coefficients corresponding to the new observation of the time series. Figure 3.10 below shows the observations from the original time series Y_t, denoted with ×, that have to be used to estimate the wavelet and scaling coefficients at time T. Using equation (3.41), it follows that for levels $j = 1, 2, 3$ the observations that have to be used span for each wavelet and scaling coefficient from T to $T - 1$, $T - 3$ and $T - 7$ respectively. When the new observation $T + 1$ arrives, the respective values of $W_{j,T+1}$ and $V_{J,T+1}$ can be estimated using only the observations depicted in grey shade, which allows for computational

efficiency compared to the filter-bank pyramid algorithm that would require estimating the wavelet and scaling coefficient for the entire time series again.

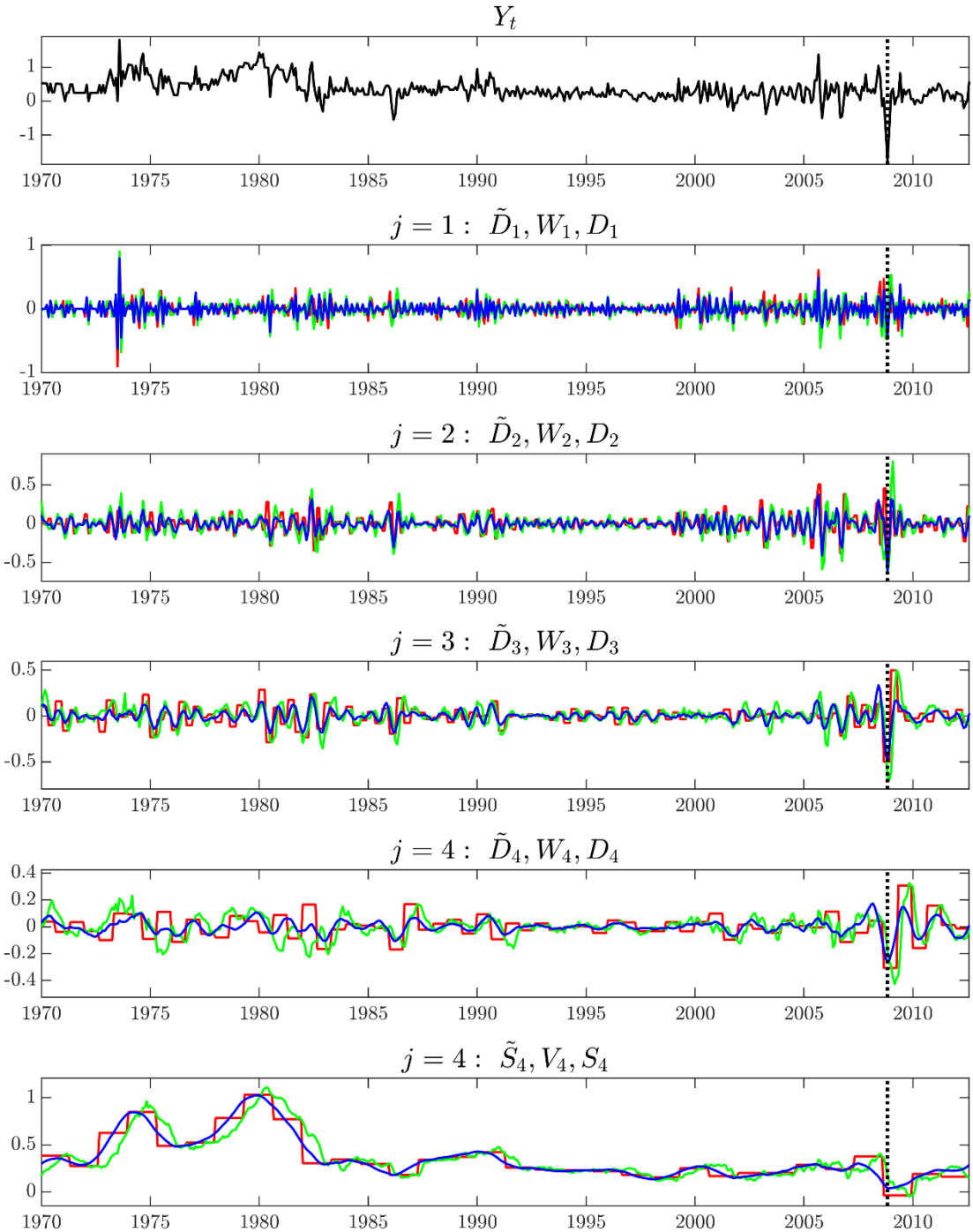
Figure 3.10: Estimation of $W_{j,t}$ and $V_{J,t}$ coefficients



Note. Estimation of $W_{j,t}$ and $V_{J,t}$ require the use of Y_t observations marked with × for each scale. The arrival of new observation Y_{t+1} and estimation of $W_{j,t+1}$ and $V_{J,t+1}$ requires only the use of Y_t observations marked in grey.

In order to understand better the impact of the different wavelet transforms, Figure 3.11 below shows the Haar DWT-MRA, MODWT and MODWT-MRA for $J = 4$. The red lines show the DWT-MRA $\tilde{D}_j$ and $\tilde{S}_J$, the green lines show the MODWT W_j and V_J, and the blue lines show the MODWT-MRA D_j and S_J. The vertical dot line during the peak of the 2008 financial crisis shows how much better the MODWT-MRA details coefficients D_4 are aligned to the original time series Y_t compared to the other two additive decompositions, i.e. the DWT-MRA ($\tilde{D}_4$) and the MODWT (W_4).

Figure 3.11: US CPI month-on-month growth rate Haar DWT-MRA, MODWT and MODWT-MRA for $J = 4$ levels



Note. First subplot Y_t shows CPI month-on-month growth rate. Subsequent subplots show the Haar DWT-MRA, MODWT and MODWT-MRA decompositions in red, green and blue lines, respectively. The vertical dot line shows the lowest value during the peak of the 2008 financial crisis. CPI is seasonally adjusted

To summarise the various discrete wavelet transforms presented in Sections 3.2.3 and 3.2.4, which are relevant to this study, Table 3.2 below compiles each one's main characteristics and the notation used to differentiate them.

Table 3.2: Discrete wavelet transforms' properties summary

Properties	DWT	DWT-MRA	MODWT	MODWT-MRA
Input Y_t length	Restricted to $T = 2^J$ Partial transform for $J_0 < J$ For arbitrary length, requires time series 'padding'.	Restricted to $T = 2^J$	Arbitrary	Arbitrary
Output	$\tilde{W}_j$ for $j = 1,\cdots, J$ of length $T_j = T / 2^j$, and $\tilde{V}_j$ of length 1	$\tilde{D}_j$ for $j = 1,\cdots, J$ $\tilde{S}_J$, all of length T	W_j for $j = 1,\cdots, J$ V_J, all of length T	D_j for $j = 1,\cdots, J$ S_J, all of length T
Additive decomposition	No	Any wavelet $Y_t = \sum_{j=1}^{J} \tilde{D}_j + \tilde{S}_J$	Only Haar wavelet $Y_t = \sum_{j=1}^{J} W_j + V_J$	Any wavelet $Y_t = \sum_{j=1}^{J} D_j + S_J$
Energy decomposition	$\|Y\| = \sum_{j=1}^{J} \|\tilde{W}_j\|^2 + \|\tilde{V}_J\|^2$	$\|Y\| = \sum_{j=1}^{J} \|\tilde{D}_j\| + \|\tilde{S}_J\|$	$\|Y\| = \sum_{j=1}^{J} \|W_j\| + \|V_J\|$	No
Used in Section(s)	3.3.1	No	3.3.2, 3.3.3	3.5

3.3 Single-frequency time series models

This section presents the three wavelet-based approaches used for multivariate time series forecasting, which are investigated in this study. In particular, Section 3.3.1 explains the role of wavelet-based time series denoising and its role in forecasting, and Section 3.3.2 explains the separate modelling of time series across the scales of the wavelet transform. Finally, Section 3.3.3 introduces the multiscale vector autoregressive model, which combines multivariate time series' wavelet transform components in a single model.

3.3.1 DWT denoising

Wavelet-based signal estimation or denoising utilizes wavelet decomposition in order to extract the signal from an observed signal corrupted by noise. The key idea lies in modifying the wavelet coefficients such that their inverse transform will result in the denoised signal. Depending on the assumed type of the underlying signal (deterministic or stochastic) and its noise (*iid* or non-*iid*), different methods have been proposed that involve thresholding, scaling or shrinking of the wavelet coefficients – see Percival and Walden (2000), Ch. 10 for a textbook treatment and Cascio (2007) for an overview of wavelet denoising in economics.

While the literature in wavelet-based signal estimation is voluminous, especially in 2-dimensional problems of imaging and information fusion, for the purpose of time series forecasting, studies using these methods are less frequent and scattered across various fields. For example, applications of linear[57] models forecasting coupled with denoising can be found in Alrumaih and Al-Fawzan (2002), Ferbar, Črešlovnik, Mojškerc, and Rajgelj (2009), Schlueter and Deuschle (2014), Herwartz and Schlüter (2017) and Bruzda (2020). Although not examined in this study, Bayesian approaches have also been proposed for wavelet-based denoising, allowing to impose shrinkage on the wavelet coefficients and their associated noise via the introduction of appropriate priors – see Chipman, Kolaczyk, and McCulloch (1997), Crouse, Nowak, and Baraniuk (1998), Abramovich, Sapatinas, and Silverman (1998), and Vidakovic (1998).

For the purpose of this study, denoising will take the form of thresholding for the cases of *iid* noise. Before proceeding further, it is useful to define the thresholding functions for hard, $F_h(x)$ and soft, $F_s(x)$ thresholding for input x and threshold value τ:

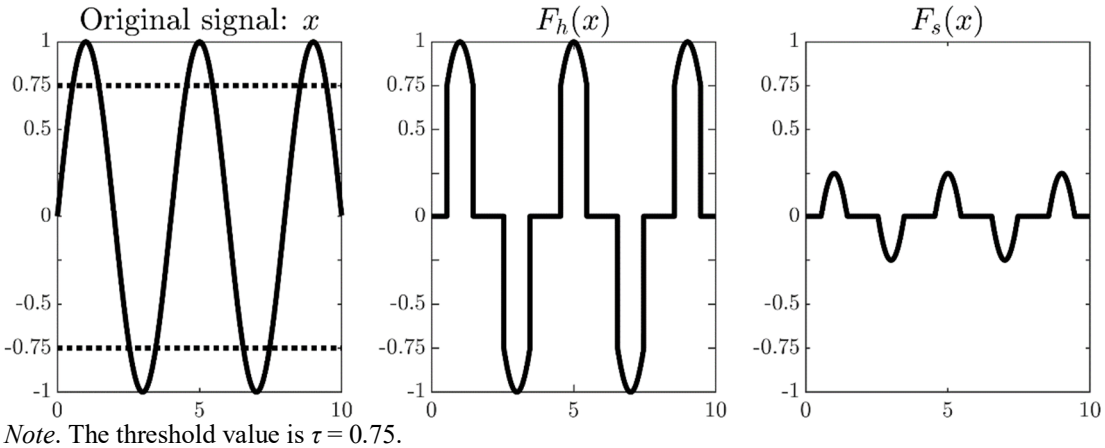
$$F_h(x) = \begin{cases} x, & |x| > \tau \\ 0, & |x| \le \tau \end{cases} \quad \text{and} \quad F_s(x) = \begin{cases} \operatorname{sgn}(x)(|x| - \tau), & |x| > \tau \\ 0, & |x| \le \tau \end{cases} \tag{3.42}$$

Where $\operatorname{sgn}(x)$ is the signum function taking the value of +1 if $x > 0$, −1 if $x < 0$, and 0 if $x = 0$.

Figure 3.12 below demonstrates, for a simple case, the output of the hard and soft thresholding functions for the threshold value $\tau = 0.75$.

[57] For studies with non-linear models see Lotrič (2004), Lotrič and Dobnikar (2005), Li, He, Lai, and Zou (2014), and Xu, Han, and Lin (2018).

Figure 3.12: Hard, $F_h(x)$ and soft, $F_s(x)$ thresholding functions



Note. The threshold value is $\tau = 0.75$.

To formalise ideas, we assume that the 1-dimensional measured time series Y_t contains some noise ε_t:

$$Y_t = G_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right) \tag{3.43}$$

It has to be noted that noise in the context of the macroeconomic time series should be interpreted with caution. For time series that are subject to revisions like, such as the GDP, noise could be interpreted indeed as a reflection of measurement errors, especially in the context of real-time data. However, a more intuitive interpretation, even for time series that are subject to revisions, is that noise ε_t actually captures transitory (temporary) economic shocks, while G_t represents an equilibrium level or long-term trend of a variable.

The denoising procedure consists of the following three steps:

1. Apply J_0 levels partial DWT using the partial transform matrix $\tilde{\mathcal{W}}^p$ and obtain:

$$\tilde{W} = \tilde{\mathcal{W}}^p Y = \tilde{\mathcal{W}}^p G + \tilde{\mathcal{W}}^p \varepsilon = \tilde{G} + \tilde{\varepsilon}, \tag{3.44}$$

Where for each submatrix $\tilde{W}_j$ of $\tilde{W}$ at each point *t,* it follows:

$$\tilde{W}_{j,t} = \tilde{G}_{j,t} + \tilde{\varepsilon}_{j,t}, \quad j = 1, \cdots, J_0, \quad t = 1, \cdots, T/2^j \tag{3.45}$$

2. Specify a universal threshold value τ or level j dependent threshold τ_j and apply it to each scale wavelet coefficients $\tilde{W}_j$ (the scaling coefficients $\tilde{V}_{J_0}$ remain intact) using a hard, $F_h(x)$ or soft, $F_s(x)$ thresholding function as defined above in equation (3.42). The resulting wavelet coefficients which have passed via the thresholding function are denoted with a superscript *t*, i.e. $\tilde{W}_j^t$ to differentiate them from the original wavelet coefficients $\tilde{W}_j$.

3. Form the $\tilde{W}^t = \left[\tilde{W}_1^t, \cdots, \tilde{W}_{J_0}^t, \tilde{V}_{J_0}\right]$, and via the inverse DWT transform defined in equation (3.18), estimate the denoised time series $\hat{G}$ as:

$$\hat{G} = \left(\tilde{\mathcal{W}}^p\right)' \tilde{W}^t \tag{3.46}$$

The key element in the above procedure is to estimate the threshold value τ_j or τ such that the estimate $\hat{G}$ of the true and unobserved time series minimises the risk, $R\left(\hat{G},G\right)$ defined as:

$$R\left(\hat{G},G\right)=E\left\|\hat{G}-G\right\|^2 \tag{3.47}$$

Donoho and Johnstone (1994) showed that the threshold should be:

$$\hat{\tau}=\left(2\hat{\sigma}_\varepsilon^2\log\left(T\right)\right)^{1/2} \tag{3.48}$$

Where $\hat{\sigma}_\varepsilon^2$ is the median absolute deviation of the $j = 1$ level wavelet coefficients $\tilde{W}_1$ of length $T/2$ scaled by an appropriate factor:

$$\hat{\sigma}_\varepsilon^2=\frac{\text{median}\left(\left|\tilde{W}_{1,1}\right|,\left|\tilde{W}_{1,2}\right|,\cdots,\left|\tilde{W}_{1,T/2}\right|\right)}{0.6745} \tag{3.49}$$

Another popular approach follows Donoho and Johnstone (1995), who showed that the threshold can be estimated by minimizing Stein's unbiased risk estimator (SURE) due to Stein (1981). Assuming $\tilde{W}\sim\mathcal{N}\left(\tilde{G},\sigma_{\tilde{\varepsilon}}^2 I_T\right)$ as a reformulation of equation (3.44) and the special case $\hat{G}=G+f\left(G\right)$, where f is differentiable, then under the Law of Large numbers and using soft thresholding, $R\left(\hat{G},G\right)$ will be equal to:

$$SURE\left(\tilde{W},\tau\right)=\sigma_\varepsilon^2T-2\sigma_\varepsilon^2\#\left\{t:\left|\tilde{W}_t\right|\le\tau\right\}+\sigma_\varepsilon^2\sum_{t=1}^{T}\left(\min\left(\left|\tilde{W}_t\right|,\tau\right)\right)^2 \tag{3.50}$$

Where $\#\{\cdot\}$ denotes the cardinality of the set.

It follows that the threshold $\hat{\tau}$ should minimise $SURE\left(\tilde{W},\tau\right)$:

$$\hat{\tau}=\underset{0\le\tau\le\hat{\sigma}_\varepsilon\sqrt{2\log(T)}}{\arg\min}\ SURE\left(\tilde{W},\tau\right) \tag{3.51}$$

Turning now to using wavelet-based denoising for the purpose of forecasting, the procedure is straightforward. Assuming a multivariate time series $Y_t=\left[Y_{1,t},Y_{2,t},\cdots,Y_{N,t}\right]'$, wavelet-based denoising is applied as described previously, i.e. for each of the univariate time series $Y_{n,t}$, for $n = 1,\cdots,N$ separately, resulting in the denoised time series $Y_{n,t}^d$. The corresponding multivariate denoised time series $Y_t^d=\left[Y_{1,t}^d,Y_{2,t}^d,\cdots,Y_{N,t}^d\right]'$ instead of the original Y_t can now be used with standard BVAR models for forecasting.

3.3.2 MODWT separate scale BVAR

Wavelet-based multiscale/multiresolution analysis has been used on numerous occasions in the literature for the purpose of forecasting. The underlying idea consists of modelling and

forecasting separately the detail and smooth coefficients and finally estimating the forecast of the original time series by their summation. The literature following this approach is mainly focused on univariate times and follows frequentist estimation. Soltani, Boichu, Simard, and Canu (2000), Conejo, Plazas, Espinola, and Molina (2005), Tan, Zhang, Wang, and Xu (2010), Schlueter and Deuschle (2014), Kriechbaumer, Angus, Parsons, and Casado (2014), Zhu, Wang, and Fan (2014) and Long and Yan (2016) model (long memory) time series with MRA detail and smooth coefficients using separate AR(I)MA processes and potentially coupled with GARCH models. More recently, Zhang, Gençay, and Yazgan (2017) and Uddin, Gençay, Bekiros, and Sahamkhadam (2019) used multiresolution analysis with automatic-ARMA modelling[58] to find the correct specified ARMA models at each scale j and find evidence of superior forecasting versus forecasting directly the original time series. In a slightly different approach, Wong, Ip, Xie, and Lui (2003) and Schlueter and Deuschle (2014) use wavelets to decompose time series into a trend, periodic/seasonal and irregular components and then model them separately.

Extending to multivariate time series, Xu and Niimura (2004), in a forecasting application, suggest for a bivariate time series of electricity price and load to use VAR modelling for the smooth coefficients whilst maintaining univariate modelling for the detail coefficients due to their low correlation. Outside a forecasting context, using wavelets transform for multivariate series trying to explain the underlying dynamics, Boubaker and Raza (2017), Khalfaoui (2018) and Gupta, Das, Hasim, and Tiwari (2018) assume a priori that the bivariate financial time series they examine maintain their interdependence at all the levels of the MODWT-MRA decomposition. The former two papers of Boubaker and Raza (2017) and Khalfaoui (2018) model the detail coefficients at each level using VARMA models with further specifications of conditional variance and examine the interdependence of oil price and BRICs stock market or gold at different scales, respectively. Similarly, Gupta, Das, Hasim, and Tiwari (2018) use a MODWT-MRA bivariate VAR to examine at different scales the causality relationship between stock returns and trading volumes.

All of the studies following independent modelling of detail and smooth coefficients so far, either for the univariate or multivariate models, have used frequentist estimation techniques. In that sense, using Bayesian techniques for the estimation of VAR models across each scale poses as a plausible alternative due to its inherent advantage of imposing shrinkage to the estimated parameters and controlling cross-variable effects.

To be more formal, supposing a multivariate time series $Y_t = \left[Y_{1,t}, Y_{2,t}, \cdots, Y_{N,t} \right]'$, and then applying the Haar MODWT, each time series can be decomposed as:

$$Y_{n,t} = \sum_{j=1}^{J} W_{n,j,t} + V_{n,J,t}$$

Now, the corresponding multivariate time series of the wavelet and scaling coefficients at each level $j = 1, \cdots, J$ can be formed as:

$$W_{j,t} = \left[W_{1,j,t}, W_{2,j,t}, \cdots, W_{N,j,t} \right]' \quad \text{and} \quad V_{J,t} = \left[V_{1,J,t}, V_{2,J,t}, \cdots, V_{N,J,t} \right]' \tag{3.52}$$

[58] Non-linear models like neural networks have also been employed both for univariate and multivariate time series forecasting in Bekiros and Marcellino (2013) and Saâdaoui and Rabbouch (2014).

Using the newly formed multivariate time series $W_{j,t}$ and $V_{j,t}$, a VAR model can be formed at each decomposition level j to forecast the multivariate wavelet and scaling coefficients time series jointly. To be more concrete for the wavelet coefficients $W_{j,t}$ at each scale $j = 1,\cdots, J$ a VAR is formed:

$$W_{j,t} = c_j + \sum_{p=1}^{P_j} B_{j,p} W_{j,t-p} + e_{j,t}, \quad e_{j,t} \sim \mathcal{N}\left(0, \Sigma_j\right) \tag{3.53}$$

Similarly, for the scaling coefficients $V_{j,t}$, a VAR is formed as follows:

$$V_{J,t} = c_{J+1} + \sum_{p=1}^{P_{J+1}} B_{J+1,p} V_{J,t-p} + e_{J+1,t}, \quad e_{J+1,t} \sim \mathcal{N}\left(0, \Sigma_{J+1}\right) \tag{3.54}$$

Using the standard VAR setting, c_j are of $N \times 1$ dimensions, $B_{j,p}$ are $N \times N$ and the innovations $e_{j,t}$ are $N \times 1$ vectors normally distributed.

Alternatively, for the wavelet coefficients $W_{j,t}$ up to a scale $j \leq j^*$, which capture the high-frequency fluctuations of the time series, the VAR model can be further extended by the inclusion of standard stochastic volatility specification as in Cogley and Sargent (2005):

$$W_{j,t} = c_j + \sum_{p=1}^{P_j} B_{j,p} W_{j,t-p} + e_{j,t}, \quad e_{j,t} \sim \mathcal{N}\left(0, \Sigma_{j,t}\right) \tag{3.55}$$

Where $\Sigma_{j,t} = \Sigma_t = A^{-1} L_t A^{-1\prime}$, with matrix A being lower diagonal[59], and the natural logarithm of the diagonal elements of $L_{t,n,n} = h_{n,t}$ is defined as:

$$\ln h_{n,t} = \ln h_{n,t-1} + \eta_{n,t}, \quad \eta_{n,t} \sim \mathcal{N}\left(0, \Phi_{n,n}\right) \tag{3.56}$$

The prior used for the coefficients B for each separate scale homoscedastic or heteroscedastic BVAR is a conventional Minnesota-type prior. Dropping the subscript j for notational simplicity, it follows that B for scale-level of decomposition j in a vectorised form follows a normal distribution:

$$\text{vec}\left(B\right) \sim \mathcal{N}\left(\underline{\text{M}}_B, \underline{\Omega}_B\right) \tag{3.57}$$

The first moment of coefficients B prior is set as:

$$E\left[\left(B_p\right)_{m,n}\right] = 0, \ \forall m, n \tag{3.58}$$

The variance of the coefficients' prior is set as:

$$\text{var}\left[\left(B_p\right)_{m,n}\right] = \begin{cases} \dfrac{\lambda_1^{\,2}}{p^{\lambda_3}}, & m = n \\ \dfrac{\lambda_1^{\,2}\lambda_2^{\,2}}{p^{\lambda_3}} \dfrac{\sigma_m^2}{\sigma_n^2}, & \text{otherwise} \end{cases} \tag{3.59}$$

The hyperparameter values are discussed in Section 3.4.2.

[59] More details regarding matrix A are provided in Section 3.3.3.2.

Moreover, for the homoscedastic BVAR, the covariance matrix Σ follows an inverse Wishart distribution with scale matrix $\underline{S}$ and degrees of freedom $\underline{d}$:

$$\Sigma \sim \mathcal{IW}\left(\underline{S}_{\Sigma}, \underline{d}_{\Sigma}\right) \tag{3.60}$$

Finally, for the heteroscedastic BVARs employing stochastic volatility, the innovations of the stochastic volatility in equation (3.56) follow an inverse Gamma prior:

$$\Phi_{n,n} \sim \mathcal{IG}\left(\underline{S}_{\Phi_{n,n}}, \underline{d}_{\Phi_{n,n}}\right) \tag{3.61}$$

More details for the stochastic volatility estimation are provided in Section 3.3.3.2 and Appendix C.1.1 Stochastic volatility estimation.

In order to obtain forecasts for the original multivariate time series Y_t, the MODWT separate scale BVAR forecasts have to be summed across each forecasted period h and Gibbs sampler draw (r):

$$\hat{Y}_{T+h}^{(r)} = \begin{bmatrix} \hat{Y}_{1,T+h}^{(r)} \\ \hat{Y}_{2,T+h}^{(r)} \\ \vdots \\ \hat{Y}_{N,T+h}^{(r)} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{J} \hat{W}_{1,j,T+h}^{(r)} + \hat{V}_{1,J,T+h}^{(r)} \\ \sum_{j=1}^{J} \hat{W}_{2,j,T+h}^{(r)} + \hat{V}_{2,J,T+h}^{(r)} \\ \vdots \\ \sum_{j=1}^{J} \hat{W}_{N,j,T+h}^{(r)} + \hat{V}_{N,J,T+h}^{(r)} \end{bmatrix} \tag{3.62}$$

The Haar MODWT-MRA additive decomposition was also used by simply substituting the wavelet, $W_{j,t}$ and scaling, $V_{J,t}$ coefficients with the detail, $D_{j,t}$ and smoothing, $S_{J,t}$ coefficients. However, the results were not favourable and are not presented in the relevant Section 3.4.4. More details about the prior and the specifications used for the BVAR models at each level j can be found in Section 3.4.2.

3.3.3 Multiscale BVAR

The third approach using wavelet decomposition for forecasting is based on the idea of a multiscale autoregressive model (MAR) introduced for univariate time series in Renaud, Starck, and Murtagh (2003) – see also Murtagh, Starck, and Renaud (2004), Benaouda, Murtagh, Starck, and Renaud (2006), and Aminghafari and Poggi (2007). MAR model, equation (3.63), is an autoregressive model where the time series Y_t instead of being explained by the lagged values of itself, is rather explained by its Haar MODWT lagged wavelet and scaling coefficients. The particular advantage of the MAR model consists of capturing the autocorrelation properties of the time series efficiently and is particularly robust to misspecification, as well as long-range dependence and fractionally integrated time series.

To be more formal, the originally proposed MAR model for a zero mean univariate time series is defined as follows:

$$Y_t = \sum_{j=1}^{J}\sum_{p=1}^{P_j} \beta_{j,p} W_{j,t-1-2^j(p-1)} + \sum_{p=1}^{P_{J+1}} \beta_{J+1,p} V_{J,t-1-2^J(p-1)} + e_t, \quad e_t \sim \mathcal{N}\left(0,\sigma^2\right) \tag{3.63}$$

Where Y_t, $W_{j,t}$, $V_{J,t}$, $\beta_{j,p}$ and e_t are all scalars.

What is crucial for the MAR model is to understand its sparsity and the exact lag pattern for the wavelet and scaling coefficients, i.e. rather than including the lags in a continuous fashion $t - p$, they are in a discontinuous fashion $t - 1 - 2^j (p - 1)$, omitting 2^j lags at each scale j. This becomes more clear in Figure 3.13 below. For example, for $P_j = 2$ for $j = 1,\cdots, J + 1$, time series observation Y_T depends, for $j = 1$ only on $W_{1,T-1}$ and $W_{1,T-3}$ omitting $W_{1,T-2}$.

Figure 3.13: MAR model lagged wavelet and scaling coefficients

	$T-9$	$T-8$	$T-7$	$T-6$	$T-5$	$T-4$	$T-3$	$T-2$	$T-1$	T
Y_t										×
$W_{1,t}$							×		×	
$W_{2,t}$					×				×	
$W_{3,t}$	×								×	
$V_{3,t}$	×								×	

Note. Lagged wavelet and scaling coefficients marked with × represent those that need to be used to estimate observation Y_T in the MAR model.

The rationale behind using the lags in a discontinuous fashion, i.e. $t - 1 - 2^j (p - 1)$, is to achieve a parsimonious representation with the least possible lagged terms that provide sufficient information to reconstruct the time series. The selected subsets of the lagged wavelet $W_{j,t-1-2^j(p-1)}$ and scaling $V_{J,t-1-2^J(p-1)}$ coefficients are part of the DWT orthogonal transform basis. This becomes evident via equation (3.37) repeated below:

$$\tilde{W}_{j,t} = 2^{j/2} W_{j,2^j t-1} \quad \text{and} \quad \tilde{V}_{j,t} = 2^{J/2} V_{j,2^j t-1} \quad \text{for } t = 1,\cdots,T_j \quad \text{where } T_j = T/2^j$$

For example, for $j = 1$, the last two observations of the DWT wavelet coefficients would be for $t = T / 2^1 = T / 2$ and $t = T / 2 - 1$. Using the equation above would yield:

$$\tilde{W}_{1,T/2} = 2^{j/2} W_{1,T-1} \quad \text{and} \quad \tilde{W}_{1,T/2-1} = 2^{j/2} W_{1,2^1\left(T/2^1-1\right)-1} = 2^{j/2} W_{1,T-3}$$

In Figure 3.13, it is clear that the MODWT wavelet coefficients for $j = 1$ are $W_{1,T-1}$ and $W_{1,T-3}$, as shown above.

Renaud, Starck, and Murtagh (2003) assume an AR(k) process and prove that the parameters of the MAR model asymptotically converge to values that result in the same predictions as would be the forecasts under the true AR process. However, the proof holds under the condition that the minimum number of lags P_j at each level j satisfies the condition $P_j \geq k / 2^j$, where k is the maximum lag in the AR(k) process. In terms of estimation, the proposed model was estimated in a frequentist approach via OLS or MLE; nonetheless, extending the model below in a multivariate setting due to coefficients proliferation, Bayesian estimation techniques are used.

3.3.3.1 Homoscedastic MBVAR

The extension of the MAR model for multivariate time series[60] using a vector autoregressive model (MVAR) using the N-dimensional vectors of Y_t, $W_{j,t}$ and $V_{J,t}$ as defined in equation (3.52) would yield the following model:

$$Y_t = B_c + \sum_{j=1}^{J}\sum_{p=1}^{P_j} B_{j,p} W_{j,t-1-2^j(p-1)} + \sum_{p=1}^{P_{J+1}} B_{J+1,p} V_{J,t-1-2^J(p-1)} + e_t, \quad e_t \sim \mathcal{N}(0,\Sigma) \tag{3.64}$$

Where B_c is $N \times 1$, $B_{j,p}$ is $N \times N$, and the innovations e_t is an $N \times 1$ vector normally distributed.

Thus, stacking all the MVAR coefficients in a single matrix B of dimensions $\left(1+N\sum_{j=1}^{J+1}P_j\right)\times N$ would result in the following:

$$B = \left[B_c, B_{1,1}, \cdots, B_{1,P_1}, \cdots, B_{J,1}, \cdots, B_{J,P_J}, B_{J+1,1}, \cdots, B_{J+1,P_{J+1}}\right]' \tag{3.65}$$

Obviously, for the simplified case where $P_j = P$, for all $j = 1,\cdots, J+1$, then matrix B would be of dimensions $(1 + NP(J+1)) \times N$.

Stacking all the right-hand side regressors in equation (3.64) in the row vector X_t of dimensions $1\times\left(1+N\sum_{j=1}^{J+1}P_j\right)$ would result in the following:

$$X_t = \left[1, W'_{1,t-1}, \cdots, W'_{1,t-1-2^1(P_1-1)}, \cdots, W'_{J,t-1}, \cdots, W'_{J,t-1-2^J(P_J-1)}, V'_{J,t-1}, \cdots, V'_{1,t-1-2^J(P_{J+1}-1)}\right] \tag{3.66}$$

The MVAR model can be rewritten in a compact form as:

$$Y_t = B'X'_t + e_t \tag{3.67}$$

Regarding coefficients B of the MBVAR, they follow two different prior specifications. The first approach follows a Minnesota family-inspired type prior further being adjusted to accommodate separate shrinkage on the wavelet and scaling coefficients and is discussed in Section 3.3.3.1.1. The second specification uses the Stochastic Search Variables Selection (SSVS) proposed by George and McCulloch (1993); George, Sun, and Ni (2008) and is discussed in Section 3.3.3.1.2.

The innovations variance Σ for both B prior specifications follows an inverse Wishart which is the standard case for a similar type of model:

$$\Sigma \sim \mathcal{IW}\left(\underline{S}_\Sigma, \underline{d}_\Sigma\right) \tag{3.68}$$

Where $\underline{S}$ is the prior scale matrix, and $\underline{d}$ are the degrees of freedom.

[60] Cekic, Grandjean, and Renaud (2019) have proposed recently a bivariate TVP-MVAR model to investigate Granger causality for brain signals.

3.3.3.1.1 Minnesota family-type prior elicitation

Before proceeding further with the MBVAR posterior estimation, it is important to present more details about the prior elicitation for the B coefficients, which follow a normal distribution.

$$B \sim \mathcal{N}\left(\underline{\mathrm{M}}_B, \underline{\Omega}_B\right) \tag{3.69}$$

As a starting point, the prior is based on conventional Minnesota-type prior further adjusted for the MBVAR model. In that sense, it is a type of prior which assumes which regressors in each variable's equation are more important for the purpose of forecasting. To be more specific, the prior has a global shrinkage hyperparameter λ_1 which controls the overall tightness for all parameters. In addition, similar to the Minnesota type prior, there is a hyperparameter λ_2 controlling the cross variables' shrinkage, as well as a lag decay parameter λ_3 controlling the impact of the more distant lags. The constant follows a diffuse prior, set by the hyperparameter λ_0.

The main difference from the Minnesota prior is the absence of the so-called scaling factor σ_m / σ_n for the coefficients of variables *n,* other than its variables' *m* own lags in equation *m.* The reasoning for this is based on the fact that as the scale *j* increases, wavelet and scaling coefficients become more and more smooth and thus, σ_n would decrease steadily, and consequently, this effect can be captured a priori by relevant hyperparameters' values. It is true, however, that for a given scale *j,* the respective wavelet/scaling coefficients are of comparable behaviour, and the scaling factor σ_m / σ_n could be maintained and have an interpretation similar to standard BVAR literature. This prior's specification was tested, but its impact was found to be negligible.

Since the regressors of the model are not any more the time series themselves but the wavelet and scaling coefficients that the time series can be decomposed into, the proposed prior introduces a new hyperparameter $\tilde{\lambda}_j$. Hyperparameter $\tilde{\lambda}_j$ essentially controls a-priori the importance of the wavelets and scaling coefficients at each scale *j*. To formalise ideas, the relevant prior's equation is defined as follows.

The first moment of coefficients B prior is set as:

$$E\left[\left(B_c\right)_{n,1}\right] = E\left[\left(B_{j,p}\right)_{m,n}\right] = 0, \quad \forall j, p, m, n \tag{3.70}$$

The standard deviation of the coefficients' prior is set as:

$$\left[\mathrm{var}\left(B_c\right)_{n,1}\right]^{1/2} = \lambda_0$$

$$\left[\mathrm{var}\left(\left(B_{j,p}\right)_{m,n}\right)\right]^{1/2} = \begin{cases} \lambda_1 \dfrac{\tilde{\lambda}_j}{p^{\lambda_3}}, \; m = n, \;\; j = 1, \cdots, J, J+1 \\ \lambda_1 \lambda_2 \dfrac{\tilde{\lambda}_j}{p^{\lambda_3}}, \; m \neq n, \;\; j = 1, \cdots, J, J+1 \end{cases} \tag{3.71}$$

Where $\tilde{\lambda}_j$ is a hyperparameter controlling the tightness for the *j*-scale wavelet and scaling coefficients. This hyperparameter essentially allows imposing the belief on which of the

wavelet or scaling coefficients dominate the time series and provide more informational content for the purpose of forecasting the original time series.

Having as a motivation that the high-frequency components represented with small scale j fade out over time and that the lowest frequency components summarise the long-term behaviour of the time series by setting appropriate values to $\tilde{\lambda}_j$ it becomes feasible to shrink more towards zero the high-frequency components (small j) while imposing a loose prior on the low-frequency components (large j and $J+1$ scaling coefficients). This intuition is also supported by the estimation of the marginal likelihood of the models. In more detail, using Chib (1995) marginal likelihood approximation method (see Appendix C.1.2 Marginal likelihood estimation for more details), it becomes evident that a less tight prior for larger scales j is also supported by the data. It has to be noted that these values are not the optimal values with respect to a global optimisation problem of the marginal likelihood for all the hyperparameters. Showing that the marginal likelihood $p(Y \mid \mathcal{M})$ increases by letting $\tilde{\lambda}_j$ to have an increasing pattern rather than a decreasing or a constant one provides evidence that the prior belief is in the proper direction.

Table 3.3 below presents the log marginal likelihood of an MBVAR for constant, increasing and decreasing patterns of the hyperparameter $\tilde{\lambda}_j$. The sensitivity of the marginal likelihood is tested against different specifications of the hyperparameters $\tilde{\lambda}_j$, while the remaining hyperparameters λ_0 - λ_3 are kept constant with the conventional values of 10^3, 0.1, 0.5 and 1 used in the literature on BVARs.

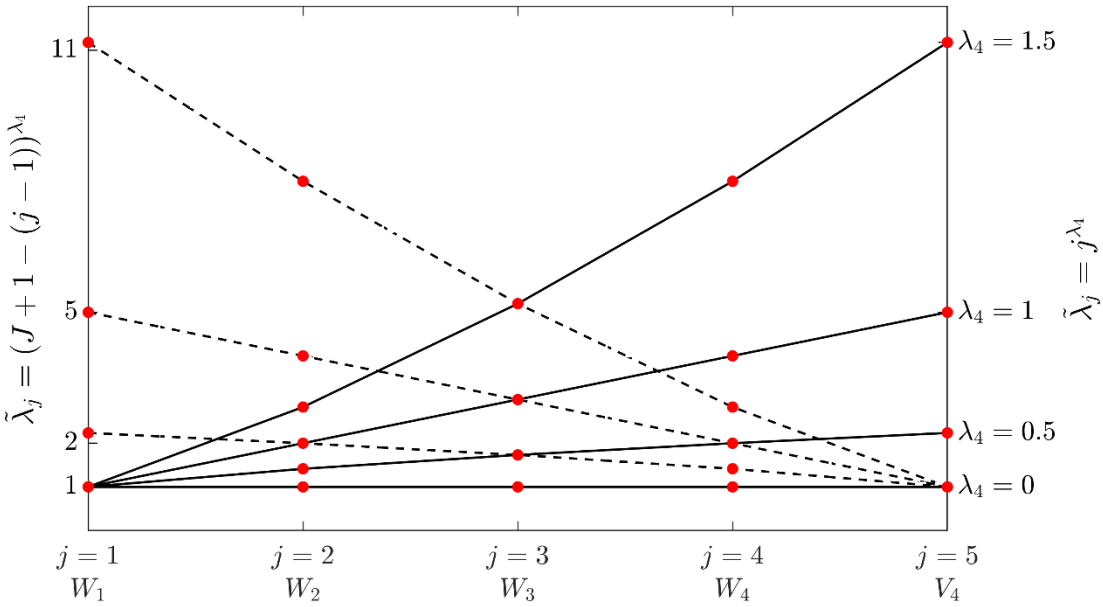
Table 3.3: Wavelet and scaling coefficients hyperparameters

Model $\mathcal{M}_i$	$\tilde{\lambda}_j$	Log marg. likelihood
1	$\tilde{\lambda}_j = j^{\lambda_4},\ \lambda_4 = 0$	25340.8
2	$\tilde{\lambda}_j = j^{\lambda_4},\ \lambda_4 = 0.5$	83188.4
3	$\tilde{\lambda}_j = \left[J+1-(j-1)\right]^{\lambda_4},\ \lambda_4 = 0.5$	28220.4
4	$\tilde{\lambda}_j = j^{\lambda_4},\ \lambda_4 = 1$	201552.2
5	$\tilde{\lambda}_j = \left[J+1-(j-1)\right]^{\lambda_4},\ \lambda_4 = 1$	31769.7
6	$\tilde{\lambda}_j = j^{\lambda_4},\ \lambda_4 = 1.5$	384163.7
7	$\tilde{\lambda}_j = \left[J+1-(j-1)\right]^{\lambda_4},\ \lambda_4 = 1.5$	34296.8

Note. Log marginal likelihood for different specifications of $\tilde{\lambda}_j$ with the remaining hyperparameters λ_0 - λ_3 kept constant.

In the first row of the table, $\mathcal{M}_1$ keeps the same shrinkage across all scales *j*. For the remaining models, every two rows, there is a pattern of increasing loosening of the prior in line with *j* and the exact opposite pattern with an increasing tightening of the prior with respect to scale *j*. To get a better understanding of the different patterns in hyperparameter $\tilde{\lambda}_j = j^{\lambda_4}$. Figure 3.14 below shows visually how the hyperparameter values change across all scales *j* for different values of λ_4 for each model $\mathcal{M}_i$. Dash lines show decreasing patterns over scale *j*, defined as $\tilde{\lambda}_j = \left[J+1-(j-1)\right]^{\lambda_4}$ in the left y-axis, while continuous lines show increasing patterns over scale *j*, defined as $\tilde{\lambda}_j = j^{\lambda_4}$ in the right y-axis. The symmetric (increasing versus decreasing) patterns with respect to the y-axis have λ_4 hyperparameter values of 0 (constant), 0.5 (concave), 1 (linear) and 1.5 (convex) shown on the right y-axis. As a reminder, scale $J + 1$ refers to the scaling coefficients V_J.

Figure 3.14: MBVAR model wavelet and scaling hyperparameters shrinking functions



Note. Left axis $\tilde{\lambda}_j = \left[J+1-\left(j-1\right)\right]^{\lambda_4}$ refers to decreasing patterns with dash lines. Right axis $\tilde{\lambda}_j = j^{\lambda_4}$ refers to increasing patterns with continuous lines. Values of λ_4 for the symmetric patterns across the y-axis are shown on the right axis.

It becomes easily evident that across the various shrinkage specifications tested, log marginal likelihood is larger when shrinkage is stronger for the high-frequency components and gradually becomes looser. The exact way of decreasing tightness with level *j*, i.e. if it is linear, concave or convex, is determined with respect to providing the best forecasting performance and is not examined as a global optimisation problem with respect to the marginal likelihood.

To formalise this finding, using proper model comparison in a Bayesian context, the Bayes factors can be used. To begin with, by a simple application of the Bayes rule, the posterior model's $\mathcal{M}_i$ probability given data Y is given as:

$$p\left(\mathcal{M}_i|Y\right) = \frac{p\left(Y|\mathcal{M}_i\right)p\left(\mathcal{M}_i\right)}{p\left(Y\right)} \tag{3.72}$$

Where $p(\mathcal{M}_i|Y)$ is the model's posterior, $p(Y|\mathcal{M}_i)$ is the model's marginal likelihood, $p(\mathcal{M}_i)$ is the model's prior and $p(Y)$ is the data likelihood.

In order to compare two models $\mathcal{M}_1$ and $\mathcal{M}_0$ the ratio of their posterior probabilities can be estimated. Using equation (3.73), the so-called posterior odds ratio is now defined as:

$$PO_{1,0} = \frac{p\left(Y|\mathcal{M}_1\right)p\left(\mathcal{M}_1\right)}{p\left(Y|\mathcal{M}_0\right)p\left(\mathcal{M}_0\right)} \tag{3.73}$$

By further assuming equal prior models probabilities, i.e. $p(\mathcal{M}_1) = p(\mathcal{M}_0)$, the posterior odds ratio simplifies essentially to the so-called Bayes factor $BF_{1,0}$:

$$BF_{1,0} = \frac{p\left(Y\middle|\mathcal{M}_1\right)}{p\left(Y\middle|\mathcal{M}_0\right)} \tag{3.74}$$

Equation (3.74) can be useful in comparing models pairwise since the log marginal likelihood, $\ln p(Y|\mathcal{M})$, has already been estimated and is available in Table 3.3. Since the estimation of the marginal likelihood from the log marginal likelihood is not possible for such large values, I resort to Kass and Raftery (1995), Section 3.2, where they provide some ranges of the quantity $2\ln BF_{1,0}$ and the corresponding qualitative characterisation for a hypothesis testing against the null of the model $\mathcal{M}_0$. As a result, equation (3.76) can be written as:

$$2\ln BF_{1,0} = 2\ln\frac{p\left(Y\middle|\mathcal{M}_1\right)}{p\left(Y\middle|\mathcal{M}_0\right)} = 2\left(\ln p\left(Y\middle|\mathcal{M}_1\right) - \ln p\left(Y\middle|\mathcal{M}_0\right)\right) \tag{3.75}$$

To make use of the log marginal likelihood presented in Table 3.3, the various models are compared in 2 ways. Firstly, they are compared for each pair of decreasing versus increasing tightness patterns, i.e. $2\ln BF_{2,3}$, $2\ln BF_{4,5}$, and $2\ln BF_{6,7}$. Secondly, the decreasing linear pattern tightness with $\tilde{\lambda}_j = j^{\lambda_4}$ for $\lambda_4 = 1$ of the model $\mathcal{M}_4$ is compared with the constant tightness across all scales ($2\ln BF_{4,1}$) and the increasing tightness models, i.e. $2\ln BF_{4,3}$, $2\ln BF_{4,5}$ and $2\ln BF_{4,7}$. It is obvious that for all cases, these values are in the range of thousands, much greater than 10, for which Kass and Raftery (1995) provide the characterisation of 'very strong' evidence against the null hypothesis of the model in the denominator in equation (3.75). It is also obvious that across all models, the decreasing tightness pattern of the model $\mathcal{M}_6$ with λ_4 =1.5 presents the largest marginal likelihood and equivalently larger Bayes factor vis-à-vis all of the other models. Nonetheless, for the empirical application part in Section 3.4.2, the linear decreasing tightness (increasing 'looseness') pattern was selected because of resulting in better forecast accuracy.

Turning now to the posterior draws of coefficients *B,* the standard approach is to sample directly from its posterior for the entire model:

$$B \sim \mathcal{N}\left(\overline{\mathrm{M}}_B, \overline{\Omega}_B\right) \tag{3.76}$$

Where $\overline{\Omega}_B = \left(\underline{\Omega}_B^{-1} + \Sigma^{-1}\otimes\left(X'X\right)\right)^{-1}$, $\overline{\mathrm{M}}_B = \overline{\Omega}_B\left(\underline{\Omega}_B^{-1}\underline{\mathrm{M}}_B + \left(\Sigma^{-1}\otimes X'\right)vec\left(Y\right)\right)$.

Consequently, the iteration (r) from the Gibbs sampler is based on the following two blocks:

1. $p\left(B^{(r)}\middle|\Sigma^{(r-1)}, Y\right)$
2. $p\left(\Sigma^{(r)}\middle|B^{(r)}, Y\right)$

Where the second block of the Gibbs sampler draws the variance-covariance matrix Σ from its conditional posterior distribution:

$$\Sigma \sim \mathcal{IW}\left(\overline{S}_\Sigma, \overline{d}_\Sigma\right) \tag{3.77}$$

Where $\overline{S}_\Sigma = \underline{S}_\Sigma + \left(\tilde{Y} - \tilde{X}B\right)'\left(\tilde{Y} - \tilde{X}B\right)$ and $\overline{d}_\Sigma = \underline{d}_\Sigma + T$.

More details on an equation-by-equation estimation of the posterior for the *B* coefficients are presented in Section 3.3.3.3.

3.3.3.1.2 Stochastic search variable selection prior

Following the Stochastic Search Variables Selection (SSVS) prior proposed by George and McCulloch (1993); George, Sun, and Ni (2008), each element B_i of the coefficients matrix *B* for $i = 1,\cdots, \left(1+N\sum_{j=1}^{J+1}P_j\right)\times N$ is distributed as a mixture of two normal distributions:

$$B_i \sim (1-\gamma_i)\mathcal{N}\left(0,\tau_{0,i}^2\right)+\gamma_i\mathcal{N}\left(0,\tau_{0,1}^2\right) \tag{3.78}$$

Where γ_i is a dummy variable taking the values of zero or one. When γ_i is equal to zero, the coefficient B_i is sampled from the first Normal distribution, whereas when it is one, it is sampled from the second Normal distribution. Hyperparameters $\tau_{0,i}^2$ and $\tau_{0,1}^2$ essentially perform the variable selection by selecting small values for $\tau_{0,i}^2$ and relatively large uninformative values for $\tau_{0,1}^2$.

Dummy variable γ_i follows an independent Bernoulli prior, i.e. $\gamma \sim Bernoulli\left(\underline{p}_i\right)$, with success prior probability $\underline{p}_i$ such that:

$$\Pr\left(\gamma_i;\underline{p}_i\right)=\begin{cases}\underline{p}_i, & if\ \gamma_i=1\\ 1-\underline{p}_i, & if\ \gamma_i=0\end{cases} \tag{3.79}$$

Rewriting more compactly equation (3.84) in a matrix notation, then coefficients *B* prior is:

$$B|\gamma \sim \mathcal{N}\left(0, DD\right) \tag{3.80}$$

Where *D* is a diagonal matrix with the *i*-th diagonal element given as

$$D_{i,i}=\begin{cases}\tau_{0,i}, & if\ \gamma_i=0\\ \tau_{1,i}, & if\ \gamma_i=1\end{cases} \tag{3.81}$$

Turning to the posterior estimation, the Gibbs sampler is based on the following three blocks:

1. $p\left(B^{(r)}\middle|\gamma^{(r-1)},\Sigma^{(r-1)},Y\right)$
2. $p\left(\gamma^{(r)}\middle|B^{(r)},\Sigma^{(r-1)},Y\right)$
3. $p\left(\Sigma^{(r)}\middle|B^{(r)},\gamma^{(r)},Y\right)$

The first block draws from *B* posterior conditional on *γ,* which controls matrix *D*:

$$vec(B)\big|\gamma \sim \mathcal{N}\left(\overline{\mathrm{M}}_B,\overline{\Omega}_B\right) \tag{3.82}$$

Where $\overline{\Omega}_B=\left((DD)^{-1}+\Sigma^{-1}\otimes(X'X)\right)^{-1}$, $\overline{\mathrm{M}}_B=\overline{\Omega}_B\left(\left(\Sigma^{-1}\otimes X'\right)vec(Y)\right)$.

The second block draws γ from its posterior assuming independent Bernoulli distribution across all regressor coefficients B_i, $\gamma_i \sim Bernoulli\left(\underline{p}_i\right)$:

$$\Pr\left(\gamma_i;\overline{p}_i\right)=\begin{cases}\overline{p}_i, \ if \ \gamma_i=1\\ 1-\overline{p}_i, \ if \ \gamma_i=0\end{cases} \tag{3.83}$$

Where

$$\overline{p}_i=\frac{\frac{1}{\tau_{1,i}}\exp\left(-\frac{B_i^2}{2\tau_{1,i}^2}\right)\underline{p}_i}{\frac{1}{\tau_{1,i}}\exp\left(-\frac{B_i^2}{2\tau_{1,i}^2}\right)\underline{p}_i+\frac{1}{\tau_{0,i}}\exp\left(-\frac{B_i^2}{2\tau_{0,i}^2}\right)\left(1-\underline{p}_i\right)}.$$

Finally, the third block draws from the posterior of the innovations[61] variance-covariance Σ, as discussed previously in Section 3.3.3.1.1. More details on an equation-by-equation estimation of the posterior for the *B* coefficients are presented in Section 3.3.3.3.

3.3.3.2 Heteroscedastic MBVAR with stochastic volatility

By further relaxing the assumption of homoscedasticity and adding stochastic volatility (SV), the previous homoscedastic MBVAR model is now defined as:

$$Y_t=B_c+\sum_{j=1}^{J}\sum_{p=1}^{P_j}B_{j,p}W_{j,t-1-2^j(p-1)}+\sum_{p=1}^{P_{J+1}}B_{J+1,p}V_{J,t-1-2^J(p-1)}+e_t, \quad e_t \sim \mathcal{N}\left(0,\Sigma_t\right) \tag{3.84}$$

The inclusion of stochastic volatility, as in Cogley and Sargent (2005), transforms the homoscedastic MBVAR model described equation (3.64) with the constant variance-covariance matrix Σ to innovations whose distribution has a time-varying structure Σ_t defined as:

$$\Sigma_t=A^{-1}L_tA^{-1\prime} \tag{3.85}$$

Matrix *A* is a lower triangular matrix $N \times N$ with ones in the diagonal[62]:

[61] George, Sun, and Ni (2008) further allow for the off-diagonal elements (covariances) of Σ to follow an SSVS prior.

[62] Primiceri (2005) allows matrix *A* to be time-varying, i.e. A_t, with individual elements components $\alpha_{i,j}$ below the diagonal following a random walk: $\alpha_{m,n,t}=\alpha_{m,n,t-1}+\varphi_t, \ \ \varphi_t \sim \mathcal{N}(0,\sigma^2)$.

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \alpha_{2,1} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \alpha_{N,1} & \alpha_{N,2} & \cdots & \alpha_{N,N-1} & 1 \end{bmatrix} \tag{3.86}$$

Elements $\alpha_{m,n}$ below the diagonal, similar to Cogley and Sargent (2005), follow a normal prior distribution. Since elements $\alpha_{m,n}$ are estimated for each row of matrix A independently; it is more convenient to define the prior for all elements $\alpha_{m,n}$, $2 \leq m \leq N$, $1 \leq n \leq m-1$ per row m:

$$\alpha_{m,\cdot} \sim \mathcal{N}\left(\underline{\mathrm{M}}_{\alpha_{m,\cdot}}, \underline{\Omega}_{\alpha_{m,\cdot}}\right) \tag{3.87}$$

In equation (1.17), L_t is a diagonal matrix $N \times N$, and its diagonal elements $L_{n,n}$ for $n = 1,\cdots, N$ are defined as $L_{n,n,t} = h_{n,t}$, where the natural logarithm of $h_{n,t}$ follows a random walk:

$$\ln h_{n,t} = \ln h_{n,t-1} + \eta_{n,t}, \quad \eta_{n,t} \sim \mathcal{N}\left(0, \Phi_{n,n}\right) \tag{3.88}$$

Stacking all equations' stochastic volatilities $h_{n,t}$, $n = 1,\cdots, N$ and their corresponding innovations, it follows that $\ln h_t = \ln h_{t-1} + \eta_t$. Cogley and Sargent (2005) define matrix Φ as a diagonal matrix whose diagonal elements $\Phi_{n,n}$ follow an inverse gamma prior with scale matrix $\underline{S}$ and degrees of freedom $\underline{d}$:

$$\Phi_{n,n} \sim \mathcal{IG}\left(\underline{S}_{\Phi_{n,n}}, \underline{d}_{\Phi_{n,n}}\right) \tag{3.89}$$

The multiscale BVAR with stochastic volatility is simply a combination of the building blocks from the previously discussed homoscedastic multiscale BVAR specifications with the Minnesota type family or the SSVS prior by further adding the blocks relevant to the stochastic volatility. Thus, for the Minnesota type family, the iteration (r) from the Gibbs sampler is based now on the following four blocks:

1. $p\left(B^{(r)} \middle| A^{(r-1)}, h^{(r-1)}, \Phi^{(r-1)}, Y\right)$
2. $p\left(A^{(r)} \middle| B^{(r)}, h^{(r-1)}, \Phi^{(r-1)}, Y\right)$
3. $p\left(h^{(r)} \middle| B^{(r)}, A^{(r)}, \Phi^{(r-1)}, Y\right)$
4. $p\left(\Phi^{(r)} \middle| B^{(r)}, A^{(r)}, h^{(r)}, Y\right)$

Similarly, the SSVS prior for iteration (r) from the Gibbs sampler is based on the following five blocks:

1. $p\left(B^{(r)} \middle| \gamma^{(r-1)}, A^{(r-1)}, h^{(r-1)}, \Phi^{(r-1)}, Y\right)$
2. $p\left(\gamma^{(r)} \middle| B^{(r)}, A^{(r-1)}, h^{(r-1)}, \Phi^{(r-1)}, Y\right)$
3. $p\left(A^{(r)} \middle| B^{(r)}, \gamma^{(r)}, h^{(r-1)}, \Phi^{(r-1)}, Y\right)$
4. $p\left(h^{(r)} \middle| B^{(r)}, \gamma^{(r)}, A^{(r)}, \Phi^{(r-1)}, Y\right)$
5. $p\left(\Phi^{(r)} \middle| B^{(r)}, \gamma^{(r)}, A^{(r)}, h^{(r)}, Y\right)$

For the Minnesota family-type prior, the first block in a vectorised form would draw the coefficients B from their conditional posterior distribution:

$$vec(B) \sim \mathcal{N}\left(\overline{\mathrm{M}}_B, \overline{\Omega}_B\right) \tag{3.90}$$

Where $\overline{\Omega} = \left(\underline{\Omega}_B^{-1} + \sum_{t=1}^{T} \Sigma_t^{-1} \otimes \left(X_t X_t'\right)\right)^{-1}$, $\overline{\mathrm{M}}B = \overline{\Omega}B\left(\underline{\Omega}_B^{-1}\underline{\mathrm{M}}_B + vec\left(\sum_{t=1}^{T} X_t' Y_t \Sigma_t^{-1}\right)\right)$.

For the SSVS specification prior, the first block would be drawn from the following conditional posterior:

$$vec(B)\big|\gamma \sim \mathcal{N}\left(\overline{\mathrm{M}}_B, \overline{\Omega}_B\right) \tag{3.91}$$

Where $\overline{\Omega}_B = \left(\left(DD\right)^{-1} + \sum_{t=1}^{T} \Sigma_t^{-1} \otimes \left(X_t X_t'\right)\right)^{-1}$, $\overline{\mathrm{M}}_B = \overline{\Omega}_B\left(vec\left(\sum_{t=1}^{T} X_t' Y_t \Sigma_t^{-1}\right)\right)$.

The remaining blocks concerning the estimation of the stochastic volatility related blocks are estimated as follows. The posterior of matrix A elements $\alpha_{m,\cdot}$ for $2 \le m \le N$ is given by the following transformed regressions:

$$\varepsilon_{m,t} h_{m,t}^{-1/2} = \sum_{k=1}^{m-1} \alpha_{m,k}\left(-\varepsilon_{k,t} h_{m,t}^{-1/2}\right) + h_{m,t}^{-1/2}\xi_{m,t}, \quad \xi_{m,t} \sim \mathcal{N}\left(0,1\right)$$

Defining as Z_m and z_m the left-hand and right-hand side variables for each equation $2 \le m \le N$ respectively, the conditional posterior distributions of elements $\alpha_{m,\cdot}$ is given:

$$\alpha_{m,\cdot} \sim \mathcal{N}\left(\overline{\mathrm{M}}\alpha_{m,\cdot}, \overline{\Omega}\alpha_{m,\cdot}\right)$$

Where $\overline{\Omega}_{\alpha_{m,\cdot}} = \left(\underline{\Omega}_{\alpha_{m,\cdot}}^{-1} + \mathrm{Z}_m \mathrm{Z}_m\right)^{-1}$, $\overline{\mathrm{M}}\alpha_{m,\cdot} = \overline{\Omega}\alpha_{m,\cdot}\left(\underline{\Omega}_{\alpha_{m,\cdot}}^{-1}\underline{\mathrm{M}}_{\alpha_{m,\cdot}} + \mathrm{Z}_m' z_m\right)$

The stochastic volatility block h is estimated following the algorithm of Kim, Shephard, and Chib (1998). More details can be found in Appendix C.1.1 Stochastic volatility estimation.

Finally, the last block related to the stochastic volatility's innovations variance posterior distribution draws from the posterior of Φ independently for each equation n:

$$\Phi_{n,n} \sim \mathcal{IG}\left(\overline{S}_{\Phi_{n,n}}, \overline{d}_{\Phi_{n,n}}\right)$$

Where $\overline{S}_{\Phi_{n,n}} = \underline{S}_{\Phi_{n,n}} + \sum_{t=1}^{T}\left(\Delta \ln h_{n,t}\right)^2$ and $\overline{d}_{\Phi_{n,n}} = \underline{d}_{\Phi_{n,n}} + T$.

3.3.3.3 Equation-by-equation estimation

In order to achieve a fast estimation of the *B* coefficients posterior, the equation-by-equation estimation procedure originally suggested in Carriero, Clark, and Marcellino (2019) and further clarified by Carriero, Chan, Clark, and Marcellino (2022) is followed. In particular, for the MBVAR heteroscedastic specification[63], assuming that the innovations covariance Σ_t can be decomposed as $\Sigma_t = A^{-1} L_t A^{-1\prime}$ where A is lower triangular and L_t is a diagonal matrix as discussed previously in Section 3.3.3.2, then equation (3.67) can be written as:

$$\tilde{Y}_t = AY_t = AB'X_t' + L_t^{1/2}\varepsilon_t = A\left(X_t B\right)' + L_t^{1/2}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, I_N\right)$$

Thus, the model can be expressed in an equation-by-equation representation as follows:

$$\begin{aligned}
\tilde{Y}_{1,t} &= X_t B_{\bullet,1} + D_{1,1,t}^{1/2}\varepsilon_{1,t} \\
&\vdots \\
\tilde{Y}_{n,t} &= \sum_{i=1}^{n-1} A_{n,i} X_t B_{\bullet,i} + X_t B_{\bullet,n} + L_{n,n,t}^{1/2}\varepsilon_{n,t} \\
&\vdots \\
\tilde{Y}_{N,t} &= \sum_{i=1}^{N-1} A_{N,i} X_t B_{\bullet,i} + X_t B_{\bullet,N} + L_{N,N,t}^{1/2}\varepsilon_{N,t}
\end{aligned} \tag{3.92}$$

Where $\tilde{Y}_t = AY_t$ is an *N*-dimensional vector with the *n*-th element $\tilde{Y}_{n,t} = Y_{n,t} + \sum_{i=1}^{n-1} A_{n,i} Y_{i,t}$.

It becomes obvious that the *n-th* equation coefficients $B_{\bullet,n}$ influence not only the *n*-th equation but the remaining $n + 1, \cdots, N$ equations. This is formally stated by the following factorization:

$$\begin{aligned}
p\left(B_{\bullet,n} \middle| B_{\bullet,(-n)}, \Sigma_{1:T}, Y\right) \propto\ & p\left(B_{\bullet,n} \middle| B_{\bullet,1}, \cdots, B_{\bullet,j-1}, \Sigma_{1:T}, Y\right) \\
& \vdots \\
& \times p\left(B_{\bullet,N} \middle| B_{\bullet,N-1}, \cdots, B_{\bullet,1}, \Sigma_{1:T}, Y\right)
\end{aligned}$$

Where $B_{\bullet,(-n)} = \left[B_{\bullet,1}, \ldots, B_{\bullet,n-1}, B_{\bullet,n+1}, \ldots, B_{\bullet,N}\right]$

Using the most recent Gibb sampler draws, it follows that in order to estimate the coefficients for all *N* equations, *N* subsystems with $N - n + 1$ equations for each one need to be estimated. To be more concrete, for the estimations of coefficients of the *n*-th equation, the $i \geq n$ equations are used:

$$\begin{aligned}
Z_{n,t} &= X_t B_{\bullet,n} + L_{n,n,t}^{1/2}\varepsilon_{n,t} \\
Z_{n+1,t} &= A_{n+1,n} L_{n+1,n}^{-1} X_t B_{\bullet,n} + L_{n+1,n+1,t}^{1/2}\varepsilon_{n+1,t} \\
&\vdots \\
Z_{N,t} &= A_{N,n} X_t B_{\bullet,n} + L_{N,N,t}^{1/2}\varepsilon_{N,t}
\end{aligned}$$

[63] For the homoscedastic MBVAR specification the innovations variance-covariance Σ can be decomposed as $\Sigma = LDL'$ where *L* is lower triangular and *D* is a diagonal matrix.

Where $Z_{i,t} = \tilde{Y}_{i,t} - \sum_{k=1}^{n-1} A_{i,k} X_t B_{\bullet,k} - \sum_{k=n+1}^{N} A_{i,k} X_t B_{\bullet,k}$.

Having assumed a prior on *B,* then following the equation-by-equation estimation procedure, each column $B_{\bullet,n}$ can be estimated from the univariate regressions in equation (3.92) equation-by-equation representation of the system with the following posterior[64]:

$$B_{\bullet,n} \sim \mathcal{N}\left(\overline{\mathrm{M}}_{B_{\bullet,n}}, \overline{\Omega}_{B_{\bullet,n}}\right) \tag{3.93}$$

Where $\overline{\Omega}_{B_{\bullet,n}} = \left(\underline{\Omega}^{-1}_{B_{\bullet,n}} + \sum_{i=n}^{N} A_{i,n} \sum_{t=1}^{T} L^{-1}_{i,i,t} X'_t X_t\right)^{-1}$,

$\overline{\mathrm{M}}_{B_{\bullet,n}} = \overline{\Omega}_{B_{\bullet,n}} \left(\underline{\Omega}^{-1}_{B_{\bullet,n}} \underline{\mathrm{M}}_{B_{\bullet,n}} + \sum_{i=n}^{N} A_{i,n} \sum_{t=1}^{T} L^{-1}_{i,i,t} X'_t Z_{i,t}\right)$.

Following this procedure, instead of estimating *B* of dimensions $\left(1 + N\sum_{j=1}^{J+1} P_j\right) \times N$ at once, at each equation *n*, one column vector of dimensions $\left(1 + N\sum_{j=1}^{J+1} P_j\right) \times 1$ from matrix *B* is estimated.

In order to perform the sampling from the multivariate distribution in equation (3.93), the standard algorithm using the Cholesky decomposition would be:

$$B_{\bullet,n} = \overline{\mathrm{M}}_{B_{\bullet,n}} + chol\left(\overline{\Omega}_{B_{\bullet,n}}\right) \mathcal{N}_K\left(0,1\right)$$

Where $\mathcal{N}_K$ denotes $K = 1 + N\sum_{j=1}^{J+1} P_j$ draws from a standard normal distribution $\mathcal{N}_K(0,1)$.

However, in practice, for high dimensional problems, the posterior can be sampled using the more computationally efficient algorithm of Bhattacharya, Chakraborty, and Mallick (2016), which is suitable for regressions with high dimensionality in the regressors. More details are available in Appendix C.1.3 Fast sampling algorithm.

[64] In practice, for the estimation using Matlab, the Carriero, Chan, Clark, and Marcellino (2022), Section 3 alternative matrix notation is used.

3.3.3.4 MBVAR forecasting

In order to perform iterative forecasting using the MBVAR model, the following procedure is used. At each Gibbs sampler draw of the coefficients $B^{(r)}$, $\Sigma^{(r)}$, the iterative forecasting procedure uses the MODWT after the estimation of each h-step ahead forecast such that the lagged values of the MODWT wavelet and scaling coefficients can be used. For example, the $\hat{Y}_{T+h}^{(r)}$ forecast at a given Gibbs sampler draw (r) would be estimated as described below.

Having obtained the forecasts previously up to $T+h-1$, i.e. $\hat{Y}_{T+1}^{(r)},\cdots,\hat{Y}_{T+h-1}^{(r)}$, append them to the initial data $\left[Y_1,\cdots,Y_T,\hat{Y}_{T+1}^{(r)},\cdots,\hat{Y}_{T+h-1}^{(r)}\right]$ and use the MODWT to estimate $\left[W_{j,1},\cdots,W_{j,T},\hat{W}_{j,T+1}^{(r)},\cdots,\hat{W}_{j,T+h-1}^{(r)}\right]$ for $j=1,\cdots,J$ and $\left[V_{J,1},\cdots,V_{J,T},\hat{V}_{J,T+1}^{(r)},\cdots,\hat{V}_{J,T+h-1}^{(r)}\right]$. Then, the $\hat{Y}_{T+h}^{(r)}$ would simply be estimated as:

$$\hat{Y}_{T+h}^{(r)}=B_c^{(r)}+\sum_{j=1}^{J}\sum_{p=1}^{P_j}B_{j,p}^{(r)}\hat{W}_{j,T+h-1-2^j(p-1)}^{(r)}+\sum_{p=1}^{P_{J+1}}B_{J+1,p}^{(r)}\hat{V}_{J,T+h-1-2^J(p-1)}^{(r)}+e_{T+h},\quad e_{T+h}\sim\mathcal{N}\left(0,\Sigma^{(r)}\right)$$

The direct forecasting method discussed in Marcellino, Stock, and Watson (2006), adjusted to the current model, which does not require the MODWT at each Gibbs iteration, was also tested, but the results were not favourable and are not presented in the relevant Section 3.4.4.

3.4 Empirical application single-frequency time series

This section presents an empirical application of the proposed wavelet-based multivariate time series in an out-of-sample forecasting exercise with US macroeconomic data. Section 3.4.1 presents in detail the data that were used, Section 3.4.2 discusses various implementation details in terms of priors and models' specifications, and finally, Section 3.4.4 presents the point and density forecast accuracy of the proposed models.

3.4.1 Data description

All model specifications are estimated with the following six variables at quarterly frequency: real GDP growth, CPI inflation, Federal Funds Rates (FFR), total non-farm employment, industrial production (IP) and real consumption expenditure (PCE). All variables are transformed into growth rates (100 Δlog), except for the FFR, which is used in differences. Particularly for the case of FFR, the choice of differencing the data is based on the empirical application findings in Renaud, Starck, and Murtagh (2003), who show the superiority of the multiscale autoregressive model (MAR) for time series, which exhibit a lack of strong trend[65].

[65] Chan, Jacobi, and Zhu (2020) for example also difference the federal fund rates in a classical VAR framework.

The variables whose original frequency is monthly are first transformed into quarterly frequency by calculating each quarter's respective average. See Table 2.4 below.

Table 3.4: Single-frequency time series – variables

Variable	Transformation	Original freq.
GDP	100 Δlog	Quarterly
CPI	100 Δlog	Monthly
FFR	Δ	Monthly
Employment	100 Δlog	Monthly
Industrial production	100 Δlog	Monthly
PCE	100 Δlog	Quarterly

Source: Archival Federal Reserve Economic Data of the Federal Reserve Bank of St. Louis.

The out-of-sample forecasting exercise is performed recursively using an expanding window since 1960Q1, with the first out-of-sample exercise performed for 1985Q1 and the last one for 2018Q4. For each of the out-of-sample exercises at period t, the last available data point is that of $t-1$. The data used are as they were available during the 2020Q1 vintage.

Table 3.5 below shows how the total variance of each variable can be decomposed across the $J = 7$ MODWT scales in line with equation (3.34). In particular, for each variable, the values show which percentage of the original time series variance can be explained by this scale. For example, for GDP, 33.72% of the time series total variance of $0.67 = 0.82^2$ can be explained by the fluctuations with a period between 2 to 4 quarters.

It should be noted, though, that for the FFR, the table shows the variance decomposition both for the differenced time series used in this study, as well as in percentage points. For the differenced time series, the wavelet-based variance decomposition shows that more than 2/3 of the total variance can be explained by components with fluctuation periods up to 4 quarters. Differencing the time series as a way to detrend the data is actually reflected in the variance decomposition by reducing the variance for large scaling coefficients. For the case of FFR in percentage points, more than 50% of the total variance can be explained by the scaling coefficients V_6, which represent the slow-moving trend with a fluctuation period greater than 32 years – see also Matthes, Lubik, and Verona (2019).

Table 3.5: Variance decomposition

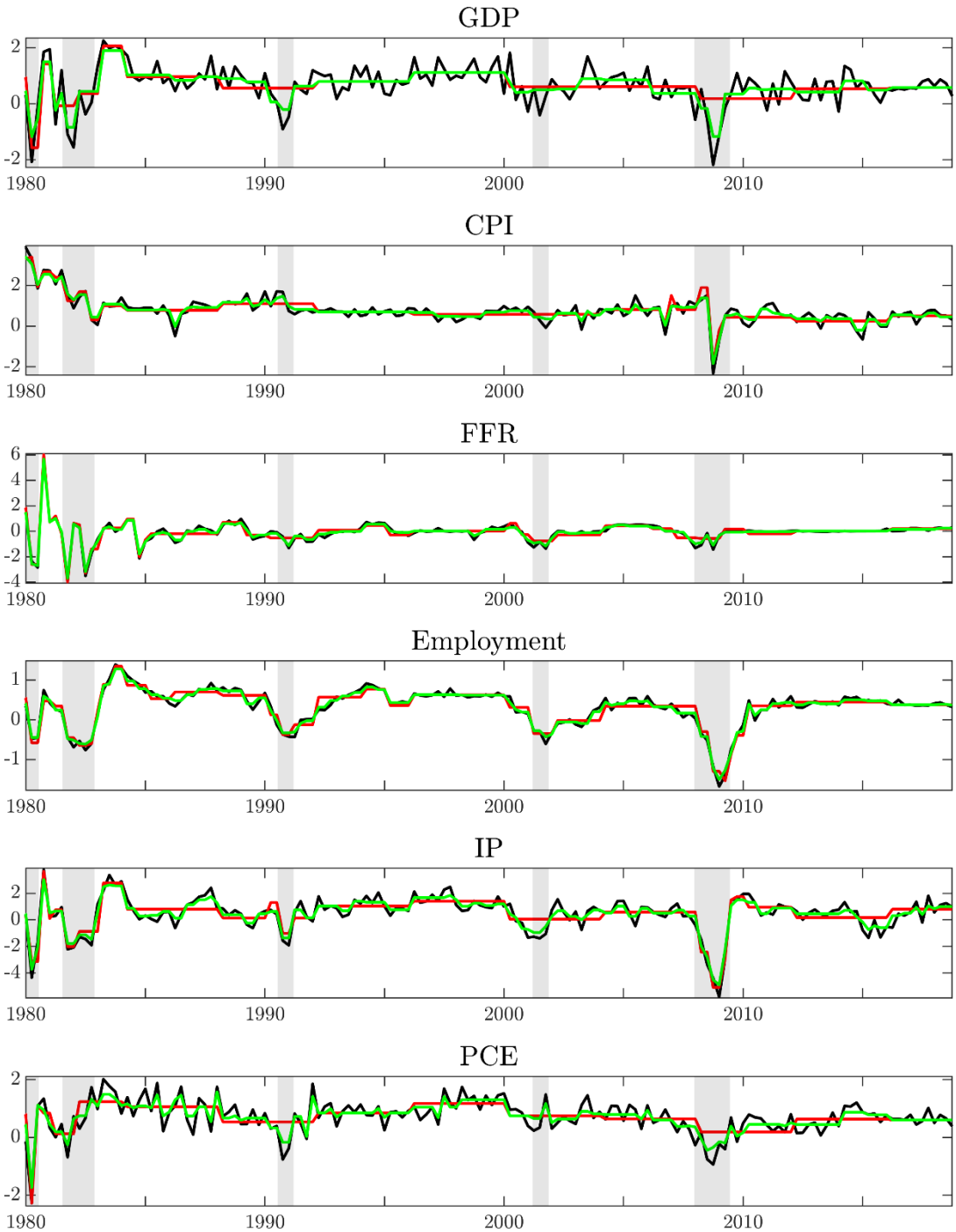
Variable	Mean	St. dev	Short term		Business cycle		Medium term		Long term
			W_1: 2-4Q	W_2: 4-8Q	W_3: 2-4Y	W_4: 4-8Y	W_5: 8-16Y	W_6: 16-32Y	V_6: >32Y
GDP	0.74	0.82	33.72	20.59	18.25	13.27	7.20	3.97	2.99
CPI	0.92	0.76	11.92	9.87	8.71	11.84	10.23	14.41	33.02
FFR (Δ)	-0.01	0.89	38.07	30.00	14.40	11.42	4.36	0.99	0.76
FFR (pc)	5.05	3.67	1.50	2.86	5.68	10.53	11.39	14.99	53.05
Empl.	0.43	0.53	8.34	13.94	21.91	24.95	15.46	5.71	9.69
IP	0.64	1.54	20.43	23.54	24.86	17.38	7.27	3.65	2.87
PCE	0.80	0.66	34.46	16.95	15.17	14.15	8.55	6.31	4.41

Note. Normalised variance decomposition across all MODWT components (may not add up to 100% due to rounding).

Figure 3.15 below shows the DWT denoised time series discussed in Section 3.3.1, which were used as input to a standard BVAR for forecasting. The time series are shown only for the post-1980 period for visibility reasons. Overall, it is clear that the SURE soft thresholding denoising method in green is much more smooth and follows more closely the original time series in black compared to the hard universal hard thresholding method in red. For example, for GDP in the 2000-2008 period, the universal hard thresholding method results in a constant line. This should be kept in mind because, as it will become evident in the empirical application in Section 3.4.4, despite the huge difference between the original time series end-point and the denoised time series end-point for each forecasting exercise, for some variables, even the 1-quarter ahead forecasts are significantly better than using the original time series.

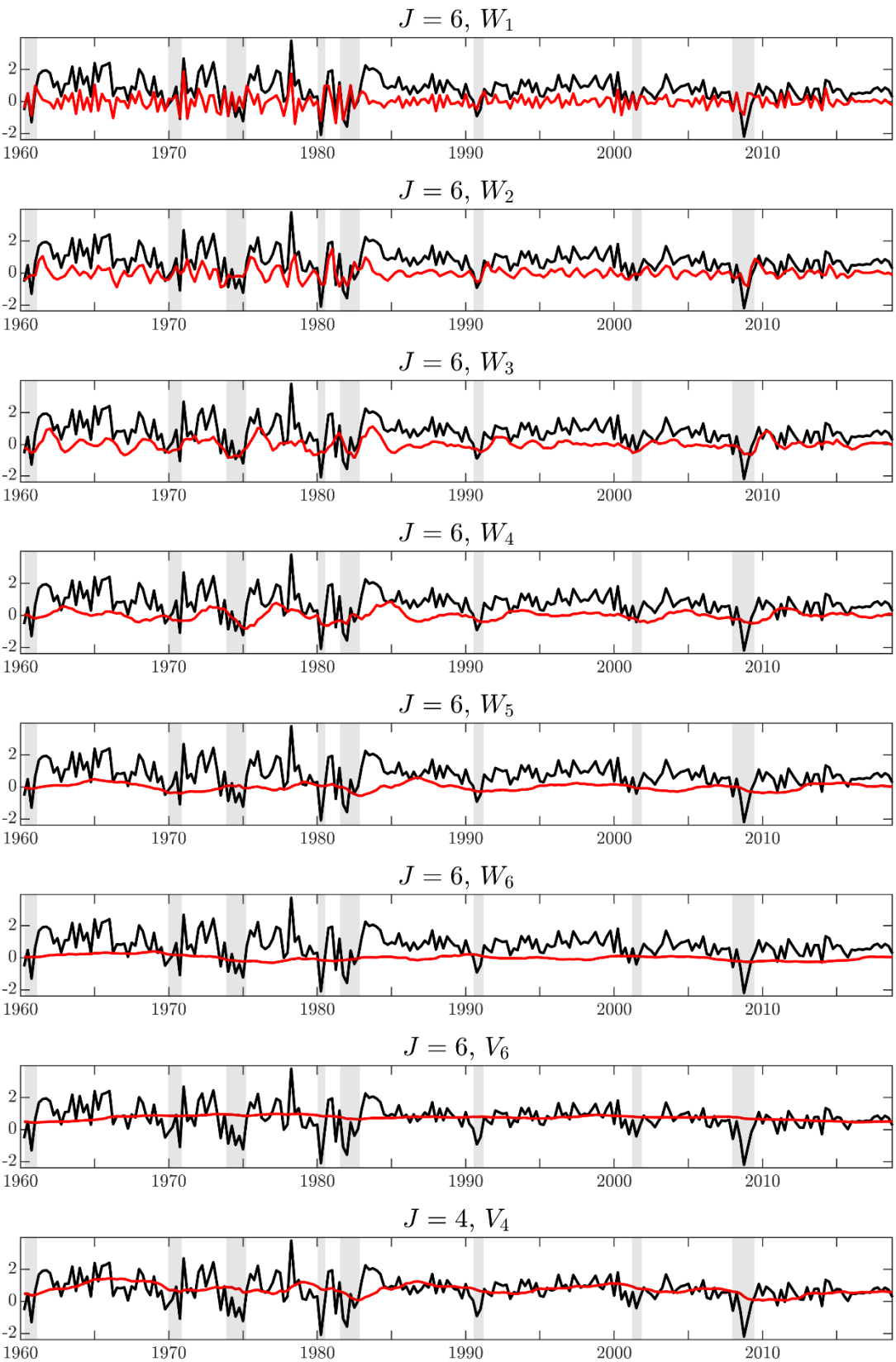
In continuation, Figure 3.16 shows, for the case of the GDP in black, the wavelet and scaling coefficients in red for $J = 6$ in the first seven panels, which were used for the separate MODWT modelling with BVARs. The eighth panel shows in red the scaling coefficients for $J = 4$, which were used for the multiscale BVAR. As a reminder, for $J = 4$, the wavelet coefficients $V_1 - V_4$ are the same as for $J = 6$. The wavelet and scaling coefficients for the remaining variables can be found in Appendix C.2.1 Wavelet and scaling coefficients.

Figure 3.15: DWT denoised time series



Note. Original time series are shown in black. Universal hard thresholding and SURE soft thresholding are shown in red and green, respectively.

Figure 3.16: GDP $J = 6$ levels Haar MODWT



Note. Black lines show actual time series. Red lines show MODWT $J = 6$ wavelet and scaling coefficients and $J = 4$ scaling coefficients.

3.4.2 Priors and other estimation details

All of the proposed wavelet-based multivariate time series forecasting discussed in Section 3.3 are evaluated against a benchmark BVAR of 4 lags with an independent normal inverse Wishart prior following the standard values for the hyperparameters set in the literature. In particular, the hyperparameters regarding the prior of the coefficients B in equation (3.59) are set as: overall shrinkage $\lambda_1 = 0.2$, cross-variable shrinkage $\lambda_2 = 0.5$, lag decay $\lambda_3 = 1$ and the constants use a diffuse prior. The innovations covariance Σ, which follows an inverse Wishart prior in equation (3.60), uses $N + 2$ degrees of freedom, and the scale is set equal to the residuals of independent AR(1) regressions for each variable.

Regarding the denoising methods discussed in Section 3.3.1, they use $J = 4$ levels (scales) of wavelet decomposition for GDP, FFR, industrial production and PCE and $J = 3$ for CPI and employment (see more details in Section 3.4.3.1). The BVAR models used for the denoised time series follow the exact same specification and priors which are used for the benchmark model.

The MODWT separate scale modelling of each level j wavelet and scaling coefficients discussed in Section 3.3.2 is based on $J = 6$ scales (see more details in Section 3.4.3.2). The BVAR models used at each scale j have the same prior specification of the benchmark with the only difference that the number of lags increases in a linear fashion in line with the scale, i.e. for $j = 1$, the BVAR model uses one lag, for $j = 2$ the model uses two lags and so forth. The MODWT separate scale modelling, which employs BVAR-SV for wavelet coefficients up to the fourth scale, i.e. W_1-W_4, uses an inverse Gamma prior in equation (1.21) with 2 degrees of freedom and a scale equal to 0.1, while for the elements of the lower triangular matrix A a diffuse prior is used.

The homoscedastic and the heteroscedastic MBVAR models discussed in Section 3.3.3 are both based on $J = 4$ levels of wavelet decomposition and one lag across all levels j, i.e. $P_j = 1$, for $j = 1,\cdots, J, J + 1$ (see more details in Section 3.4.3.3). The Minnesota family type prior of coefficients B use the following hyperparameters discussed in equation (3.71): $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $\lambda_3 = 1$, $\lambda_0 = 10^3$ and $\tilde{\lambda}_j = j^{\lambda_4}$ with $\lambda_4 = 1$ for $j = 1,\cdots, J, J + 1$. The homoscedastic and heteroscedastic MBVAR specifications using SSVS have the following prior parameters: $\underline{p}_i = 0.5$, $\tau_{0,i} = 0.01$, and $\tau_{1,i} = 1$. The prior of the stochastic volatility innovations in equation (3.89) uses 2 degrees of freedom and a scale equal to 0.1, while for the elements of the lower triangular matrix A, a diffuse prior is used.

For all the estimated models, the first 5000 draws are discarded as a burn-in sample, and the subsequent 5000 draws are used for posterior inference.

3.4.3 Optimal decomposition level selection

The current section presents how the optimal level of decomposition has been selected across each of the three wavelet-based forecasting approaches.

3.4.3.1 DWT denoising optimal decomposition level selection

The optimal number of the DWT partial decomposition level J_0 for each variable is set according to Srivastava, Anderson, and Freed (2016), using the 'peak-to-sum' ratio of the wavelet coefficients at level j defined as:

$$PS_j = \frac{\max\left(\left|W_j\right|\right)}{\sum_{t=1}^{T_j}\left|W_{j,t}\right|} \tag{3.94}$$

The peak-to-sum ratio reflects the sparsity of the wavelet coefficients and is bounded as $0 \leq PS_j < 1$. Its interpretation suggests that a small PS_j reveals noise with a large number of small wavelet coefficient values, whereas a large value implies only a few large wavelet coefficient values at level j. Srivastava, Anderson, and Freed (2016) suggest using the threshold value $T_r \approx 0.2$ such that $PS_j \leq T_r$ and $PS_{j+1} > T_r$ provide the DWT optimal decomposition level $j = J_{0,\,\text{opt}}$.

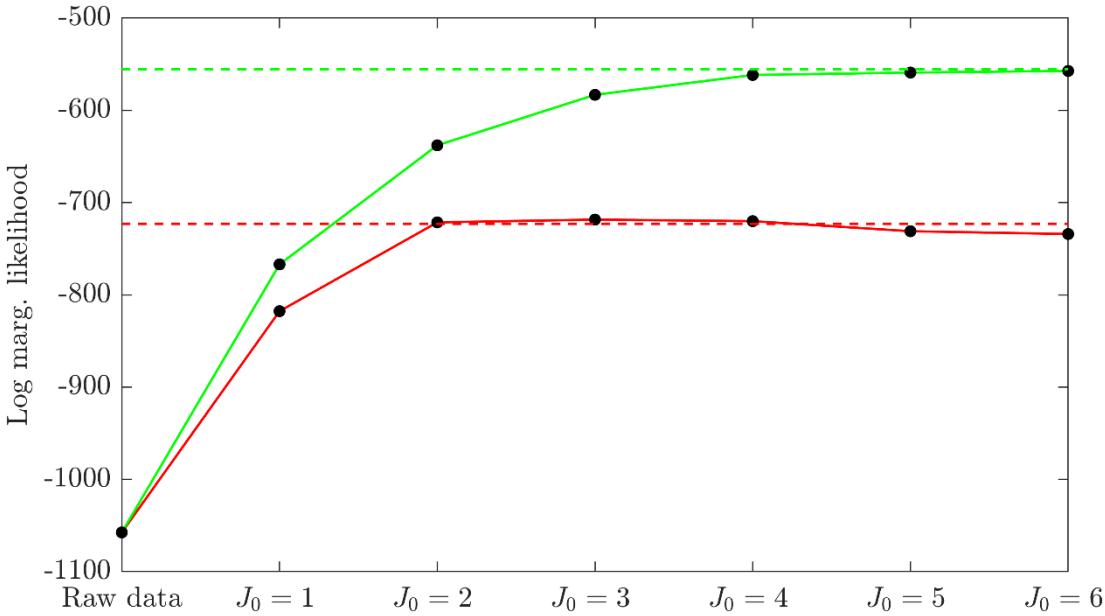
The following table shows, for each variable n, the relevant values of PS_j at each level j, where the decomposition level j corresponding to the shaded values is the optimal decomposition level $J_{0,\,\text{opt},\,n}$. The maximum potential decomposition level $J_{0,\,\max}$ is set equal to 6 and is dictated by the time series length of $T = 100$ observations used for the first out-of-sample exercise for 1985Q1 using the estimation sample beginning in 1960Q1, i.e. $6.64 = \log_2(100)$ rounded to the smallest integer.

Table 3.6: Peak-to-sum ratio

	GDP	**CPI**	**FFR**	**Employment**	**IP**	**PCE**
$j = 1$	0.05	0.04	0.10	0.05	0.04	0.05
$j = 2$	0.11	0.16	0.16	0.09	0.09	0.07
$j = 3$	0.11	0.12	0.14	0.14	0.14	0.16
$j = 4$	0.20	0.28	0.19	0.23	0.19	0.20
$j = 5$	0.23	0.29	0.49	0.26	0.24	0.26
$j = 6$	0.38	0.46	0.43	0.47	0.34	0.36
$J_{0,\,opt}$	4	3	4	3	4	4

As an additional criterion, given that ultimately what is of interest are the forecasts produced from the BVARs estimated with the denoised series, the marginal likelihoods of the BVARs estimated with denoised time series using as decomposition levels $J_0 = 1,\cdots,6$ across all variables are compared with the BVAR marginal likelihood estimated with the denoised time series using the optimal decomposition level $J_{0,\,\text{opt},\,n}$ given in the above table.

Figure 3.17: BVAR log marginal likelihood for raw and DWT denoised time series



Note. BVAR log marginal likelihood for raw time series and J_0 =1,⋯,6 universal threshold and SURE DWT denoising shown in red and green continuous lines, respectively. Dashed lines show log marginal likelihood estimated with variable specific optimally selected $J_{0,\,\mathrm{opt},\,n}$ according to Srivastava, Anderson, and Freed (2016).

3.4.3.2 Separate scale BVAR optimal decomposition level selection

The optimal decomposition level J for the MODWT wavelet transform follows the minimum entropy criterion originally suggested in Coifman and Wickerhauser (1992), which has been used in the context of macroeconomic/financial times series in Bekiros and Marcellino (2013), Uddin, Bekiros, and Ahmed (2018) and Uddin, Gençay, Bekiros, and Sahamkhadam (2019) for the MODWT-MRA. The optimal decomposition level J_0 seeks to minimise the cost function C, which is an additive function over the orthonormal basis of the DWT transform:

$$C\left(\tilde{W}\right) = \sum_{j=1}^{J_0} C\left(\tilde{W}_j\right) \tag{3.95}$$

Formally, the optimal decomposition level J_0 is the solution to the following minimisation problem:

$$\min_{J_0} C\left(\tilde{W}\right) \tag{3.96}$$

Replacing the cost function C, with the entropy function E, for a given level j, the wavelet coefficients' entropy is defined as:

$$E\left(\tilde{W}_j\right) = -\sum_{t=1}^{T_j} p_t^2 \log\left(p_t^2\right) \tag{3.97}$$

Where $p_t = \tilde{W}_{j,t} \Big/ \left\| \tilde{W}_{j,t} \right\|$ with $\| \ \|$ denoting the l^2 (Euclidian) norm[66]. Entropy follows the convention $E(0) = 0 \log(0) = 0$.

Recursively beginning from the bottom, if the sum of the $j + 1$ wavelet and scaling coefficients (children) entropy is larger than the j scaling coefficients (parent node), then $j + 1$ decomposition level reveals non-redundant information and is used; otherwise, level j scaling coefficients node is set as the new terminal decomposition level Coifman and Wickerhauser (1992). This is equivalent to finding the optimal decomposition level $J_{0,\,opt}$, which has the minimum entropy for $J_0 = 1,\cdots, J_{0,\,max}$ estimated for each J_0 using all the wavelet coefficients $j = 1,\cdots, J_0$ and the J_0 scaling coefficients. The following table presents the optimal decomposition level $J_{0,\,opt,\,n}$ with respect to the minimum entropy in shaded numbers across each variable n for a maximum decomposition level of $J_{0,\,max} = 6$.

Table 3.7: Optimal minimum entropy criterion

	GDP	**CPI**	**FFR**	**Employment**	**IP**	**PCE**
$J_0 = 1$	4.605	4.173	3.690	4.452	4.283	4.661
$J_0 = 2$	4.307	3.626	3.514	4.064	4.171	4.265
$J_0 = 3$	4.198	3.238	3.527	3.757	4.094	4.010
$J_0 = 4$	4.010	2.890	3.526	3.521	4.005	3.734
$J_0 = 5$	3.749	2.515	3.519	3.483	3.956	3.415
$J_0 = 6$	3.524	2.426	3.512	3.324	3.987	3.056

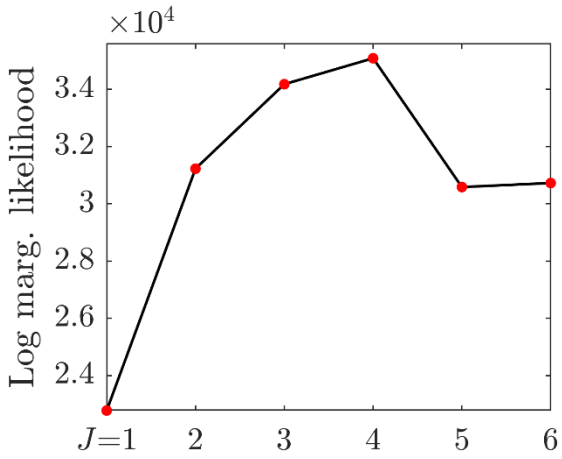
The optimally selected decomposition levels for each variable n can be used for separate univariate time series modelling. However, due to using a VAR process across each scale j, and in order to keep symmetry across all variables, the MODWT global scale J selected is set equal to the maximum of all the variable-specific optimally selected decomposition levels, i.e. $J = \max(J_{0,\,opt,\,n}) = \max(6, 6, 6, 6, 5, 6) = 6$.

3.4.3.3 MBVAR optimal decomposition level selection

The MBVAR optimum decomposition level $J = 4$ is selected based on maximising the marginal likelihood using a simple grid search over $J = 1,\cdots, 6$ for a fixed set of hyperparameters. The following figure shows how the homoscedastic MBVAR – Minnesota type prior log marginal likelihood changes for the different values of the decomposition level J while hyperparameters are kept constant.

[66] The $-l^2 \log l^2$ cost functional is also known as Shannon entropy.

Figure 3.18: Homoscedastic MBVAR log marginal likelihood for different decomposition levels J



Note. All the remaining hyperparameters are kept constant.

3.4.4 Out-of-sample results – Forecast evaluation metrics

Point forecasts for variable n and horizon h are evaluated according to the root mean squared error (RMSE) :

$$\text{RMSE}_{n,h} = \left[\frac{1}{T-h-t_0-1} \sum_{t=t_0}^{T-h} \left(Y_{n,t+h} - \hat{Y}_{n,t+h|t} \right)^2 \right]^{1/2}$$

Where t_0 refers to 1985Q4 and T refers to 2018Q4.

Density forecasts are evaluated according to the continuous ranked probability score (CRPS). Following, Gneiting and Raftery (2007), the CPRS can be evaluated in a closed form as:

$$\text{CRPS}(F, y) = \frac{1}{2} E_F \left| \hat{y} - \hat{y}' \right| - E_F \left| y - \hat{y} \right|$$

Where $\hat{y}$ are independent draws from the predictive density with cumulative distribution F and y is the outcome.

The exact formula[67] for the retained draws R of the MCMC algorithm for variable n and horizon h is given as:

$$\text{CRPS}_{n,h} = \frac{1}{T-h-t_0-1} \sum_{t=t_0}^{T-h} \left[\frac{1}{R} \sum_{r=1}^{R} \left| \hat{Y}_{n,t+h|t}^{(r)} - Y_{n,t+h} \right| - \frac{1}{2R^2} \sum_{r=1}^{R} \sum_{r'=1}^{R} \left| \hat{Y}_{n,t+h|t}^{(r)} - \hat{Y}_{n,t+h|t}^{(r')} \right| \right]$$

Furthermore, density forecasts are also evaluated using the log predictive score (LPS), and the results are available in Appendix C.2.2 . The time variation of the point and density forecast accuracy is presented in Appendix C.2.3 Time variation of results.

[67] More information can be found in Dieppe, Legrand, and Van Roye (2016).

In order to provide a rough measure of whether the differences in the forecast accuracy are significant, the Diebold and Mariano (1995) test is employed. The hypothesis testing is always one-sided, i.e. rejection of the null of equal performance versus the benchmark model provides evidence of the proposed model outperforming the benchmark. The test applies in the same manner both for point forecasts (MSE) and density forecasts[68].

Turning now to the point forecasts accuracy results in Table 3.8 and examining the suggested models, the following are observed. Regarding the denoising methods, overall, the hard denoising method exhibits more gains than the SURE soft denoising method. For given GDP forecast horizons and PCE across all horizons for the hard denoising method, results are found to be significant. What is of interest is the medium to long-term forecast improvement for CPI bearing in mind the big differences between the original and the denoised time series in Figure 3.15.

Regarding the MODWT separate scale BVAR modelling approaches, both homoscedastic and heteroscedastic, except for FFR and the 1-quarter ahead forecasts for CPI, industrial production and employment, the results overall are in favour of the proposed models in terms of RMSE ratios and further found to be significant for some horizons. The inclusion of stochastic volatility for $j \leq 4$ provides some noticeable improvements for the 1-quarter ahead forecast across all variables except for the FFR; however, for medium to long-term forecasts, there are not any improvements.

Finally, regarding the third approach of the multiscale BVAR, for the homoscedastic specification, both for the proposed Minnesota type prior and the SSVS, the results are in favour of the proposed model across all variables and horizons except for the 1-quarter ahead forecasts where the model does not exhibit any forecast gains. The improved forecast accuracy is also found to be significant and is the only approach so far that provides robust results across all variables. Adding stochastic volatility to the multiscale BVAR does not yield any noticeable improvements. In particular, the 1-quarter ahead forecasts are improved but not satisfactory since the RMSE ratios remain above unity except for FFR, while the medium to long-term forecasts deteriorate on some occasions. CPI forecast accuracy beyond the 1-quarter ahead forecast improves significantly across all four specifications of the multiscale BVAR model.

Focusing on the density forecasts as evaluated by the CRPS in Table 3.9, the overall picture with few exceptions is that wavelet-based forecasting improves density forecasts significantly on many occasions, and similar results can be found in Appendix C.2.2 Log predictive score, where the log predictive score results are presented.

Examining the forecasting approaches proposed one by one, for both denoising methods used, density forecasts are improved with respect to the benchmark, except for the 1-quarter ahead forecasts of CPI and medium to long-term forecasts for employment. The reason for this improvement is that the variance estimate of the BVAR's innovations Σ is much smaller for the denoised time series, and thus the uncertainty over the density forecasts is reduced over non-recessionary periods.

The MODWT separate scale BVAR modelling approach also improves forecast accuracy across variables and forecast horizons in a robust way, except for a few cases. For the

[68] The results for density forecasts should be interpreted with care since this type of statistical testing applies to rolling estimation schemes and not to recursive re-estimation of the models under an expanding window (Amisano & Giacomini, 2007).

homoscedastic BVAR specifications, in particular, there are only 4 cases out of the 24 that the benchmark performs better, while the inclusion of stochastic volatility for small scales seems to improve the overall picture, with only 2 out of the 24 cases now being in favour of the benchmark. Stochastic volatility brings the largest improvements for FFR, where for all horizons, the differences now in favour of the proposed model are found to be significant.

Finally, examining the multiscale BVAR density forecast results, they present the same behaviour as for the point forecasts, i.e. the proposed model improves density forecasts robustly across the majority of the variables and the forecast horizons except for the 1-quarter ahead forecasts for the case of the Minnesota type prior. For 9 out of the 24 cases, the improvement is found to be significant for the homoscedastic case. Adding stochastic volatility helps the 1-quarter ahead forecasts marginally but is not satisfactory enough with respect to the benchmark, except for the case of the FFR, where the results for all horizons are significant.

Table 3.8: Point forecast evaluation

	GDP	CPI	FFR	Employment	IP	PCE
			BVAR			
$h = 1$	0.545	0.503	0.424	0.211	1.027	0.522
$h = 4$	0.644	0.565	0.441	0.451	1.321	0.578
$h = 8$	0.659	0.584	0.434	0.519	1.332	0.567
$h = 12$	0.641	0.623	0.443	0.524	1.298	0.557
			Universal hard threshold denoising BVAR			
$h = 1$	0.996	1.023	1.055	1.193	0.993	0.900**
$h = 4$	0.985	0.989	1.017	1.022	0.993	0.938***
$h = 8$	0.981***	0.987	1.014	1.001	0.996	0.969**
$h = 12$	0.995	0.957**	0.992	1.002	1.005	0.986**
			SURE soft threshold denoising BVAR			
$h = 1$	0.973	1.020	1.006	1.032	0.945*	0.918**
$h = 4$	0.990	1.019	1.015	1.004	1.005	0.972**
$h = 8$	1.007	1.005	1.017	1.022	1.026	1.003
$h = 12$	1.010	0.973	0.993	1.025	1.014	1.018
			MODWT separate scale BVAR			
$h = 1$	0.960	1.036	1.050	1.056	1.006	0.913*
$h = 4$	0.889***	0.906**	1.119	0.959	0.993	0.842***
$h = 8$	0.915*	0.867**	1.071	0.894	0.959	0.902*
$h = 12$	0.982	0.784***	1.050	0.866	0.975	0.972
			MODWT separate scale BVAR-SV ($j \leq 4$) / BVAR($j > 4$)			
$h = 1$	0.973	1.028	0.922*	0.986	0.951	0.908*
$h = 4$	0.895**	0.912**	1.110	0.955	1.002	0.842***
$h = 8$	0.905**	0.868***	1.060	0.891	0.953	0.906*
$h = 12$	0.986	0.779***	1.038	0.864	0.977	0.975
			Multiscale BVAR-Minnesota			
$h = 1$	1.102	1.034	1.016	1.416	1.129	0.970
$h = 4$	0.946***	0.927**	0.962**	0.888***	0.945**	0.885***
$h = 8$	0.948***	0.879***	0.967*	0.876**	0.951**	0.924***
$h = 12$	0.973***	0.836***	0.955**	0.886***	0.967***	0.953***
			Multiscale BVAR-Minnesota-SV			
$h = 1$	1.037	1.027	0.914	1.366	1.085	0.899**
$h = 4$	0.946***	0.902**	0.948**	0.913**	0.968	0.875***
$h = 8$	0.972*	0.859***	0.964*	0.935**	0.967*	0.945**
$h = 12$	0.993	0.805***	0.949**	0.974*	1.003	0.986
			Multiscale BVAR-SSVS			
$h = 1$	0.992	1.012	0.998	1.251	1.032	0.954
$h = 4$	0.892***	0.931**	1.052	0.854**	0.943	0.861***
$h = 8$	0.934**	0.888**	0.952*	0.847**	0.966**	0.897**
$h = 12$	0.975	0.827***	0.957	0.856**	0.984*	0.957*
			Multiscale BVAR-SSVS-SV			
$h = 1$	0.954	1.037	0.905*	1.169	1.025	0.946
$h = 4$	0.898***	0.885**	0.972	0.846*	0.973	0.867***
$h = 8$	0.962*	0.887**	1.003	0.907	0.996	0.935**
$h = 12$	1.010	0.851**	0.997	0.935	1.021	0.989

Note. First panel shows benchmark's raw RMSEs. Remaining panels show RMSE ratios of the proposed models with respect to the benchmark. Stars refer to the p-values of the Diebold and Mariano (1995) one-sided test with respect to the benchmark. *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively. The evaluation period is 1985-2018.

Table 3.9: Density forecast evaluation with the CRPS metric

	GDP	CPI	FFR	Employment	IP	PCE
			BVAR			
$h = 1$	0.312	0.244	0.293	0.119	0.559	0.291
$h = 4$	0.357	0.289	0.309	0.232	0.687	0.322
$h = 8$	0.362	0.315	0.307	0.268	0.690	0.318
$h = 12$	0.355	0.337	0.312	0.273	0.679	0.314
			Universal hard threshold denoising BVAR			
$h = 1$	0.936**	1.034	0.998	1.136	0.967*	0.887***
$h = 4$	0.942***	0.969*	0.971***	1.014	0.972**	0.919***
$h = 8$	0.939***	0.960	0.968***	1.000	0.980**	0.948**
$h = 12$	0.952**	0.943***	0.949***	0.999	0.980**	0.960*
			SURE soft threshold denoising BVAR			
$h = 1$	0.923***	1.049	0.939***	0.986	0.929***	0.910***
$h = 4$	0.951**	0.990	0.933***	0.997	0.982*	0.963**
$h = 8$	0.966*	0.969	0.938***	1.029	1.004	0.987
$h = 12$	0.973	0.950**	0.922***	1.030	0.988	0.998
			MODWT separate scale BVAR			
$h = 1$	0.933*	1.057	0.925***	0.962	0.980	0.916**
$h = 4$	0.892***	0.895**	1.008	1.006	1.027	0.848***
$h = 8$	0.922**	0.849***	0.990	0.933	0.984	0.904*
$h = 12$	0.978*	0.768***	0.982	0.901	0.979	0.963
			MODWT separate scale BVAR-SV ($j \leq 4$) / BVAR($j > 4$)			
$h = 1$	0.947	1.029	0.705***	0.881***	0.928	0.919*
$h = 4$	0.878***	0.898***	0.884***	0.999	1.033	0.838***
$h = 8$	0.893***	0.846***	0.850***	0.933	0.975	0.903**
$h = 12$	0.960**	0.775***	0.836***	0.898	0.967**	0.953
			Multiscale BVAR-Minnesota			
$h = 1$	1.101	1.053	1.046	1.326	1.132	0.998
$h = 4$	0.975	0.911**	0.994	0.875***	0.966	0.914**
$h = 8$	0.978	0.856***	1.003	0.863***	0.970	0.941***
$h = 12$	1.000	0.796***	0.991	0.861***	0.976**	0.964**
			Multiscale BVAR-Minnesota-SV			
$h = 1$	0.966	1.025	0.662***	1.126	1.007	0.879***
$h = 4$	0.914***	0.899***	0.734***	0.866***	0.954	0.868***
$h = 8$	0.946**	0.847***	0.741***	0.949	0.976	0.939**
$h = 12$	0.971	0.784***	0.715***	1.018	0.995	0.967
			Multiscale BVAR-SSVS			
$h = 1$	0.992	1.051	1.003	1.190	1.033	0.966
$h = 4$	0.915***	0.933**	1.037	0.867**	0.965	0.888***
$h = 8$	0.965	0.885**	1.020	0.866**	0.982	0.925*
$h = 12$	1.009	0.830***	1.020	0.891	1.006	0.984
			Multiscale BVAR-SSVS-SV			
$h = 1$	0.918**	1.059	0.654***	1.052	0.990	0.918*
$h = 4$	0.879***	0.903***	0.747***	0.854**	0.999	0.862***
$h = 8$	0.935**	0.884**	0.793***	0.923	1.008	0.929**
$h = 12$	0.993	0.877	0.782***	0.961	1.009	0.985

Note. First panel shows benchmark's raw CRPS. Remaining panels show the CRPS ratios of the proposed models with respect to the benchmark. Stars refer to the p-values of the Diebold and Mariano (1995) one-sided test with respect to the benchmark. *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively. The evaluation period is 1985-2018.

3.5 Mixed-frequency time series models

Modelling mixed-frequency (MF) time series usually adheres to two main approaches in the VAR literature. The first approach, which is observation driven, involves the representation and estimation of the VAR model in the lowest frequency, where the high-frequency time series are treated as multiple lower frequency time series. For example, for a VAR using annual and quarterly frequency time series, the quarterly time series would be treated as four separate annual time series, one for each quarter of the calendar year. This approach is usually called the stacked VAR approach and has been studied thoroughly by Carriero, Clark, and Marcellino (2015b), Ghysels (2016), Koop, McIntyre, and Mitchell (2020), and McCracken, Owyang, and Sekhposyan (2021).

The second approach is based on a state-space model representation. The lowest frequency time series is treated as if its corresponding high-frequency time series latent process that is not observed was known, and the VAR model is essentially written in the highest frequency. Kalman filtering allows the estimation of the latent process based on the observations in the lowest frequency. In other words, the lower frequency time series are interpolated such that they satisfy an intertemporal aggregation restriction depending on whether the time series enter the VAR in levels or growth rates. Indicative literature following this approach includes Mariano and Murasawa (2010), Schorfheide and Song (2015) and Koop, McIntyre, Mitchell, and Poon (2020). The state-space model representation approach presents the main advantage that through the appropriate adjustment of the Kalman filter update equations, it can handle various irregularities which are frequent in macroeconomic data, such as missing[69] observations and ragged edge data. However, this comes at the expense of the large computational cost of the Kalman filtering and smoothing algorithms which need to be run at its iteration of the MCMC algorithm.

For the purpose of this study, I will focus on the second approach, i.e. representation of the model in the higher frequency via a state-space model, and I will show how wavelet transform can be utilised in that respect. MF-VARs can be used like single-frequency VARs for the purpose of forecasting, structural inference and policy/scenario analysis (Sims, 1980; Stock & Watson, 2001); however, the majority of their applications are focused on nowcasting. Especially for the case of MF-VARs casted in a state-space representation, which can address the ragged edge data problem effectively, a natural by-product is the estimation of the high-frequency latent process corresponding to the low-frequency time series, which can be used separately at a later stage.

Due to the high computational cost of re-estimating the mixed-frequency VAR by Kalman techniques for an out-of-sample forecasting exercise, this study will rather focus as an illustrative example on the in-sample accuracy of the monthly estimates for the proposed wavelet-based MF-VAR method versus the standard MF-VAR approach (henceforth simply named as MF-VAR). In order to prove the proposed model's improved accuracy, time series, which are known both in quarterly and monthly frequency, will be treated as if the observed monthly frequency time series were the latent processes which need to be estimated with the proposed model. The estimated monthly time series will be compared with their actual

[69] For a detailed textbook treatment of missing observations refer to Harvey (1989), p. 143 and Durbin and Koopman (2012), p. 110.

(observed) monthly time series values for the cases of the MF-VAR approach and the proposed wavelet-based MF-VAR approach.

Of course, for a real-world application like GDP, investment and government spending time series' monthly estimates that their actual monthly values are unknown, it would not be possible to know with certainty the superiority of the proposed model. However, by showing that for a number of quarterly time series whose monthly time series are known, the proposed model's accuracy cannot get worse than the MF-VAR approach, it provides some empirical evidence of its advantages that could hold for the series that are of interest for higher frequency estimates.

The following Sections 3.5.1 and 3.5.2, explain in detail the MF-VAR and the wavelet-based MF-VAR, respectively.

3.5.1 Mixed-frequency VAR

Assuming a N-multivariate time series $Y_t = \left[Y_{1,t}, Y_{2,t}, \cdots, Y_{N,t} \right]'$ at a monthly frequency, then it follows that it can be modelled using a VAR process:

$$Y_t = c + \sum_{p=1}^{P} B_p \, Y_{t-p} + \varepsilon_t \tag{3.98}$$

Using the standard VAR setting, c is of $N \times 1$ dimensions, B_p are $N \times N$ matrices, and the innovations ε_t is $N \times 1$ column vector normally distributed as $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$.

Now assume that out of the $n = 1, \cdots, N$ in total univariate time series comprising the N-multivariate time series, the first n_1 time series are observed only in quarterly frequency, while the remaining $n_2 = N - n_1$ are observed directly in monthly frequency:

$$Y_t = \left[Y_{1,t}^{q}, \cdots, Y_{n_1,t}^{q}, Y_{1,t}^{m}, \cdots, Y_{n_2,t}^{m} \right]'$$

Where in terms of notation, superscript q in $Y_{1,t}^{q}$ denotes that the underlying univariate time series is observed in quarterly frequency, whereas superscript m in $Y_{1,t}^{m}$ that the univariate time series is observed directly in monthly frequency. The subvectors of dimensions n_1 and n_2, respectively, for the quarterly and monthly time series, are denoted as:

$$Y_t^{q} = \left[Y_{1,t}^{q}, \cdots, Y_{n_1,t}^{q} \right]' \quad \text{and} \quad Y_t^{m} = \left[Y_{1,t}^{m}, \cdots, Y_{n_2,t}^{m} \right]' \tag{3.99}$$

The observed values of the n_1 quarterly frequency time series, where t denotes the month that the quarterly values are observed and not the respective quarter, are further denoted as:

$$Y_t^{o} = \left[Y_{1,t}^{o}, \cdots, Y_{n_1,t}^{o} \right]' \tag{3.100}$$

Following Mariano and Murasawa (2003, 2010), a univariate quarterly time series transformed into growth rates can be mapped on its corresponding monthly latent process via the following approximation:

$$Y_{1,t}^{o} = \frac{1}{3}Y_{1,t}^{q} + \frac{2}{3}Y_{1,t-1}^{q} + Y_{1,t-2}^{q} + \frac{2}{3}Y_{1,t-3}^{q} + \frac{1}{3}Y_{1,t-4}^{q} \tag{3.101}$$

Where $Y_{1,t}^{o} = \Delta_3\left(\ln \tilde{Y}_{1,t}^{o}\right) = \ln \tilde{Y}_{1,t}^{o} - \ln \tilde{Y}_{1,t-3}^{o}$ and $Y_{1,t}^{q} = \Delta\left(\ln \tilde{Y}_{1,t}^{q}\right) = \ln \tilde{Y}_{1,t}^{q} - \ln \tilde{Y}_{1,t-1}^{q}$ with a tilde denoting the time series in levels.

For US data, following the national income and product accounts (NIPA) convention, higher frequency flow variables are annualised, and as a consequence, equation (3.101) which treats the variable as a stock, remains valid. For example, the GDP or personal consumption expenditure which are flow variables originally following the NIPA convention, are reported as annualised, and thus the quarterly frequency value is the average and not the sum of the corresponding monthly observations.

Now, following the literature on mixed-frequency VARs, a state-space model can be formed as follows[70]:

$$Z_t = GS_t^{ssm} + e_t, \quad e_t \sim \mathcal{N}(0, R) \tag{3.102}$$

$$S_t^{ssm} = C + FS_{t-1}^{ssm} + v_t, \quad v_t \sim \mathcal{N}(0, Q) \tag{3.103}$$

The state equation in equation (3.103) can easily accommodate the VAR process in equation (3.98) once it is re-written in the usual VAR companion form where the matrices S_t^{ssm}, C, F, v_t and Q are defined as follows:

$$S_t^{ssm} = \left[Y_t', Y_{t-1}', \cdots, Y_{t-P+1}'\right]', \; C = \left[c', 0_{1\times N(P-1)}\right]', \; F = \left[\begin{array}{c|c} B_1, \cdots, B_{P-1} & B_P \\ \hline I_{N(P-1)} & 0_{N(P-1)\times N} \end{array}\right], \; v_t = \left[\varepsilon_t', 0_{1\times N(P-1)}\right]',$$

and $Q = \left[\begin{array}{c|c} \Sigma_{N\times N} & 0_{N\times N(P-1)} \\ \hline 0_{N(P-1)\times N} & 0_{N(P-1)\times N(P-1)} \end{array}\right]$

Regarding the observation equation (3.102), it can be further separated into two blocks, q and m, corresponding to the n_1 quarterly variables that we want to interpolate in monthly frequency and the n_2 monthly variables, which are observed directly in monthly frequency. In more detail, equation (3.102) can be written as follows:

$$\left[\begin{array}{c} Z_t^q \\ \hline Z_t^m \end{array}\right] = \left[\begin{array}{c} G^q \\ \hline G^m \end{array}\right] S_t + \left[\begin{array}{c} e_t^q \\ \hline e_t^m \end{array}\right], \; \left[\begin{array}{c} e_t^q \\ \hline e_t^m \end{array}\right] \sim \mathcal{N}\left(\left[\begin{array}{c} 0_{n_1} \\ \hline 0_{n_2} \end{array}\right], \left[\begin{array}{c|c} R^q & 0_{n_1\times n_2} \\ \hline 0_{n_2\times n_1} & R^m \end{array}\right]\right) \tag{3.104}$$

Block q in detail corresponds to the $1,\cdots, n_1$ quarterly variables, thus $Z_t^q = Y_t^o$:

$$Z_t^q = Y_t^o = \begin{bmatrix} Y_{1,t}^o \\ \vdots \\ Y_{n_1,t}^o \end{bmatrix} = \begin{bmatrix} G_1^q \\ \vdots \\ G_{n_1}^q \end{bmatrix} S_t + \begin{bmatrix} e_{1,t}^q \\ \vdots \\ e_{n_1,t}^q \end{bmatrix}, \; \begin{bmatrix} e_{1,t}^q \\ \vdots \\ e_{n_1,t}^q \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} R_1^q & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_{n_1}^q \end{bmatrix}\right)$$

[70] To avoid confusion with the MODWT-MRA smooth coefficients denoted also with S, all quantities related to the state-space model's state vector are denoted using superscript *ssm* as S_t^{ssm}.

Where $G_i^q = \left[0_{1\times d_1}, 1/3, 0_{1\times d_2}, 2/3, 0_{1\times d_2}, 1, 0_{1\times d_2}, 2/3, 0_{1\times d_2}, 1/3, 0_{1\times d_3}\right]$ with $d_1 = (i-1)$ for $i = 1,\cdots, n_1$, $d_2 = N-1$ and $d_3 = n_1 - n + n_2 + N(P-5)$ applies the intertemporal restriction discussed in equation (3.101). Moreover, matrix R^q is a square zero matrix of dimensions $n_1 \times n_1$.

In continuation, block m corresponds to $1,\cdots, n_2$ monthly variables, which are observed directly in monthly frequency, i.e. $Z_t^m = \left[Y_{1,t}^m, \cdots, Y_{n_2,t}^m\right]'$. Furthermore, $G^m = \left[0_{n_2\times n_1}, I_{n_2}, 0_{n_2\times N(P-1)}\right]$ and R^m is a square zero matrix of dimensions $n_2 \times n_2$.

As a convention, since the application of this study focuses on achieving ex-post the highest accuracy of the estimated monthly time series originally observed in quarterly frequency, the following are assumed. For the low frequency (quarterly) time series described with block q in equation (3.104), their observations Z_t^q are available at the last (third) month of each respective quarter, while the remaining months it is assumed to have missing observations. The high frequency (monthly) time series' observations Z_t^m described with block m in equation (3.104) are available for the corresponding month without any publication delay. To be more concrete, Table 3.10 below shows the observations vector $Z_t = \left[Z_t^{q\prime}, Z_t^{m\prime}\right]'$ available for the first and second quarters of a given calendar year.

Table 3.10: MF-VAR quarterly and monthly time series observations

$\cdots Q_4$	Q_1			Q_2			$Q_3 \cdots$
$\cdots M_{12}$	M_1	M_2	M_3	M_4	M_5	M_6	$M_7 \cdots$
Y_t^o	$\varnothing$	$\varnothing$	Y_{t+3}^o	$\varnothing$	$\varnothing$	Y_{t+6}^o	$\varnothing$
Y_t^m	Y_{t+1}^m	Y_{t+2}^m	Y_{t+3}^m	Y_{t+4}^m	Y_{t+5}^m	Y_{t+6}^m	Y_{t+7}^m

Where $Z_t^q = Y_t^o$ and $Z_t^m = Y_t^m$ as defined in equations (3.99) and (3.100), refer to the n_1 quarterly and n_2 monthly variables, respectively.

The MF-VAR is estimated using standard Bayesian techniques mentioned in the relevant literature. More details on the estimation algorithm are presented in the following section discussing the wavelet-based MF-VAR, which is a more generic case and encapsulates the MF-VAR approach.

3.5.2 Wavelet-based mixed-frequency VAR

Now, turning to the proposed wavelet-based MF-VAR, the main idea is to interpolate separately all or a subgroup of the higher frequency wavelet coefficients using a VAR model while maintaining the same intertemporal restriction as for the case of the MF-VAR. The main argument behind this approach lies in the fact that the detail and scaling coefficients describe time series fluctuations in defined ranges of frequencies.

To be more precise, for the case of an MF-VAR with monthly and quarterly frequency time series, Table 3.11 below describes the range of frequencies captured at each level of the MODWT-MRA decomposition.

Table 3.11: Monthly and quarterly series $J = 1, 2$ Haar MODWT-MRA components frequency range

Decomp. level	Monthly freq. time series	Quarterly freq. time series
	$J = 1$	
D_1	2-4 months	2-4 quarters (6-12 months)
S_1	>4 months	>4 quarters (>12 months)
	$J = 2$	
D_1	2-4 months	2-4 quarters (6-12 months)
D_2	4-8 months	4-8 quarters (12-24 months)
S_2	>8 months	>8 quarters (>24 months)

As the table shows, there is not a perfect overlap between the range of frequencies described by monthly and quarterly frequency time series' detail and smooth coefficients. The monthly time series S_1 smooth coefficients provide informational content for fluctuations greater than four months which can be captured sufficiently by quarterly time series D_1 detail coefficients. On the other hand, quarterly time series D_1 detail coefficients cannot describe fluctuations with a period smaller than six months[71]. Nonetheless, in order to produce the most accurate estimates of the underlying monthly latent processes, it is important to make the best use of the available monthly time series informational content.

Consequently, by separating the time series informational content into components with fluctuations in a range of frequencies, the wavelet-based MF-VAR aims to better inform the quarterly time series' monthly latent process from the high-frequency fluctuations from the remaining time series observed originally in monthly frequency. The slower moving trends captured in the smooth coefficients both for monthly and quarterly frequency time series, as shown in the table above, in theory, should not affect and interfere with the higher frequency fluctuations that are of interest to make the best usage of monthly time series informational content.

To become more concrete, suppose an N-multivariate time series $Y_t = \left[Y_{1,t}, Y_{2,t}, \cdots, Y_{N,t} \right]'$ in monthly frequency, then applying the Haar MODWT-MRA[72], each univariate time series $Y_{n,t}$ can be decomposed as:

[71] This is a direct outcome of the Nyquist-Shannon sampling theorem stating that in order to capture information from a time series with maximum frequency f_{max}, the time series should be sampled at twice this frequency: $f_s > 2 f_{max}$ (Unser, 2000). In other words, a quarterly time series (three month sampling interval) cannot have informational content with periodicity smaller than six months (two quarters), which is the lower limit of the quarterly series D_1 detail coefficients.

[72] The Haar MODWT additive decomposition of wavelet and scaling coefficients was also used but resulted in inferior results. In principle, when using the MODWT-MRA any wavelet beyond the Haar can be used instead.

$$Y_{n,t} = \sum_{j=1}^{J} D_{n,j,t} + S_{n,J,t} \tag{3.105}$$

Now, the corresponding N-multivariate time series of the detail and smooth coefficients at each level $j = 1,\cdots, J$ can be formed as:

$$D_{j,t} = \left[D_{1,j,t}, D_{2,j,t}, \cdots, D_{N,j,t} \right]' \quad \text{and} \quad S_{J,t} = \left[S_{1,J,t}, S_{2,J,t}, \cdots, S_{N,J,t} \right]' \tag{3.106}$$

Using the newly formed N-multivariate time series $D_{j,t}$ and $S_{J,t}$, a separate VAR model can be formed at each scale j to model/forecast jointly the N-multivariate detail and smooth coefficients time series. To be more concrete for the detail coefficients $D_{j,t}$ at each scale $j = 1,\cdots, J$ a VAR is formed:

$$D_{j,t} = c_j + \sum_{p=1}^{P_j} B_{j,p} D_{j,t-p} + \varepsilon_{j,t} \tag{3.107}$$

Similarly, for the smooth coefficients $S_{J,t}$, a VAR is formed:

$$S_{J,t} = c_{J+1} + \sum_{p=1}^{P_{J+1}} B_{J+1,p} S_{J,t-p} + \varepsilon_{J+1,t} \tag{3.108}$$

Using the standard VAR setting, c_j are of $N \times 1$ dimensions, $B_{j,p}$ are $N \times N$ matrices and the innovations $\varepsilon_{j,t}$ are $N \times 1$ vectors normally distributed as $e_{j,t} \sim \mathcal{N}(0, \Sigma j)$.

Now assume that out of the $n = 1,\cdots, N$ univariate time series comprising the N-multivariate time series, the first n_1 time series are observed in quarterly frequency, while the remaining $n_2 = N - n_1$ are observed directly in monthly frequency:

$$Y_t = \left[Y_{1,t}^q, \cdots, Y_{n_1,t}^q, Y_{1,t}^m, \cdots, Y_{n_2,t}^m \right]'$$

Where in terms of notation $Y_{1,t}^q$ denotes the underlying time series which are observed in quarterly frequency, whereas $Y_{1,t}^m$ denotes the time series observed directly in monthly frequency. Following this notation, it also implied that detail and smooth coefficients can be divided into quarterly and monthly subvectors, respectively:

$D_{j,t} = \left[D_{j,t}^{q\prime}, D_{j,t}^{m\prime} \right]'$ where $D_{j,t}^q = \left[D_{1,j,t}^q, \cdots, D_{n_1,j,t}^q \right]'$ and $D_{j,t}^m = \left[D_{1,j,t}^m, \cdots, D_{n_2,j,t}^m \right]'$.

$S_{J,t} = \left[S_{J,t}^{q\prime}, S_{J,t}^{m\prime} \right]'$ where $S_{J,t}^q = \left[S_{1,J,t}^q, \cdots, S_{n_1,J,t}^q \right]'$ and $S_{J,t}^m = \left[S_{1,J,t}^m, \cdots, S_{n_2,J,t}^m \right]'$.

The observed values of the time series observed in the quarterly frequency are denoted as follows:

$$Y_t^o = \left[Y_{1,t}^o, \cdots, Y_{n_1,t}^o \right]' \tag{3.109}$$

As already discussed previously, for the case of the MF-VAR, following Mariano and Murasawa (2003, 2010), a univariate quarterly time series transformed into growth rates can be mapped on its corresponding monthly latent process via the following approximation:

$$Y_{1,t}^{o} = \frac{1}{3}Y_{1,t}^{q} + \frac{2}{3}Y_{1,t-1}^{q} + Y_{1,t-2}^{q} + \frac{2}{3}Y_{1,t-3}^{q} + \frac{1}{3}Y_{1,t-4}^{q} \tag{3.110}$$

By substituting in equation (3.110) the Haar MODWT-MRA additive decomposition of equation (3.105), it implies:

$$\begin{aligned} Y_{1,t}^{o} = {} & \frac{1}{3}\left(\sum_{j=1}^{J} D_{1,j,t}^{q} + S_{1,j,t}^{q}\right) + \frac{2}{3}\left(\sum_{j=1}^{J} D_{1,j,t-1}^{q} + S_{1,j,t-1}^{q}\right) + \sum_{j=1}^{J} D_{1,j,t-2}^{q} + S_{1,j,t-2}^{q} \\ & + \frac{2}{3}\left(\sum_{j=1}^{J} D_{1,j,t-3}^{q} + S_{1,j,t-3}^{q}\right) + \frac{1}{3}\left(\sum_{j=1}^{J} D_{1,j,t-4}^{q} + S_{1,j,t-4}^{q}\right) \end{aligned} \tag{3.111}$$

Now, following the literature on mixed-frequency VARs, a state-space model can be formed as follows:

$$Z_t = GS_t^{ssm} + e_t, \quad e_t \sim \mathcal{N}(0, R) \tag{3.112}$$

$$S_t^{ssm} = C + FS_{t-1}^{ssm} + v_t, \quad v_t \sim \mathcal{N}(0, Q) \tag{3.113}$$

Beginning with the state equation (3.113), it is straightforward that the separate scale j VAR models can be casted easily by writing each VAR as a separate block in the state equation and, in continuation, stacking them all together:

$$\underbrace{\begin{bmatrix} \boldsymbol{S}_{1,t}^{ssm} \\ \vdots \\ \boldsymbol{S}_{J,t}^{ssm} \\ \boldsymbol{S}_{J+1,t}^{ssm} \end{bmatrix}}_{S_t^{ssm}} = \underbrace{\begin{bmatrix} \boldsymbol{C}_1 \\ \vdots \\ \boldsymbol{C}_{J,t} \\ \boldsymbol{C}_{J+1,t} \end{bmatrix}}_{C} + \underbrace{\begin{bmatrix} \boldsymbol{F}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \boldsymbol{F}_J & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{F}_{J+1} \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} \boldsymbol{S}_{1,t-1}^{ssm} \\ \vdots \\ \boldsymbol{S}_{J,t-1}^{ssm} \\ \boldsymbol{S}_{J+1,t-1}^{ssm} \end{bmatrix}}_{S_{t-1}^{ssm}} + \underbrace{\begin{bmatrix} \boldsymbol{v}_{1,t} \\ \vdots \\ \boldsymbol{v}_{J,t} \\ \boldsymbol{v}_{J+1,t} \end{bmatrix}}_{v_t}, \tag{3.114}$$

$$\underbrace{\begin{bmatrix} \boldsymbol{v}_{1,t} \\ \vdots \\ \boldsymbol{v}_{J,t} \\ \boldsymbol{v}_{J+1,t} \end{bmatrix}}_{v_t} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}}_{0}, \underbrace{\begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \boldsymbol{Q}_J & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Q}_{J+1} \end{bmatrix}}_{Q} \right)$$

Each block $\boldsymbol{S}_{j,t}^{ssm}$ for $j = 1, \cdots, J$ corresponds to the separate scale detail coefficients $D_{j,t}$ VAR defined in equation (3.107). Additionally, the block $\boldsymbol{S}_{J+1,t}^{ssm}$ for $j = J+1$ corresponds to the smooth coefficients $S_{J,t}$ VAR defined in equation (3.108).

For the case of exposition, below, it is shown in more detail how each block $\boldsymbol{S}_{j,t}^{ssm}$ is formed. For the case of $j = J+1$ referring to block $\boldsymbol{S}_{J+1,t}^{ssm}$, obviously, in what follows, the detail coefficients $D_{j,t}$ should be substituted with the smooth coefficients $S_{J,t}$.

$$\boldsymbol{S}_{j,t}^{ssm} = \boldsymbol{C}_j + \boldsymbol{F}_j \boldsymbol{S}_{j,t-1}^{ssm} + \boldsymbol{v}_{j,t}, \ \boldsymbol{v}_{j,t} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{Q}_j\right)$$

Where: $\boldsymbol{S}_{j,t}^{ssm} = \left[D'_{j,t}, D'_{j,t-1}, \cdots, D'_{j,t-P_j+1} \right]'$, $\boldsymbol{C}_j = \left[c'_j, 0_{1\times N(P_j-1)} \right]'$,

$\boldsymbol{F}_j = \left[\begin{array}{c|c} B_{j,1}, \cdots, B_{j,P_j-1} & B_{j,P_j} \\ \hline I_{N(P_j-1)} & 0_{N(P_j-1)\times N} \end{array}\right]$, $\boldsymbol{v}_{j,t} = \left[\varepsilon'_{j,t}, 0_{1\times N(P_j-1)} \right]'$ and $\boldsymbol{Q}_j = \left[\begin{array}{c|c} \Sigma_j & 0_{N\times N(P_j-1)} \\ \hline 0_{N(P_j-1)\times N} & 0_{N(P_j-1)\times N(P_j-1)} \end{array}\right]$

.

Turning now to the measurement equation (3.113), for $j = 1 \sim j^*$ for the detail coefficients, $D_{j,t}^q$ the vector enters with missing observations, while for $j = j^* + 1, \cdots, J+1$, the relevant components are assumed to be observed. In more detail, the measurement column vector Z_t of dimension $n_1 + (J+1)N$ is given below:

$$Z_t = \left[Y_t^{o\prime}, D'_{1,t}, \cdots, D'_{j^*,t}, D'_{j^*+1,t}, \cdots, D'_{J,t}, S'_{J,t} \right]' \tag{3.115}$$

Then the measurement equation can be partitioned into two blocks, q (referring to the quarterly time series) and w (referring to the wavelet transform detail and smooth coefficients), as follows:

$$\underbrace{\left[\begin{array}{c} \boldsymbol{Z}_t^q \\ \hline \boldsymbol{Z}_t^w \end{array}\right]}_{Z_t} = \underbrace{\left[\begin{array}{c} \boldsymbol{G}^q \\ \hline \boldsymbol{G}^w \end{array}\right]}_{G} S_t^{ssm} + \underbrace{\left[\begin{array}{c} \boldsymbol{e}_t^q \\ \hline \boldsymbol{e}_t^w \end{array}\right]}_{e_t}, \underbrace{\left[\begin{array}{c} \boldsymbol{e}_t^q \\ \hline \boldsymbol{e}_t^w \end{array}\right]}_{e_t} \sim \mathcal{N}\left(\underbrace{\left[\begin{array}{c} \boldsymbol{0} \\ \hline \boldsymbol{0} \end{array}\right]}_{0}, \underbrace{\left[\begin{array}{c|c} \boldsymbol{R}^q & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{R}^w \end{array}\right]}_{R} \right) \tag{3.116}$$

Block q in equation (3.116) applies the intertemporal restriction discussed in equations (3.110) and (3.111). The observations vector is set equal to the observed quarterly time series, i.e. $\boldsymbol{Z}_t^q = \boldsymbol{Y}_t^o$. Expanding block q for each univariate time series $n = 1, \cdots, n_1$ yields:

$$\boldsymbol{Z}_t^q = \boldsymbol{Y}_t^o = \underbrace{\begin{bmatrix} Y_{1,t}^o \\ \vdots \\ Y_{n_1,t}^o \end{bmatrix}}_{Z_t^q} = \underbrace{\begin{bmatrix} G_1^q \\ \vdots \\ G_{n_1}^q \end{bmatrix}}_{G^q} S_t^{ssm} + \underbrace{\begin{bmatrix} e_{1,t}^q \\ \vdots \\ e_{n_1,t}^q \end{bmatrix}}_{e^q}, \underbrace{\begin{bmatrix} e_{1,t}^q \\ \vdots \\ e_{n_1,t}^q \end{bmatrix}}_{e^q} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}}_{0}, \underbrace{\begin{bmatrix} R_1^q & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_{n_1}^q \end{bmatrix}}_{R^q} \right)$$

Where row vector $G_n^q = \left[G_{n,1}^q, \cdots, G_{n,J}^q, G_{n,J+1}^q \right]$ for $n = 1, \cdots, n_1$ contains $J+1$ row subvectors in line with equation (3.111):

$$G_{n,j}^q = \left[0_{1\times d_1}, 1/3, 0_{1\times d_2}, 2/3, 0_{1\times d_2}, 1, 0_{1\times d_2}, 2/3, 0_{1\times d_2}, 1/3, 0_{1\times d_3} \right]$$

Where $d_1 = (n-1)$ for $i = 1, \cdots, n_1$, $d_2 = n_1 - 1 + n_2$, $d_3 = n_1 - n + n_2 + N(P_j - 5)$.

Since there is not any noise corrupting the observations, matrix R^q is a square zero matrix of dimensions $n_1 \times n_1$.

Turning now to block w in equation (3.116), it contains the detail and smooth coefficients, which for the quarterly time series up to level j^* are assumed to be completely missing. Thus, the observation column vector Z_t^w is given as:

$$\boldsymbol{Z}_t^w = \left[Z_{1,t}^{w\prime}, \cdots, Z_{J+1,t}^{w\prime} \right]' = \left[D_{1,t}', \cdots, D_{j^*,t}', D_{j^*+1,t}', \cdots, D_{J,t}', S_{J,t}' \right]'$$

Consequently, the entire block w in equation (3.116) can be further partitioned into the $J+1$ blocks described below:

$$\underbrace{\begin{bmatrix} Z_{1,t}^w \\ \vdots \\ Z_{J,t}^w \\ Z_{J+1,t}^w \end{bmatrix}}_{Z_t^w} = \underbrace{\begin{bmatrix} G_1^w \\ \vdots \\ G_J^w \\ G_{J+1}^w \end{bmatrix}}_{G^w} S_t + \underbrace{\begin{bmatrix} e_{1,t}^w \\ \vdots \\ e_{J,t}^w \\ e_{J+1,t}^w \end{bmatrix}}_{e^w}, \underbrace{\begin{bmatrix} e_{1,t}^w \\ \vdots \\ e_{J,t}^w \\ e_{J+1,t}^w \end{bmatrix}}_{e^w} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}}_{0}, \underbrace{\begin{bmatrix} R_1^w & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & 0 \\ \vdots & \vdots & R_J^w & 0 \\ 0 & 0 & 0 & R_{J+1}^w \end{bmatrix}}_{R^w} \right)$$

For a given scale j, the relevant partition of the observation equation of dimension N is given as:

$$Z_{j,t}^w = G_j^w S_t^{ssm} + e_j^w, \quad e_j^w \sim \mathcal{N}\left(0, R_j^w\right)$$

Where $G_j^w = \left[0_{N \times d_1}, I_N, 0_{N \times d_2} \right]$ with $d_1 = \sum_{i=1}^{j-1} NP_j$ and $d_2 = N\left(P_j - 1\right) + N \sum_{i=j+1}^{J+1} P_i$ and

$R_j^w = 0_{N \times N}$.

To be more concrete, the entire observations vector Z_t defined in equation (3.115) for subsequent observations over the first and second quarter of a calendar year is given as follows:

Table 3.12: State-space model observations vector

$\cdots Q_4$	Q_1			Q_2			$Q_3 \cdots$
$\cdots M_{12}$	M_1	M_2	M_3	M_4	M_5	M_6	$M_7 \cdots$
Y_t^o	$\varnothing$	$\varnothing$	Y_{t+3}^o	$\varnothing$	$\varnothing$	Y_{t+6}^o	$\varnothing$
$\begin{bmatrix} D_{1,t} \\ \vdots \\ D_{J,t} \\ S_{J,t} \end{bmatrix}$	$\begin{bmatrix} D_{1,t+1} \\ \vdots \\ D_{J,t+1} \\ S_{J,t+1} \end{bmatrix}$	$\begin{bmatrix} D_{1,t+2} \\ \vdots \\ D_{J,t+2} \\ S_{J,t+2} \end{bmatrix}$	$\begin{bmatrix} D_{1,t+3} \\ \vdots \\ D_{J,t+3} \\ S_{J,t+3} \end{bmatrix}$	$\begin{bmatrix} D_{1,t+4} \\ \vdots \\ D_{J,t+4} \\ S_{J,t+4} \end{bmatrix}$	$\begin{bmatrix} D_{1,t+5} \\ \vdots \\ D_{J,t+5} \\ S_{J,t+5} \end{bmatrix}$	$\begin{bmatrix} D_{1,t+6} \\ \vdots \\ D_{J,t+6} \\ S_{J,t+6} \end{bmatrix}$	$\begin{bmatrix} D_{1,t+7} \\ \vdots \\ D_{J,t+7} \\ S_{J,t+7} \end{bmatrix}$

Where for $j = 1 \sim j^*$, $D_{j,t} = \left[D_{j,t}^{q\prime}, D_{j,t}^{m\prime} \right]' = \left[\varnothing_{1 \times n_1}, D_{j,t}^{m\prime} \right]'$ with $\varnothing_{1 \times n_1}$ denoting missing observations for the n_1 quarterly time series. Regarding the n_2 monthly time series, it is straightforward that the MODWT-MRA decomposition can be applied, and the resulting detail and smooth coefficients can be used as observations at all scales J, i.e. $D_{j,t}^m$ for $j = 1,\cdots, J$ and $S_{J,t}^m$.

Turning now to the quarterly time series, in principle, it can be assumed that $j^* = J + 1$, such that for all of the detail and smooth coefficients, latent monthly processes corresponding to the quarterly time series enter as missing observations. Alternatively, it can be assumed that for $j = j^* + 1,\cdots, J + 1$, the detail and the smooth coefficients are known in monthly frequency. Such an approach requires the application of the MODWT-MRA decomposition as if the monthly latent process of the quarterly time series were known. In that case, the MODWT-MRA needs to be applied to monthly estimates of quarterly time series estimated from a previous external step. For instance, in the simplest case, it can be assumed that all months of a quarter have values equal to 1/3 of the corresponding quarterly observation or use the Chow and Lin (1971) interpolation method. Alternatively, the MF-VAR monthly latent process estimates could be used instead.

What is critical to understand is that, as it will be shown in the empirical application Section 3.6, even for a $J = 2$ level wavelet decomposition where the unknown monthly values are assumed to be equal to the 1/3 of quarterly observation, the smooth coefficients $S_{2,t}$ are highly correlated with what would have been the actual $S_{2,t}$ estimated from the actual monthly time series.

Concerning the model estimation, this can be performed easily via a standard Gibbs sampler consisting of three blocks. Each iteration (r) is based on the following three blocks:

1. $p\left(S^{ssm(r)} \middle| B^{(r-1)}, \Sigma^{(r-1)} \right)$
2. $p\left(B_j^{(r)} \middle| S^{ssm(r)}, \Sigma_j^{(r-1)} \right)$, $j = 1,\cdots, J + 1$
3. $p\left(\Sigma_j^{(r)} \middle| S^{ssm(r)}, B_j^{(r)} \right)$, $j = 1,\cdots, J + 1$

Where B_j and Σ_j refer to the set of parameters for each one of the $J + 1$ in total separate VARs estimated for the detail and smooth coefficients respectively: $B = \{B_j, \cdots, B_{J+1}\}$ and $\Sigma = \{\Sigma_j, \cdots, \Sigma_{J+1}\}$. It has to be noted that since each one of the VARs which needs to be estimated is orthogonal to the remaining, the second and the third blocks described above can be easily parallelised for computational efficiency. Parallelisation can occur intuitively by estimating the $J + 1$ VARs in parallel rather than in a serial manner. Furthermore, to increase the computational efficiency, the equation-by-equation estimation technique for large VARs (Carriero, Clark, & Marcellino, 2019; Carriero, Chan, Clark, & Marcellino, 2022) can be employed. The prior used for the $J + 1$ VARs is a conventional Minnesota prior (independent normal - inverse Wishart) whose hyperparameters follow conventional values from the literature. More information can be found in Section 3.6.2, discussing priors and other estimation details.

Finally, regarding the posterior estimation of the states, while there exist numerous Kalman filter simulation smoother algorithms, following the findings of Bańbura, Giannone, and Lenza

(2015), I use the Durbin and Koopman (2002) simulation smoother with the Kalman filter fixed interval smoother of de Jong (1988). More details can be found in Appendix C.1.4 Kalman filtering and smoothing.

3.6 Empirical application mixed-frequency time series

This section presents an empirical application of using the proposed wavelet-based MF-VAR for modelling mixed-frequency time series. Section 3.6.1 describes the US mixed-frequency macroeconomic time series which were used, and Section 3.6.2 provides more details about the proposed model specifications and priors used. In continuation, Section 3.6.3 discusses the external monthly estimates which were used as input to some specifications of the wavelet-based MF-VAR. Finally, Section 3.6.4 presents the findings on the proposed model's forecast accuracy for the monthly estimates of the originally assumed quarterly frequency time series.

3.6.1 Data description

All models, i.e. the MF-VAR approach and the various specifications of the proposed wavelet-based MF-VAR model, are estimated using a set of 7 variables. The first 4 are quarterly frequency variables which are estimated in monthly frequency and consist of the real personal consumption expenditure (PCE), industrial production (IP), CPI inflation and West Texas Intermediate (WTI) crude oil price. The remaining three monthly frequency variables, which are used to inform the monthly estimates of the quarterly variables, are the Federal Funds Rates (FFR), 10-year bond yield, and total nonfarm employment. All the time series, both quarterly and monthly that enter the MF-VAR models are transformed into growth rates (100 Δlog), except for the FFR and 10-year bond yield, which are expressed in percentage points – see Table 3.13, first panel.

In addition, both the MF-VAR and the wavelet-based MF-VAR monthly estimates of the four quarterly variables are compared with their corresponding actual monthly time series expressed in growth rates (100 Δlog), as well – see Table 3.13, the second panel below.

Table 3.13: Mixed frequency time series – variables

Variable	Transformation	Frequency
MF-VAR models		
PCE	100 Δlog	Quarterly
Industrial production	100 Δlog	Quarterly
CPI	100 Δlog	Quarterly
WTI crude oil price	100 Δlog	Quarterly
FFR	-	Monthly
10-year bond yield	-	Monthly
Employment	100 Δlog	Monthly
Time series to evaluate MF-VAR monthly estimates		
PCE	100 Δlog	Monthly
Industrial production	100 Δlog	Monthly
CPI	100 Δlog	Monthly
WTI crude oil price	100 Δlog	Monthly

Source: Archival Federal Reserve Economic Data of the Federal Reserve Bank of St. Louis.

In order to have a better understanding of the monthly time series properties which will be estimated based on their corresponding quarterly time series, variance decomposition (see equation (3.34) for details) is performed by applying the MODWT to the actual monthly time series in Table 3.14 below.

Table 3.14: Variance decomposition

Variables	Mean	St. dev.	W_1: 2-4M	V_1: >4M
PCE	0.27	0.51	58.97	41.03
	(0.26)	(0.84)	(49.06)	(50.94)
IP	0.21	0.73	34.02	65.98
	(0.21)	(0.95)	(33.66)	(66.34)
CPI	0.31	0.31	18.65	81.35
	(0.3)	(0.31)	(18.51)	(81.49)
Oil	0.43	7.7	38.12	61.88
	(0.39)	(8.49)	(38.02)	(61.98)

Note. Normalised variance decomposition across $J = 1$ MODWT components (may not add up to 100% due to rounding). Data sample is 1960M1-2019M12 (excluding COVID-19 data) and 1960M1-2020M12 (including COVID-19 data) in parentheses.

Overall, it is clear that there is a huge variation among the variables under examination. For example, CPI variance is mostly explained by the scaling coefficients V_1 variance, by 81.35%, while on the contrary, PCE short-term fluctuations captured in wavelet coefficients W_1 explain more than half (58.97%) of the total time series variance, excluding the impact of COVID-19 data. Ordering the four variables by increasing variance in W_1 (and decreasing variance in V_1, respectively) as CPI, industrial production, oil price and PCE is important to explain the relative differences between variables in Sections 3.6.3 and 3.6.4 discussing the results. Finally, regarding the impact of the COVID-19 data on the time series, the effect is noticeable for PCE by an increase of almost 10% for frequencies greater than four months which explains the long-lasting behaviour of the COVID-19 shock.

3.6.2 Priors and other estimation details

In order to evaluate the behaviour of the wavelet-based MF-VAR versus the MF-VAR approach, which is used as the benchmark, numerous specifications are employed. In particular, the specifications span across three main dimensions, with the total specifications tested amounting to a total of 20 (5×2×2).

The first dimension, which includes five specifications, compares whether assuming all detail and smooth coefficient across all decomposition levels are missing provides better forecasting accuracy than assuming the detail levels up to level j^* are missing only, and the remaining $j = j^*+1,\cdots, J+1$ detail and smooth coefficients are based on external monthly estimates, i.e. the MF-VAR, Chow-Lin, or constant 1/3 approaches. In the same dimension, the fifth specification explores the theoretical case where only the $j = j^* = 1$ level detail coefficients are unknown, where all the remaining detail and smooth coefficients for $j = 2,\cdots, J+1$ are assumed to be estimated using the actual monthly time series. This allows to establish a lower bound in interpolating the quarterly time series only for the informational content included in the range of two to four months, whereas all the information for fluctuations above four months is assumed to be known with certainty since the actual monthly time series are used.

The second dimension investigates at which level J of the MODWT-MRA decomposition the forecast accuracy of the monthly estimates is higher. Two specifications for the $J = 1$ and $J = 2$ MODWT-MRA were tested in that respect. It is important to have in mind here that from a theoretical perspective, there exists a trade-off between the J levels of decomposition and the forecast accuracy. The larger the level of decomposition that is used, the more efficient information usage can be potentially achieved since independent VARs formed model narrower frequency bands; however, at the expense of parameters proliferation and the respective estimation uncertainty, which could undermine the theoretically more efficient use of the time series informational content across the different scales.

Finally, the third dimension investigates the impact of the latest COVID-19 data on the proposed model's accuracy. In particular, the same exercise is performed for the period 1960-2019, i.e. excluding COVID-19 data and the full sample 1960-2020, which includes COVID-19 data. For the latter case, the models are not re-estimated, but instead, the parameters estimated for the 1960-2020 period are used, and the 2020 time series observations are appended to the dataset, similar to the case Kalman filtering is used for the case of ragged-edged data or conditional forecasting.

Regarding the VARs prior, the MF-VAR and the various wavelet-based MF-VAR specifications use a conventional independent normal inverse Wishart – Minnesota type prior, which follows the standard hyperparameter values set in the literature. In particular, the hyperparameters regarding the prior of the coefficients B in equation (3.59) are set as follows: overall shrinkage $\lambda_1 = 0.1$, cross-variable shrinkage $\lambda_2 = 0.5$, lag decay $\lambda_3 = 1$ and the constants when used in the VAR specification use a diffuse prior (see below for more details about the wavelet-based MF-VAR). The innovations covariance Σ, which follows an inverse Wishart prior in equation (3.60), uses $N+2$ degrees of freedom, and the scale is set equal to the residuals of independent AR(1) regressions for each variable.

For the case of the MF-VAR, the prior of the variables available only at a quarterly frequency for which we want to estimate their monthly estimates are based on preliminary monthly time series with monthly values set equal to 1/3 of the corresponding quarterly value in growth rates. Regarding the wavelet-based MF-VARs, for the missing detail/smooth coefficients up

to level j^*, the VARs priors distribution uses the MODWT-MRA coefficients assuming preliminary monthly time series with monthly values set equal to 1/3 of the corresponding quarterly value in growth rate as previously.

The MF-VAR specification of equation (3.98) includes a constant. However, for the wavelet-based MF-VAR specifications, the VAR equations (3.107) modelling the detail coefficients do not need to include a constant since the detail coefficients are zero mean time series – see more in Section 3.2.4. Regarding the MF-VAR lag length, the MF-VAR uses $P = 12$ lags. The wavelet-based MF-VARs for all specifications (MODWT-MRA and $J = 1,2$) use 12 lags for $j \geq 2$. For $j = 1$, for which the detail coefficients refer to fluctuations in the range of 2-4 months, the lag length is set to $P_1 = 5$, which is the minimum lag length that can satisfy the time aggregation restriction discussed in the equations (3.110) and (3.111).

The simulation smoother algorithm follows an exact initialisation with the initial values of the time series. In order to compare the MF-VAR monthly estimates with the wavelet-based MF-VAR specifications for the same sample, the first 12 monthly estimate observations, which are used for the initialisation of the wavelet-based MF-VAR, are discarded.

For all the estimated models, the first 10000 draws are discarded as a burn-in sample, and the subsequent 5000 draws are used for posterior inference.

3.6.3 Results on external monthly estimates used as input to the wavelet-based MF-VAR

This section presents the results of the external monthly estimates, which are used later as input for the wavelet-based MF-VAR. External monthly estimates, in principle, can include any method that provides monthly estimates, which are later used as input to the wavelet-based MF-VAR to apply the MODWT-MRA on these external monthly estimates and treat the resulting details and/or smooth coefficients as observables.

In this study, external monthly estimates, as already discussed previously, refer to three approaches. The first one, called 'MF-VAR' below, denotes the MF-VAR monthly estimates, the second one to the Chow and Lin (1971) interpolation method and finally, the third one, called 'constant' assumes constant monthly observations set equal to 1/3 of the corresponding quarterly value in growth rates as a rough approximation.

In order to evaluate how good these external monthly estimates are such that their corresponding smooth coefficients can be treated as observables in the wavelet-based MF-VAR, the correlation between the MODWT-MRA coefficients based on the actual monthly time series and the external monthly estimates is estimated as a rough measure of their accuracy.

Table 3.15 below presents in more detail these correlation values for the MODWT-MRA detail and smooth coefficients between the actual monthly time series and the three approaches used as external monthly estimates for the quarterly frequency variables. As a reminder, for $J = 2$, the values of D_1 are the same as for the case of $J = 1$, and this is the reason that the relevant panel is missing from the table below to avoid repetition.

It is clearly evident that for $J = 1, 2$, the correlation of the smooth coefficients is strong enough and above 0.7 for any option of the monthly external estimates used. This finding is expected and in line with Table 3.11, discussed in Section 3.5.2. Even for the simplest case where the external monthly estimates are set equal to 1/3 of the corresponding quarterly variable, it is sufficient to capture the salient futures of the wavelet components which describe fluctuations with a period larger than four months, i.e. greater than the observation frequency of 3 months for a quarterly time series.

Turning to the $J = 1$ MODWT-MRA D_1, overall, it is observed that the correlation is less strong as expected across all external monthly estimate approaches used since these components capture fluctuations in the range of 2-4 months. Additionally, what stands out is how well MF-VAR monthly estimates inform the S_1 and D_2 coefficients across all variables except for CPI, in contrast to the other less computationally demanding methods. However, for the D_1 coefficients, quite surprisingly, MF-VAR monthly estimates do not present the highest correlation with the D1 coefficients obtained from the true monthly time series.

In order to reconcile these results with the findings of the variance decomposition in Table 3.14, we can focus on the $J = 1$ and the constant 1/3 external monthly estimate approach for S_1. It becomes easily evident that the order of the variables in terms of decreasing correlation (CPI, industrial production, oil price, PCE) is exactly the same as the variables' order in Table 3.14 in terms of decreasing scaling coefficients variance percentage. In other words, the larger the part of the time series total variance that can be explained by fluctuations with a period greater than four months, the better the accuracy of the constant external monthly estimate with respect to the actual monthly time series. Since the total variance explained by the high-frequency fluctuations in D_1 is small as a percentage, the sampling frequency every three months for quarterly time series is sufficient to capture fluctuations of the time series with periods greater than four months adequately.

Table 3.15: Correlation between the actual and the estimated monthly time series Haar MODWT-MRA coefficients

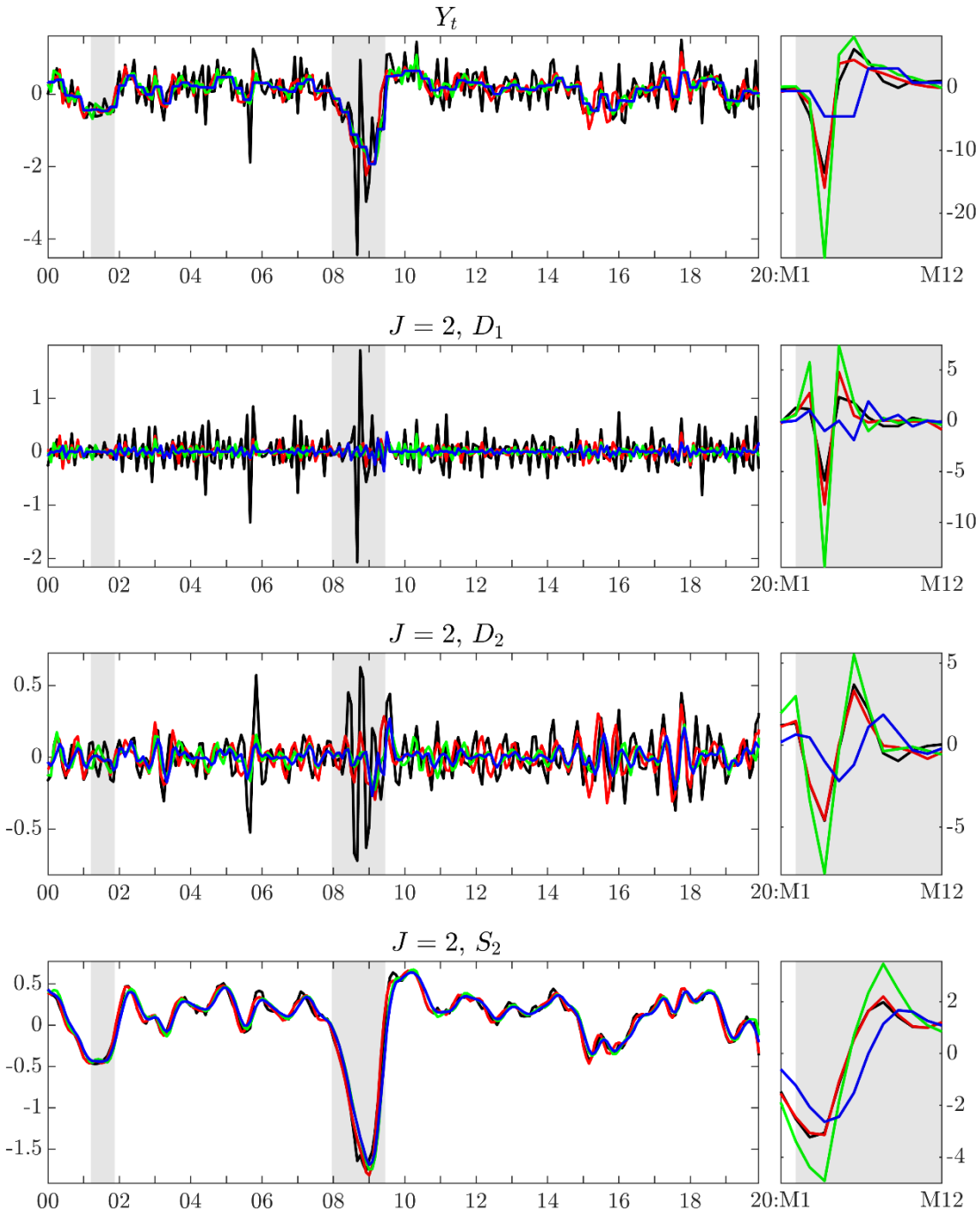
	PCE	**IP**	**CPI**	**OIL**
			***J* = 1**	
			D_1	
MF-VAR	0.09 (0.66)	0.37 (0.58)	0.03 (0.05)	0.1 (0.09)
Chow-Lin	0.09 (0.64)	0.34 (0.6)	0.11 (0.11)	0.16 (0.33)
Constant 1/3	0.11 (0.1)	0.13 (0.14)	0.11 (0.11)	0.16 (0.14)
			S_1	
MF-VAR	0.82 (0.92)	0.94 (0.95)	0.87 (0.88)	0.83 (0.8)
Chow-Lin	0.71 (0.81)	0.89 (0.91)	0.92 (0.92)	0.77 (0.77)
Constant 1/3	0.72 (0.41)	0.87 (0.74)	0.91 (0.91)	0.77 (0.69)
			***J* = 2**	
			D_2	
MF-VAR	0.51 (0.85)	0.64 (0.83)	0.16 (0.21)	0.47 (0.44)
Chow-Lin	0.27 (0.73)	0.57 (0.84)	0.44 (0.46)	0.43 (0.5)
Constant 1/3	0.27 (-0.15)	0.32 (0.14)	0.39 (0.38)	0.4 (0.23)
			S_2	
MF-VAR	0.98 (0.98)	0.99 (1)	0.99 (0.99)	0.98 (0.97)
Chow-Lin	0.93 (0.91)	0.96 (0.95)	0.98 (0.98)	0.9 (0.91)
Constant 1/3	0.93 (0.8)	0.96 (0.93)	0.98 (0.98)	0.91 (0.89)

Note. Data sample is 1960M1-2019M12 and 1960M1-2020M12 in parentheses.

Figure 3.19 below shows indicatively for the case of industrial production, the $J = 2$ Haar MODWT-MRA detail and smooth coefficients for the actual monthly time series in black and the monthly MF-VAR external estimates in red, the Chow-Lin monthly estimates in green, as well as the case of the 'constant' external monthly estimates approach in blue. The time series are shown only for the post-2000 period for visibility reasons.

It is clearly visible that all external monthly estimate methods fail to track the D_1 and D_2 coefficients closely, while the actual time series S_2 coefficients almost coincide with the S_2 coefficients estimated using the external monthly estimates. Focusing on periods where nowcasting could be highly informative in a real-time setting, comparing the 2008 financial crisis with the COVID-19 recession, it can be seen that for the former, the shock using the MF-VAR approach is underestimated, while for the latter is overestimated.

Figure 3.19: Industrial production actual and estimated monthly time series



Note. First subplot shows the actual industrial production actual (black), MF-VAR monthly estimates (red), Chow-Lin monthly estimates (green) and monthly values set equal to constant 1/3 of the corresponding quarterly values in growth rate (blue). Second, third and fourth subplots show the $J = 2$ Haar MODWT-MRA detail D_1, D_2 and smooth S_2 coefficients for the actual series in black and its monthly estimates in red and green and blue respectively.

3.6.4 Wavelet-based MF-VAR results

This section presents the key findings of the various wavelet-based MF-VAR specifications versus the MF-VAR approach in terms of in-sample point forecast accuracy measured with the root mean squared error (RMSE). The point estimates for the MF-VAR are estimated as the mean of the retained posterior draws. For the case of the wavelet-based MF-VAR, the point estimates are estimated by summing all the detail and smooth coefficients to form the monthly estimate for each posterior draw (r) of the Gibbs sampler and, in continuation, by averaging all of those monthly estimates. To be more specific, using (r) to denote a posterior draw, then the n_1-vector of the monthly estimates Y_t^q of the time series corresponding to the quarterly observed time series is defined as:

$$\left(Y_t^q\right)^{(r)} = \sum_{j=1}^{J} \left(D_{j,t}^q\right)^{(r)} + \left(S_{J,t}^q\right)^{(r)}$$

The point estimate is consequently the mean of the R in total $\left(Y_t^q\right)^{(r)}$ estimates:

$$Y_t^q = \frac{1}{R}\sum_{r=1}^{R}\left(Y_t^q\right)^{(r)}$$

Table 3.16 below shows the RMSE ratios of the various wavelet-based MF-VARs versus the MF-VAR used as the benchmark. To have a clearer image of the RMSE ratios in terms of the benchmark, raw RMSEs are also presented in the row[73]. As already discussed in Section 3.6.2, the wavelet-based MF-VAR specifications investigate the impact of external monthly estimates used as input for the wavelet-based MF-VAR, i.e. the MF-VAR monthly estimates, the Chow-Lin monthly estimates and the 'constant' 1/3 method, as well as the level of decomposition $J = 1, 2$. Additionally, to set a lower bound on the maximum accuracy that could potentially be achieved using as input to the wavelet-based MF-VAR, the actual monthly time series for the estimation of the MODWT-MRA smooth coefficients are used, and this case is referred in the table as 'actual'. Finally, the last specification in each subpanel referred to as 'missing', is the case where no external monthly estimates are assumed. This specification, in principle, provides as input to the wavelet-based MF-VAR exactly the same information as the MF-VAR.

Overall, it is evident, as expected, that for $J = 1, 2$, the forecast accuracy deteriorates (RMSE ratio values closer to unity) in a predictable way for the various specifications wavelet-based MF-VAR specifications. The 'actual' specification provides the maximum informational content to the wavelet-based MF-VAR, which estimates only the detail coefficients at $j = 1$, while the remaining wavelet coefficients (S_1 for $J = 1$ and D_1, S_2 for $J = 2$) are estimated based on the actual time series. This specification presents the lowest RMSE ratios, and the results are significant, with one exception across all variables and levels of decomposition J for the evaluation period excluding COVID-19 data.

In continuation, excluding the COVID-19 data, the specification using the MF-VAR external monthly estimates provides significant results across all variables for $J = 1$, while for $J = 2$, the results hold only for CPI and oil price. The remaining two external monthly estimate

[73] The MF-VAR raw RMSE large value for oil price is due to a large increase in the monthly time series between 1973M12 and 1974M12 from 4.31 to 10.11 USD per barrel.

methods, i.e. the Chow-Lin and 'constant 1/3', provide worse results vis-a-vis the MF-VAR method at a comparable level.

Finally, looking at the 'all missing' specification, where all of the wavelet coefficients for the quarterly variable need to be estimated by the Kalman filter, forecast accuracy overall deteriorates compared to the 'MF-VAR' specification for all of the variables as expected. For the case of PCE and industrial production, the proposed model does not provide any advantage at all compared to the benchmark with RMSE ratios close to unity. For the remaining variables, i.e. CPI and oil price, the forecast accuracy improvement remains still significant both for $J = 1, 2$ and the two evaluation periods excluding and including the COVID-19 data.

Examining the impact of the COVID-19 data on the proposed methods, it becomes clear that it has a profound worsening effect on the proposed model's in-sample forecast accuracy for the cases of PCE and industrial production, while CPI and oil price remain unaffected. Finally, trying to reconcile the following results with the variance decomposition, presented in Section 3.6.1, Table 3.14, for the two extreme cases of PCE and CPI, the results remain indeed consistent. CPI, for which the variance largest part is explained by scaling coefficients, i.e. less energy in high-frequency fluctuations, presents the biggest improvement across all specifications. On the contrary, PCE, for which more than half of the time series energy is contained in the high frequency 2-4 months detail coefficients, presents the least improvement versus the benchmark across all specifications.

Abstracting for a while from the detailed results for each specification and focusing on the greater picture, the proposed model provides advantages versus the benchmark and established MF-VAR approach. All of the proposed model specifications in the first subpanel ($J = 1$, i.e. as system of 2 VARs with D_1, S_1) do not exhibit forecast accuracy significantly worse than the benchmark model for the evaluation period, excluding COVID-19 data. The RMSE ratios across all specifications have as an upper bound the unity, which provides some empirical evidence that the proposed model, in the worst case, will be as good as the MF-VAR. There are few marginal cases with RMSE ratios slightly higher than unity; however, these small differences can be explained by the computational aspect of the problem and the parameter proliferation for the proposed models.

Nonetheless, the improvement of the proposed model across all variables requires, as a previous step, the external estimation of monthly estimates based on the MF-VAR approach, which is also used as the benchmark. It should be recognised that this 2-step approach, i.e. using external monthly estimates as input to the wavelet-based MF-VAR model, has some extra computational cost. It remains open for future research, which other more efficient external monthly estimate techniques could be used instead.

Table 3.16: Wavelet-based MF-VAR versus MF-VAR in-sample RMSE ratios

	PCE	IP	CPI	Oil
	Evaluation period 1960M1-2019M12 (Excluding COVID-19 data)			
Raw MF-VAR RMSE	0.445	0.486	0.256	6.231
	$J = 1$ (System of 2 VARs: D_1, S_1)			
Actual $S_1, j^*=1$	0.767***	0.736***	0.416***	0.618***
MF-VAR $S_1, j^*=1$	0.989*	0.958*	0.797***	0.932***
Chow-Lin $S_1, j^*=1$	1.003	0.965	0.649***	0.889***
Constant 1/3 $S_1, j^*=1$	1.001	1.022	0.653***	0.892***
Missing D_1, $S_1, j^*=2$	1.022	1.034	0.662***	0.877***
	$J = 2$ (System of 3 VARs: D_1, D_2, S_2)			
Actual $S_2, j^*=1$	0.966**	0.975	0.594***	0.822***
MF-VAR $S_2, j^*=1$	1.000	1.010	0.673***	0.881***
Chow-Lin $S_2, j^*=1$	1.006	1.019	0.645***	0.877***
Constant 1/3 $S_2, j^*=1$	1.006	1.032	0.644***	0.875***
Missing D_1, D_2, $S_2, j^*=3$	1.038	1.083	0.675***	0.881***
	Evaluation period 1960M1-2020M12 (Including COVID-19 data)			
Raw MF-VAR RMSE	0.505	0.596	0.237	7.029
	$J = 1$ (System of 2 VARs: D_1, S_1)			
Actual $S_1, j^*=1$	0.966	0.716***	0.449***	0.597***
MF-VAR $S_1, j^*=1$	1.178	0.860	0.859***	0.945***
Chow-Lin $S_1, j^*=1$	1.263	0.835*	0.699***	0.889***
Constant 1/3 $S_1, j^*=1$	1.594	1.161	0.711***	0.946*
Missing D_1, $S_1, j^*=2$	1.494	1.036	0.786***	0.889***
	$J = 2$ (System of 3 VARs: D_1, D_2, S_2)			
Actual $S_2, j^*=1$	1.444	1.002	0.649***	0.827***
MF-VAR $S_2, j^*=1$	1.426	1.002	0.760***	0.901***
Chow-Lin $S_2, j^*=1$	1.472	1.035	0.700***	0.896***
Constant 1/3 $S_2, j^*=1$	1.569	1.104	0.699***	0.901***
Missing D_1, D_2, $S_2, j^*=3$	1.544	1.106	0.848**	0.890***

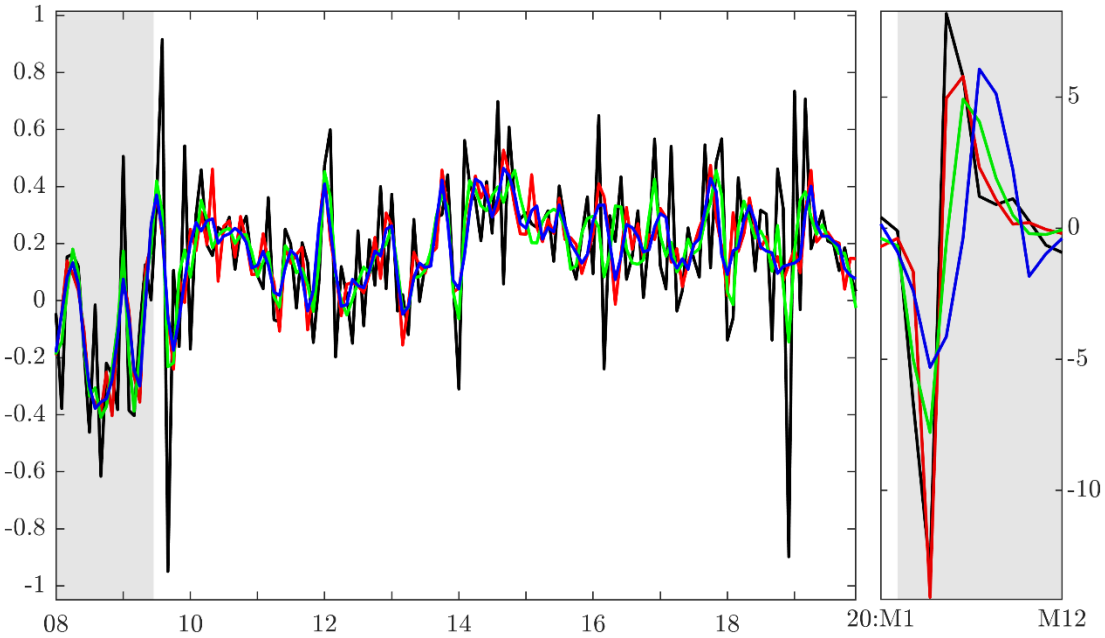
Note. Panels show RMSE ratios of wavelet-based MF-VARs with respect to the benchmark MF-VAR. Detail (*D.*) and smooth (*S.*) coefficients refer to the Haar MODWT-MRA. Wavelet-based MF-VAR input specifications include the Haar MODWT-MRA coefficients for *i*) Actual monthly time series, *ii*) MF-VAR monthly estimates, *iii*) Chow and Lin (1971) monthly estimates, *iv*) constant monthly values set equal to 1/3 of the respective quarterly value, and *v*) missing time series (all wavelet coefficients are missing). j^* refers up to which level j the wavelet coefficients are set as missing. Stars refer to the p-values of the Diebold and Mariano (1995) test with respect to the benchmark. *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively.

The figures below show across all quarterly variables, the actual monthly time series (black), the MF-VAR monthly estimates (red), and the monthly estimates of wavelet-based MF-VAR specifications using: *i*) the 'actual' time series for the estimation of S_1 coefficients (green), and *ii*) the MF-VAR external monthly estimates (blue). For visibility reasons, the figures show the post-2008 period, which includes the 2008 financial crisis and the COVID-19 recession.

As expected and already discussed in the findings of Table 3.16 above, the 'actual' wavelet-based MF-VAR specification monthly estimates in green are the ones closer to the actual monthly time series. However, even for this case, it is clear that the model is not able to capture the actual monthly time series behaviour closely, for example, during the recession periods. In other words, even allowing for the estimation only of the monthly detail coefficients D_1 capturing 2-4 months of the time series fluctuations, there still remains much pure statistical noise representing temporary economic shocks with periods smaller than two months that cannot be captured by wavelet analysis. The inclusion of stochastic volatility could be found to be beneficial during these periods.

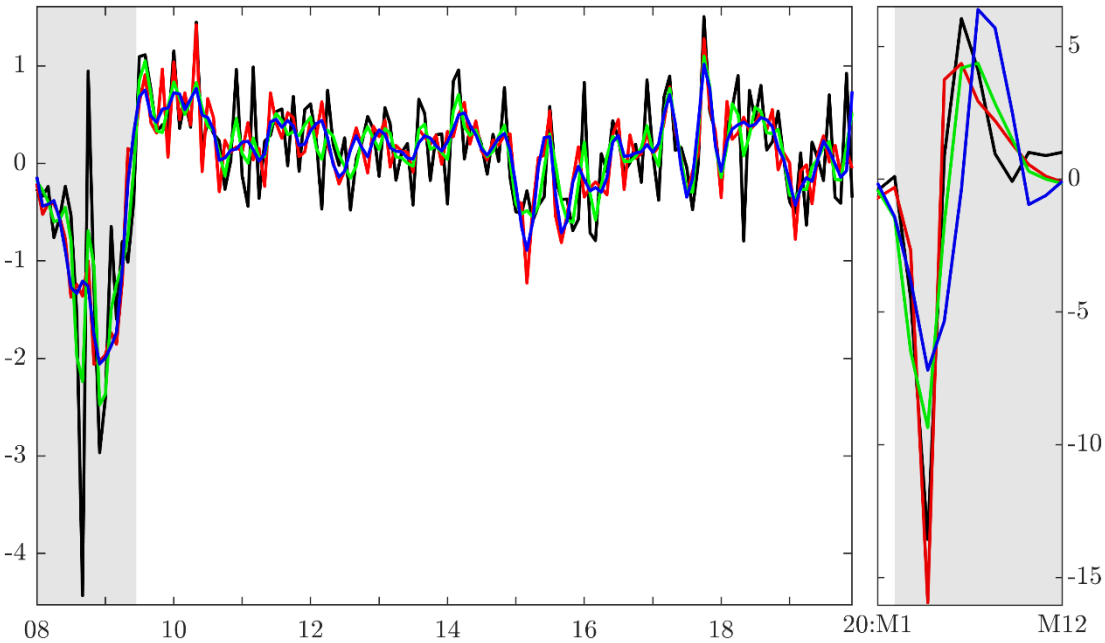
Turning to the two competing models, i.e. the MF-VAR VAR model and the wavelet-based MF-VAR based on the external monthly estimates of the MF-VAR move closer to each other, however as already indicated previously in the full sample analysis, their differences are found to be statistically significant. A closer examination reveals the MF-VAR monthly estimates tend to be sharper ('spiky') in changes although quite often not capturing the actual time series turning points, contrary to the wavelet-based MF-VAR estimates, which present a more smooth behaviour. For example, for PCE and industrial production, this behaviour is particularly accentuated for the post-2016 period, where the wavelet-based MF-VAR estimates in blue capture the underlying slow-moving trend for the time series rather than the exact turning points.

Figure 3.20: Personal consumption expenditure actual monthly time series, MF-VAR and wavelet-based MF-VAR monthly estimates



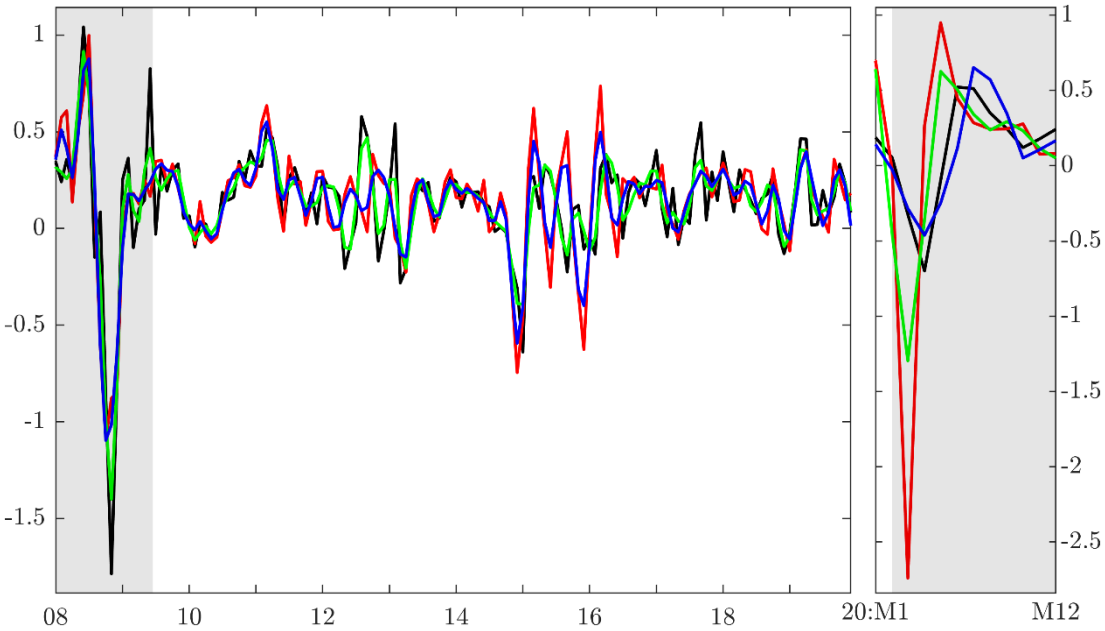
Note. Actual personal consumption expenditure is in black. MF-VAR monthly estimates are in red. Wavelet-based MF-VAR $J = 1$ using external monthly estimates with *i*) Actual monthly time series in green *ii*) MF-VAR monthly estimates in blue.

Figure 3.21: Industrial production actual monthly time series, MF-VAR and wavelet-based MF-VAR monthly estimates



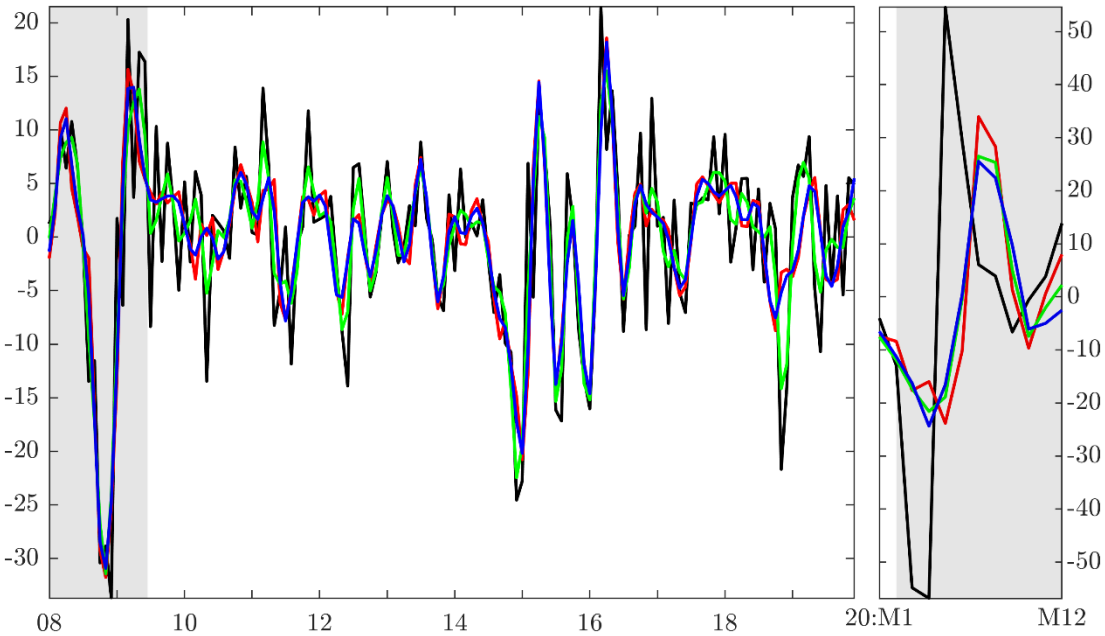
Note. See Figure 3.20 note details.

Figure 3.22: CPI actual monthly time series, MF-VAR and wavelet-based MF-VAR monthly estimates



Note. See Figure 3.20 note details.

Figure 3.23: Oil price actual monthly time series, MF-VAR and wavelet-based MF-VAR monthly estimates



Note. See Figure 3.20 note details.

3.7 Conclusion

So far in the literature, wavelet-based forecasting is focused mainly on univariate time series. This paper attempts to go beyond the case of univariate time series and extends the existing forecasting approaches to multivariate single and mixed-frequency time series relying on vector autoregressive techniques under a Bayesian estimation framework.

The three forecasting approaches which are investigated for single-frequency time series include *i*) discrete wavelet transform (DWT) based denoising and subsequent conventional BVAR forecasting, *ii*) Haar Maximal Overlap Discrete Wavelet Transform (MODWT) separate scale BVAR forecasting and subsequent aggregation of the separate scale forecasts, and *iii*) extension of the Multiscale Autoregressive model of Renaud et al. (2003) into the Multiscale BVAR both in a homoscedastic and a heteroscedastic specification employing a Minnesota inspired prior which allows a varying degree of shrinkage across scales, as well as the stochastic search variable selection prior.

In an out-of-sample forecasting exercise using US macroeconomic variables, the three forecasting approaches are found to outperform a conventional BVAR used as a benchmark in terms of point and density accuracy on many occasions. Wavelet-based denoising forecasting presents a robust superior performance for density forecasts, and the multiscale BVAR forecasting approach outperforms the benchmark across all variables for medium to long-term forecasts.

I leave for further research the investigation of more advanced wavelet-based denoising techniques incorporating prior beliefs in a Bayesian framework and a more thorough examination of the role of the hyperparameters optimisation on the MODWT separate scale BVAR modelling. Finally, for the case of the multiscale BVAR going beyond optimising the proposed prior's hyperparameter values, alternative priors like the global-local family prior could be employed to verify the extent to which the a priori additional shrinkage imposed on the wavelet high-frequency components is supported from the underlying data. Finally, the multiscale BVAR behaviour of impulse response functions and forecast error value decomposition can be tested against a conventional BVAR.

Regarding modelling mixed-frequency time series using a wavelet-based approach, the proposed wavelet-based MF-VAR model, which comprises separate scale MF-VARs in a single system, exhibits increased in-sample forecast accuracy for known monthly time series in a statistical sense compared to the standard MF-VAR; however, this behaviour reverses during recessionary periods like the latest COVID-19 recession. Further research, like examining the out-of-sample forecasting performance in a nowcasting recursive estimation exercise, could shed more light on the potential merits of the proposed wavelet-based MF-VAR.

References

Abramovich, F., Sapatinas, T., & Silverman, B. W. (1998). Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 60*(4), 725-749.

Adolfson, M., Laséen, S., Lindé, J., & Villani, M. (2005). Are constant interest rate forecasts modest policy interventions? Evidence from a dynamic open-economy model. *International Finance, 8*(3), 509-544.

Aguiar-Conraria, L., & Soares, M. J. (2011). Business cycle synchronization and the Euro: A wavelet analysis. *Journal of Macroeconomics, 33*(3), 477-489.

Aguiar-Conraria, L., & Soares, M. J. (2014). The continuous wavelet transform: Moving beyond uni-and bivariate analysis. *Journal of Economic Surveys, 28*(2), 344-375.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723.

Alrumaih, R. M., & Al-Fawzan, M. A. (2002). Time series forecasting using wavelet denoising an application to Saudi stock index. *Journal of King Saud University-Engineering Sciences, 14*(2), 221-233.

Altavilla, C., Giacomini, R., & Ragusa, G. (2017). Anchoring the yield curve using survey expectations. *Journal of Applied Econometrics, 32*(6), 1055-1068.

Aminghafari, M., & Poggi, J.-M. (2007). Forecasting time series using wavelets. *International Journal of Wavelets, Multiresolution and Information Processing, 5*(05), 709-724.

Amisano, G., & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics, 25*(2), 177-190.

Andel, J. (1976). Autoregressive series with random parameters. *Mathematische Operationsforschung und Statistik, 7*(5), 735-741.

Anderson, H. M., & Low, C. N. (2006). Random walk smooth transition autoregressive models. *Contributions to Economic Analysis, 276*, 247-281.

Andersson, M. K., Palmqvist, S., & Waggoner, D. F. (2010). Density-conditional forecasts in dynamic multivariate models. *Sveriges Riksbank, Working Paper Series, 2010-247*.

Andrieș, A. M., Ihnatov, I., & Tiwari, A. K. (2014). Analyzing time-frequency relationship between interest rate, stock price and exchange rate through continuous wavelet. *Economic Modelling, 41*, 227-238.

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72*(3), 269-342.

Angelini, E., Lalik, M., Lenza, M., & Paredes, J. (2019). Mind the gap: A multi-country BVAR benchmark for the Eurosystem projections. *International Journal of Forecasting, 35*(4), 1658-1668.

Ankargren, S., Unosson, M., & Yang, Y. (2018). A mixed-frequency Bayesian vector autoregression with a steady-state prior. *Uppsala University, Department of Statistics, Working Paper Series, 2018-3*.

Ankargren, S., Unosson, M., & Yang, Y. (2020). A flexible mixed-frequency vector autoregression with a steady-state prior. *Journal of Time Series Econometrics, 12*(2).

Antolin-Diaz, J., Drechsel, T., & Petrella, I. (2017). Tracking the slowdown in long-run GDP growth. *Review of Economics and Statistics, 99*(2), 343-356.

Antolin-Diaz, J., Petrella, I., & Rubio Ramírez, J. (2021). Structural scenario analysis with SVARs. *Journal of Monetary Economics, 117*, 798-815.

Antolin-Diaz, J., & Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for SVARs. *American Economic Review, 108*(10), 2802-2829.

Arias, J. E., Rubio-Ramírez, J. F., & Waggoner, D. F. (2018). Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications. *Econometrica, 86*(2), 685-720.

Ascari, G., & Sbordone, A. M. (2014). The macroeconomics of trend inflation. *Journal of Economic Literature, 52*(3), 679-739.

Assenmacher-Wesche, K., & Gerlach, S. (2007). Money at low frequencies. *Journal of the European Economic Association, 5*(2-3), 534-542.
Assenmacher-Wesche, K., & Gerlach, S. (2008). Interpreting euro area inflation at high and low frequencies. *European Economic Review, 52*(6), 964-986.
Auerbach, A. J., & Gorodnichenko, Y. (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy, 4*(2), 1-27.
Bańbura, M., Giannone, D., & Lenza, M. (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting, 31*(3), 739-756.
Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics, 25*(1), 71-92.
Bańbura, M., & van Vlodrop, A. (2018). Forecasting with Bayesian vector autoregressions with time variation in the mean. *Tinbergen Institute, Discussion Paper Series, 2018-025/IV*.
Barnett, A., Mumtaz, H., & Theodoridis, K. (2014). Forecasting UK GDP growth and inflation under structural change. A comparison of models with time-varying parameters. *International Journal of Forecasting, 30*(1), 129-143.
Baumeister, C., & Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica, 83*(5), 1963-1999.
Baxter, M., & King, R. G. (1999). Measuring business cycles: approximate band-pass filters for economic time series. *The Review of Economic Studies, 81*(4), 575-593.
Beaudry, P., Galizia, D., & Portier, F. (2020). Putting the cycle back into business cycle analysis. *American Economic Review, 110*(1), 1-47.
Beechey, M., & Österholm, P. (2010). Forecasting inflation in an inflation-targeting regime: A role for informative steady-state priors. *International Journal of Forecasting, 26*(2), 248-264.
Bekiros, S. (2014). Forecasting with a state space time-varying parameter VAR model: Evidence from the euro area. *Economic Modelling, 38*, 619-626.
Bekiros, S., & Marcellino, M. (2013). The multiscale causal dynamics of foreign exchange markets. *Journal of International Money and Finance, 33*, 282-305.
Benaouda, D., Murtagh, F., Starck, J.-L., & Renaud, O. (2006). Wavelet-based non-linear multiscale decomposition model for electricity load forecasting. *Neurocomputing, 70*(1-3), 139-154.
Bhattacharya, A., Chakraborty, A., & Mallick, B. K. (2016). Fast sampling with Gaussian scale mixture priors in high-dimensional regression. *Biometrika, 103*(4), 985-991.
Blake, A. P., & Mumtaz, H. (2017). *Applied Bayesian econometrics for central bankers*: Bank of England, Centre for Central Banking Studies
Bruzda, J. (2020). The wavelet scaling approach to forecasting: Verification on a large set of Noisy data. *Journal of Forecasting, 39*(3), 353-367.
Caggiano, G., Castelnuovo, E., & Groshenny, N. (2014). Uncertainty shocks and unemployment dynamics in US recessions. *Journal of Monetary Economics, 67*, 78-92.
Campbell, E. P. (2004). Bayesian selection of threshold autoregressive models. *Journal of Time Series Analysis, 25*(4), 467-482.
Caraiani, P. (2015). Estimating DSGE models across time and frequency. *Journal of Macroeconomics, 44*, 33-49.
Carriero, A., Chan, J., Clark, T. E., & Marcellino, M. (2022). Corrigendum to "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors"[J. Econometrics 212 (1)(2019) 137-154]. *Journal of Econometrics, 227*(2), 506-512.
Carriero, A., Clark, T. E., & Marcellino, M. (2015a). Bayesian VARs: Specification choices and forecast accuracy. *Journal of Applied Econometrics, 30*(1), 46-73.

Carriero, A., Clark, T. E., & Marcellino, M. (2015b). Realtime nowcasting with a Bayesian mixed-frequency model with stochastic volatility. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 178*(4), 837.

Carriero, A., Clark, T. E., & Marcellino, M. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*.

Carter, C. K., & Kohn, R. (1994). On Gibbs sampling for state-space models. *Biometrika, 81*(3), 541-553.

Cascio, I. L. (2007). Wavelet analysis and denoising: New tools for economists. *Queen Mary University of London, Department of Economics, Working Paper Series, 2007-600*.

Cekic, S., Grandjean, D., & Renaud, O. (2019). Multiscale Bayesian state-space model for Granger causality analysis of brain signal. *Journal of Applied Statistics, 46*(1), 66-84.

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis, 1*(4), 651-673.

Champagne, J., Poulin-Bellisle, G., & Sekkel, R. (2020). Introducing the Bank of Canada staff economic projections database. *Journal of Applied Econometrics, 35*(1), 114-129.

Chan, J. C. (2018). Specification tests for time-varying parameter models with stochastic volatility. *Econometric Reviews, 37*(8), 807-823.

Chan, J. C., & Grant, A. L. (2015). Pitfalls of estimating the marginal likelihood using the modified harmonic mean. *Economics Letters, 131*, 29-33.

Chan, J. C., & Grant, A. L. (2016). Fast computation of the deviance information criterion for latent variable models. *Computational Statistics and Data Analysis, 100*, 847-859.

Chan, J. C., Jacobi, L., & Zhu, D. (2020). Efficient selection of hyperparameters in large Bayesian VARs using automatic differentiation. *Journal of Forecasting, 39*(6), 934-943.

Chan, J. C., & Jeliazkov, I. (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, C. W. (1998). A Bayesian analysis of generalized threshold autoregressive models. *Statistics and Probability Letters, 40*(1), 15-22.

Chen, C. W., & Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis, 16*(5), 483-492.

Chen, D., & Bunn, D. W. (2010). Analysis of the non-linear response of electricity prices to fundamental and strategic factors. *IEEE Transactions on Power Systems, 25*(2), 595-606.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association, 90*(432), 1313-1321.

Chipman, H. A., Kolaczyk, E. D., & McCulloch, R. E. (1997). Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association, 92*(440), 1413-1421.

Chiu, C.-W. J., Mumtaz, H., & Pinter, G. (2017). Forecasting with VAR models: Fat tails and stochastic volatility. *International Journal of Forecasting, 33*(4), 1124-1143.

Chopin, N., & Papaspiliopoulos, O. (2020). *An introduction to sequential Monte Carlo*: Springer.

Chow, G. C., & Lin, A. L. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 372-375.

Christiano, L. J., & Fitzgerald, T. J. (2003). The band pass filter. *International Economic Review, 44*(2), 435-465.

Cimadomo, J., Giannone, D., Lenza, M., Monti, F., & Sokol, A. (2022). Nowcasting with large Bayesian vector autoregressions. *Journal of Econometrics, 231*(2), 500-519.

Clarida, R. H., & Coyle, D. (1984). Conditional projection by means of Kalman filtering. *National Bureau of Economic Research, Technical Working Paper Series, 1984-36*.

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics, 29*(3), 327-341.

Clark, T. E., & McCracken, M. W. (2017). Tests of predictive ability for vector autoregressions used for conditional forecasting. *Journal of Applied Econometrics, 32*(3), 533-553.
Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics, 30*(4), 551-575.
Clements, M. P., & Krolzig, H.-M. (2003). Business cycle asymmetries: Characterization and testing based on Markov-switching autoregressions. *Journal of Business and Economic Statistics, 21*(1), 196-211.
Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics, 8*(2), 262-302.
Coifman, R. R., & Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory, 38*(2), 713-718.
Conejo, A. J., Plazas, M. A., Espinola, R., & Molina, A. B. (2005). Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Transactions on Power Systems, 20*(2), 1035-1042.
Crouse, M. S., Nowak, R. D., & Baraniuk, R. G. (1998). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing, 46*(4), 886-902.
Crowley, P. M. (2007). A guide to wavelets for economists. *Journal of Economic Surveys, 21*(2), 207-267.
D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics, 28*(1), 82-101.
de Jong, P. (1988). A cross-validation filter for time series models. *Biometrika, 75*(3), 594-600.
de Jong, P., & Shephard, N. (1995). The simulation smoother for time series models. *Biometrika, 82*(2), 339-350.
Del Negro, M., & Schorfheide, F. (2012). DSGE model-based forecasting, in the Handbook of Economic Forecasting Vol. II, edited by G. Elliott and A. Timmerman. In: Elsevier.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics, 13*(3), 253-263.
Dieppe, A., Legrand, R., & Van Roye, B. (2016). The BEAR toolbox. *ECB, Working Paper Series, 2016-1934*.
Doan, T., Litterman, R., & Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews, 3*(1), 1-100.
Donoho, D. L., & Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association, 90*(432), 1200-1224.
Donoho, D. L., & Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika, 81*(3), 425-455.
Dueker, M., Owyang, M., & Sola, M. (2010). A time-varying threshold STAR model of unemployment and the natural rate. *Federal Reserve Bank of St. Louis, Working Paper Series, 2010-29*.
Durbin, J., & Koopman, S. J. (2002). A simple and efficient simulation smoother for state-space time series analysis. *Biometrika, 89*(3), 603-616.
Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*: Oxford University Press.
Dutilleux, P. (1989). An implementation of the 'algorithme à trous' to compute the wavelet transform. In *Wavelets* (pp. 298-304): Springer.
ECB. (2016). A guide to the Eurosystem/ECB staff macroeconomic projection exercises. *ECB publications, July 2016*.
El-Shagi, M., Giesen, S., & Jung, A. (2016). Revisiting the relative forecast performances of Fed staff and private forecasters: A dynamic approach. *International Journal of Forecasting, 32*(2), 313-323.

Faria, Á. E., & Santos, A. J. (2018). Dynamic Bayesian smooth transition autoregressive models applied to hourly electricity load in Brazil. In O. Valenzuela, F. Rojas, H. Pomares, & I. Rojas (Eds.), *Theory and applications of time series analysis. Selected Contributions from ITISE 2018*: Springer.

Faust, J., & Wright, J. H. (2008). Efficient forecast tests for conditional policy forecasts. *Journal of Econometrics, 146*(2), 293-303.

Ferbar, L., Črešlovnik, D., Mojškerc, B., & Rajgelj, M. (2009). Demand forecasting methods in a supply chain: Smoothing and denoising. *International Journal of Production Economics, 118*(1), 49-54.

Fernández-Villaverde, J., & Rubio-Ramírez, J. F. (2005). Estimating dynamic equilibrium economies: linear versus non-linear likelihood. *Journal of Applied Econometrics, 20*(7), 891-910.

Ferrara, L., Marcellino, M., & Mogliani, M. (2015). Macroeconomic forecasting during the Great Recession: The return of non-linearity? *International Journal of Forecasting, 31*(3), 664-679.

Frühwirth-Schnatter, S., & Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state-space models. *Journal of Econometrics, 154*(1), 85-100.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis, 15*(2), 183-202.

Fryzlewicz, P., Van Bellegem, S., & Von Sachs, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics, 55*(4), 737-764.

Fuentes-Albero, C., & Melosi, L. (2013). Methods for computing marginal data densities from the Gibbs output. *Journal of Econometrics, 175*(2), 132-141.

Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering, 93*(26), 429-441.

Galí, J. (2011). Are central banks' projections meaningful? *Journal of Monetary Economics, 58*(6-8), 537-550.

Gallegati, M., Giri, F., & Palestrini, A. (2019). DSGE model with financial frictions over subsets of business cycle frequencies. *Journal of Economic Dynamics and Control, 100*, 152-163.

Gallegati, M., & Semmler, W. (2014). *Wavelet applications in economics and finance*: Springer.

Galvão, A. B. (2006). Structural break threshold VARs for predicting US recessions using the spread. *Journal of Applied Econometrics, 21*(4), 463-487.

Galvão, A. B., & Owyang, M. T. (2018). Financial stress regimes and the macroeconomy. *Journal of Money, Credit and Banking, 50*(7), 1479-1505.

Gefang, D., & Strachan, R. (2010). Non-linear impacts of international business cycles on the UK: A Bayesian smooth transition VAR approach. *Studies in Nonlinear Dynamics and Econometrics, 14*(1).

Gelfand, A. E., & Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 56*(3), 501-514.

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association, 88*(423), 881-889.

George, E. I., Sun, D., & Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics, 142*(1), 553-580.

Gerlach, R., & Chen, C. W. (2008). Bayesian inference and model comparison for asymmetric smooth transition heteroskedastic models. *Statistics and Computing, 18*(4), 391.

Geweke, J. (1999). Using simulation methods for Bayesian econometric models: Inference, development, and communication. *Econometric Reviews, 18*(1), 1-73.

Geweke, J., Koop, G., & van Dijk, H. (2011). *The Oxford handbook of Bayesian econometrics*: Oxford University Press.
Geweke, J., & Tanizaki, H. (2001). Bayesian estimation of state-space models using the Metropolis-Hastings algorithm within Gibbs sampling. *Computational Statistics and Data Analysis, 37*(2), 151-170.
Geweke, J., & Terui, N. (1993). Bayesian threshold autoregressive models for non-linear time series. *Journal of Time Series Analysis, 14*(5), 441-454.
Ghysels, E. (2016). Macroeconomics and the reality of mixed-frequency data. *Journal of Econometrics, 193*(2), 294-314.
Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics, 25*(4), 595-620.
Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *The Review of Economics and Statistics, 97*(2), 436-451.
Giannone, D., Lenza, M., & Primiceri, G. E. (2019). Priors for the long run. *Journal of the American Statistical Association, 114*(526), 565-580.
Giannone, D., Lenza, M., & Reichlin, L. (2019). Money, credit, monetary policy and the business cycle in the euro area: What has changed since the crisis? *International Journal of Central Banking, 15*(5), 137-173.
Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association, 102*(477), 359-378.
Gordon, N. J., Salmond, D. J., & Smith, A. F. (1993). *Novel approach to non-linear/non-Gaussian Bayesian state estimation.* Paper presented at the IEE proceedings F (radar and signal processing).
Götz, T. B., & Hauzenberger, K. (2021). Large mixed-frequency VARs with a parsimonious time-varying parameter structure. *The Econometrics Journal, 24*(3), 442-461.
Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika, 82*(4), 711-732.
Gülerce, M., & Ünal, G. (2016). Using wavelet analysis to uncover the co-movement behaviour of multiple energy commodity prices. *International Journal of Wavelets, Multiresolution and Information Processing, 14*(06), 1650047.
Gupta, S., Das, D., Hasim, H., & Tiwari, A. K. (2018). The dynamic relationship between stock returns and trading volume revisited: A MODWT-VAR approach. *Finance Research Letters, 27*, 91-98.
Gustafsson, O., Villani, M., & Stockhammar, P. (2020). Bayesian optimization of hyperparameters when the marginal likelihood is estimated by MCMC. *arXiv preprint arXiv:2004.10092*.
Haar, A. (1910). On the theory of orthogonal function systems. *Mathematische Annalen, 69*(3), 331-371.
Haggan, V., & Ozaki, T. (1981). Modelling non-linear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika, 68*(1), 189-196.
Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*, 357-384.
Hamilton, J. D. (1994). *Time series analysis*: Princeton University Press.
Hamilton, J. D., Harris, E. S., Hatzius, J., & West, K. D. (2016). The equilibrium real funds rate: Past, present, and future. *IMF Economic Review, 64*(4), 660-707.
Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*: Cambridge University Press.
Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting, 13*(2), 281-291.
Herbst, E. P., & Schorfheide, F. (2015). *Bayesian estimation of DSGE models*: Princeton University Press.
Herwartz, H., & Schlüter, S. (2017). On the predictive information of futures' prices: A wavelet-based assessment. *Journal of Forecasting, 36*(4), 345-356.

Hkiri, B., Hammoudeh, S., Aloui, C., & Shahbaz, M. (2018). The interconnections between US financial CDS spreads and control variables: New evidence using partial and multivariate wavelet coherences. *International Review of Economics and Finance, 57*, 237-257.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar US business cycles: an empirical investigation. *Journal of Money, Credit and Banking*, 1-16.

Holschneider, M., Kronland-Martinet, R., Morlet, J., & Tchamitchian, P. (1989). A real-time algorithm for signal analysis with the help of the wavelet transform. In *Wavelets* (pp. 286-297): Springer.

Holston, K., Laubach, T., & Williams, J. C. (2017). Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics, 108*, S59-S75.

Huber, F., Koop, G., Onorante, L., Pfarrhofer, M., & Schreiner, J. (2020). Nowcasting in a pandemic using non-parametric mixed-frequency VARs. *Journal of Econometrics*.

Iseringhausen, M., & Vierke, H. (2019). What drives output volatility? The role of demographics and government size revisited. *Oxford Bulletin of Economics and Statistics, 81*(4), 849-867.

Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics, 12*(4), 371-389.

Jarociński, M. (2010). Conditional forecasts and uncertainty about forecast revisions in vector autoregressions. *Economics Letters, 108*(3), 257-259.

Jarociński, M. (2015). A note on implementing the Durbin and Koopman simulation smoother. *Computational Statistics and Data Analysis, 91*, 1-3.

Jarociński, M., & Smets, F. (2008). House prices and the stance of monetary policy.

Jorgenson, D. W., Ho, M. S., & Stiroh, K. J. (2008). A retrospective look at the US productivity growth resurgence. *Journal of Economic Perspectives, 22*(1), 3-24.

Julier, S. J., & Uhlmann, J. K. (2004). Unscented filtering and non-linear estimation. *Proceedings of the IEEE, 92*(3), 401-422.

Julier, S. J., Uhlmann, J. K., & Durrant-Whyte, H. F. (1995). *A new approach for filtering non-linear systems.* Paper presented at the Proceedings of 1995 American Control Conference-ACC'95.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

Kennedy, J. (1997). *The particle swarm: Social adaptation of knowledge.* Paper presented at the Proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC'97).

Kennedy, J., & Eberhart, R. (1995). *Particle swarm optimization.* Paper presented at the Proceedings of ICNN'95-International Conference on Neural Networks.

Khalfaoui, R. (2018). Oil-gold time varying nexus: A time-frequency analysis. *Physica A: Statistical Mechanics and its Applications, 503*, 86-104.

Khan, M. Y., & Mittnik, S. (2018). Non-linear time series modelling and forecasting the seismic data of the Hindu Kush region. *Journal of Seismology, 22*(1), 353-376.

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies, 65*(3), 361-393.

Knotek II, E. S., & Zaman, S. (2019). Financial nowcasts and their usefulness in macroeconomic forecasting. *International Journal of Forecasting, 35*(4), 1708-1724.

Kontogeorgos, G., & Lambrias, K. (2022). Evaluating the Eurosystem/ECB staff macroeconomic projections: The first 20 years. *Journal of Forecasting, 41*(2), 213-229.

Koop, G. (2003). *Bayesian econometrics*: Wiley.

Koop, G. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics, 28*(2), 177-203.

Koop, G., & Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics, 177*(2), 185-198.
Koop, G., McIntyre, S., & Mitchell, J. (2020). UK regional nowcasting using a mixed-frequency vector autoregressive model with entropic tilting. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 183*(1), 91-119.
Koop, G., McIntyre, S., Mitchell, J., & Poon, A. (2020). Regional output growth in the United Kingdom: More timely and higher frequency estimates from 1970. *Journal of Applied Econometrics, 35*(2), 176-197.
Koop, G., & Potter, S. M. (2003). Bayesian analysis of endogenous delay threshold models. *Journal of Business and Economic Statistics, 21*(1), 93-103.
Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association, 92*(440), 1630-1638.
Kriechbaumer, T., Angus, A., Parsons, D., & Casado, M. R. (2014). An improved wavelet-ARIMA approach for forecasting metal prices. *Resources Policy, 39*, 32-41.
Krüger, F., Clark, T. E., & Ravazzolo, F. (2017). Using entropic tilting to combine BVAR forecasts with external nowcasts. *Journal of Business and Economic Statistics, 35*(3), 470-485.
Li, X., He, K., Lai, K. K., & Zou, Y. (2014). Forecasting crude oil price with multiscale denoising ensemble model. *Mathematical Problems in Engineering, 2014*.
Lim, K., & Tong, H. (1980). Threshold autoregressions, limit cycles, and data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 42*, 245-292.
Lindsten, F., Jordan, M. I., & Schon, T. B. (2014). Particle Gibbs with ancestor sampling. *Journal of Machine Learning Research, 15*, 2145-2184.
Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions: Five years of experience. *Journal of Business and Economic Statistics, 4*(1), 25-38.
Livingston, G., & Nur, D. (2020). Bayesian inference of smooth transition autoregressive (STAR)(k)-GARCH (l,m) models. *Statistical Papers, 61*(6), 2449-2482.
Livingston Jr, G., & Nur, D. (2019). Bayesian estimation and model selection of a multivariate smooth transition autoregressive model. *Environmetrics*, e2615.
Long, H., & Yan, G. (2016). Forecasting import and export volume with a combined model based on wavelet filtering. *International Journal of Wavelets, Multiresolution and Information Processing, 14*(03), 1650018.
Lopes, H. F., & Salazar, E. (2006a). Bayesian model uncertainty in smooth transition autoregressions. *Journal of Time Series Analysis, 27*(1), 99-117.
Lopes, H. F., & Salazar, E. (2006b). Time series mean level and stochastic volatility modelling by smooth transition autoregressions: A Bayesian approach. In *Econometric Analysis of Financial and Economic Time Series* (Vol. 20, pp. 225-238): Emerald Publishing.
Lotrič, U. (2004). Wavelet-based denoising integrated into multilayered perceptron. *Neurocomputing, 62*, 179-196.
Lotrič, U., & Dobnikar, A. (2005). Neural networks with wavelet-based denoising layer for time series prediction. *Neural Computing and Applications, 14*(1), 11-17.
Louzis, D. P. (2019). Steady-state modelling and macroeconomic forecasting quality. *Journal of Applied Econometrics, 34*(2), 285-314.
Lubrano, M. (2000). *Bayesian analysis of non-linear time series models with a threshold.* Paper presented at the Non-linear Econometric Modelling in Time Series: Proceedings of the Eleventh International Symposium in Economic Theory.
Lundbergh, S., Teräsvirta, T., & Van Dijk, D. (2003). Time-varying smooth transition autoregressive models. *Journal of Business and Economic Statistics, 21*(1), 104-121.
Lütkepohl, H. (2005). *New introduction to multiple time series analysis*: Springer.
Maih, J. (2010). Conditional forecasts in DSGE models. *Norges Bank, Working Paper Series, 2010-7*.
Mallat, S. (1999). *A wavelet tour of signal processing*: Elsevier.

Marcellino, M. (2002). Instability and non-linearity in the EMU. *CEPR, Discussion Paper Series, 2002-3312*.
Marcellino, M. (2004). Forecasting EMU macroeconomic variables. *International Journal of Forecasting, 20*(2), 359-372.
Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics, 135*(1-2), 499-526.
Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics, 18*(4), 427-443.
Mariano, R. S., & Murasawa, Y. (2010). A coincident index, common factors, and monthly real GDP. *Oxford Bulletin of Economics and Statistics, 72*(1), 27-46.
Martins, M. M., & Verona, F. (2020). Forecasting inflation with the New Keynesian Phillips curve: Frequency matters. *Bank of Finland, Research Discussion Paper Series, 2020-4*.
Matthes, C., Lubik, T., & Verona, F. (2019). *Assessing US Aggregate Fluctuations Across Time and Frequencies*. Retrieved from
McCracken, M. W., & McGillicuddy, J. T. (2019). An empirical investigation of direct and iterated multistep conditional forecasts. *Journal of Applied Econometrics, 34*(2), 181-204.
McCracken, M. W., Owyang, M., & Sekhposyan, T. (2021). Real-time forecasting and scenario analysis using a large mixed-frequency Bayesian VAR. *International Journal of Central Banking, 17*(5), 327-367.
Mertins, A. (1999). *Signal analysis: wavelets, filter banks, time-frequency transforms and applications*: Wiley.
Morley, J., & Piger, J. (2012). The asymmetric business cycle. *The Review of Economics and Statistics, 94*(1), 208-221.
Mumtaz, H. (2018). A generalised stochastic volatility in mean VAR. *Economics Letters, 173*, 10-14.
Murtagh, F., Starck, J.-L., & Renaud, O. (2004). On neuro-wavelet modelling. *Decision Support Systems, 37*(4), 475-484.
Nicholls, D. F., & Quinn, B. G. (1980). The estimation of random coefficient autoregressive models. *Journal of Time Series Analysis, 1*(1), 37-46.
Nonejad, N. (2016). Replicating the Results in 'A New Model of Trend Inflation'Using Particle Markov Chain Monte Carlo. *Journal of Applied Econometrics, 31*(7), 1478-1483.
Omay, T., Shahbaz, M., & Hasanov, M. (2020). Testing PPP hypothesis under temporary structural breaks and asymmetric dynamic adjustments. *Applied Economics*, 1-19.
Percival, D. B., & Walden, A. T. (2000). *Wavelet methods for time series analysis*: Cambridge University Press.
Primiceri, G. E. (2005). Time-varying structural vector autoregressions and monetary policy. *The Review of Economic Studies, 72*(3), 821-852.
Qu, H., Chen, W., Niu, M., & Li, X. (2016). Forecasting realized volatility in electricity markets using logistic smooth transition heterogeneous autoregressive models. *Energy Economics, 54*, 68-76.
Ramsey, J. B. (2002). Wavelets in economics and finance: Past and future. *Studies in Nonlinear Dynamics and Econometrics, 6*(3), 1-29.
Ramsey, J. B., & Lampart, C. (1998a). The decomposition of economic relationships by time scale using wavelets: expenditure and income. *Studies in Nonlinear Dynamics and Econometrics, 3*(1), 23-42.
Ramsey, J. B., & Lampart, C. (1998b). Decomposition of economic relationships by timescale using wavelets. *Macroeconomic Dynamics, 2*(1), 49-71.

Renaud, O., Starck, J.-L., & Murtagh, F. (2003). Prediction based on a multiscale decomposition. *International Journal of Wavelets, Multiresolution and Information Processing, 1*(02), 217-232.
Robertson, J. C., Tallman, E. W., & Whiteman, C. H. (2005). Forecasting using relative entropy. *Journal of Money, Credit and Banking, 37*(3), 383-401.
Rua, A. (2010). Measuring comovement in the time-frequency space. *Journal of Macroeconomics, 32*(2), 685-691.
Rua, A. (2011). A wavelet approach for factor-augmented forecasting. *Journal of Forecasting, 30*(7), 666-678.
Rua, A. (2017). A wavelet-based multivariate multiscale approach for forecasting. *International Journal of Forecasting, 33*(3), 581-590.
Rubio-Ramírez, J. F., Waggoner, D. F., & Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies, 77*(2), 665-696.
Saâdaoui, F., & Rabbouch, H. (2014). A wavelet-based multiscale vector-ANN model to predict comovement of econophysical systems. *Expert Systems with Applications, 41*(13), 6017-6028.
Sala, L. (2015). DSGE models in the frequency domains. *Journal of Applied Econometrics, 30*(2), 219-240.
Särkkä, S. (2013). *Bayesian filtering and smoothing*: Cambridge University Press.
Sato, J. R., Morettin, P. A., Arantes, P. R., & Amaro Jr, E. (2007). Wavelet based time-varying vector autoregressive modelling. *Computational Statistics and Data Analysis, 51*(12), 5847-5866.
Schlueter, S., & Deuschle, C. (2014). Wavelet-based forecasting of ARIMA time series-an empirical comparison of different methods. *Managerial Economics, 15*(1), 107.
Schorfheide, F., & Song, D. (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business and Economic Statistics, 33*(3), 366-380.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.
Shensa, M. J. (1992). The discrete wavelet transform: wedding the a trous and Mallat algorithms. *IEEE Transactions on Signal Processing, 40*(10), 2464-2482.
Shi, Y., & Eberhart, R. (1998). *A modified particle swarm optimizer.* Paper presented at the 1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360).
Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 1-48.
Soltani, S., Boichu, D., Simard, P., & Canu, S. (2000). The long-term memory prediction by multiscale decomposition. *Signal Processing, 80*(10), 2195-2205.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583-639.
Srivastava, M., Anderson, C. L., & Freed, J. H. (2016). A new wavelet denoising method for selecting decomposition levels and noise thresholds. *IEEE Access, 4*, 3862-3877.
Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 1135-1151.
Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives, 15*(4), 101-115.
Stock, J. H., & Watson, M. W. (2012). *Disentangling the Channels of the 2007-2009 Recession*. Retrieved from
Strohsal, T., Proaño, C. R., & Wolters, J. (2019). Characterizing the financial cycle: Evidence from a frequency domain analysis. *Journal of Banking and Finance, 106*, 568-591.
Sundararajan, D. (2016). *Discrete wavelet transform: A signal processing approach*: Wiley.

Tallman, E. W., & Zaman, S. (2019). Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy. *International Journal of Forecasting, 36*(2), 373-398.

Tan, Z., Zhang, J., Wang, J., & Xu, J. (2010). Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Applied Energy, 87*(11), 3606-3610.

Telatar, E., & Hasanov, M. (2009). Purchasing power parity in Central and East European countries. *Eastern European Economics, 47*(5), 25-41.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association, 89*(425), 208-218.

Teräsvirta, T. (2006). Forecasting economic variables with non-linear models. *Handbook of economic forecasting, 1*, 413-457.

Tiwari, A. K. (2013). Oil prices and the macroeconomy reconsideration for Germany: Using continuous wavelet. *Economic Modelling, 30*, 636-642.

Tong, H. (1990). *Non-linear time series: A dynamical system approach*: Oxford University Press.

Tong, H., Thanoon, B., & Gudmundsson, G. (1985). Threshold time series modelling of two Icelandic river flow systems *Journal of the American Water Resources Association, 21*(4), 651-662.

Trezzi, R. (2013). A wavelet analysis of international risk-sharing. *Economics Letters, 118*(2), 330-333.

Tsay, R. S., & Chen, R. (2018). *Non-linear time series analysis*: Wiley.

Tulip, P. (2009). Has the economy become more predictable? Changes in Greenbook forecast accuracy. *Journal of Money, Credit and Banking, 41*(6), 1217-1231.

Uddin, G. S., Bekiros, S., & Ahmed, A. (2018). The nexus between geopolitical uncertainty and crude oil markets: An entropy-based wavelet analysis. *Physica A: Statistical Mechanics and its Applications, 495*, 30-39.

Uddin, G. S., Gençay, R., Bekiros, S., & Sahamkhadam, M. (2019). Enhancing the predictability of crude oil markets with hybrid wavelet approaches. *Economics Letters, 182*, 50-54.

Unser, M. (2000). Sampling-50 years after Shannon. *Proceedings of the IEEE, 88*(4), 569-587.

Verona, F. (2016). Time-frequency characterization of the US financial cycle. *Economics Letters, 144*, 75-79.

Vidakovic, B. (1998). Non-linear wavelet shrinkage with Bayes rules and Bayes factors. *Journal of the American Statistical Association, 93*(441), 173-179.

Villani, M. (2009). Steady-state priors for vector autoregressions. *Journal of Applied Econometrics, 24*(4), 630-650.

Waggoner, D. F., & Zha, T. (1999). Conditional forecasts in dynamic multivariate models. *The Review of Economics and Statistics, 81*(4), 639-651.

Wan, E. A., & Van Der Merwe, R. (2001). The unscented Kalman filter. In S. Haykin (Ed.), *Kalman filtering and neural networks* (pp. 221-280).

Wieland, V., & Wolters, M. (2013). Forecasting and policy making. In *Handbook of economic forecasting* (Vol. 2, pp. 239-325): Elsevier.

Wong, H., Ip, W.-C., Xie, Z., & Lui, X. (2003). Modelling and forecasting by wavelets, and the application to exchange rates. *Journal of Applied Statistics, 30*(5), 537-553.

Wright, J. H. (2013). Evaluating real-time VAR forecasts with an informative democratic prior. *Journal of Applied Econometrics, 28*(5), 762-776.

Xu, H., & Niimura, T. (2004). *Short-term electricity price modelling and forecasting using wavelets and multivariate time series.* Paper presented at the IEEE PES Power Systems Conference and Exposition, 2004.

Xu, M., Han, M., & Lin, H. (2018). Wavelet-denoising multiple echo state networks for multivariate time series prediction. *Information Sciences, 465*, 439-458.
Yang, L., Lee, C., & Chen, I. P. (2021). Threshold model with a time-varying threshold based on Fourier approximation. *Journal of Time Series Analysis, 42*(4), 406-430.
Yogo, M. (2008). Measuring business cycles: A wavelet analysis of economic time series. *Economics Letters, 100*(2), 208-212.
Zhang, K., Gençay, R., & Yazgan, M. E. (2017). Application of wavelet decomposition in time-series forecasting. *Economics Letters, 158*, 41-46.
Zhu, L., Wang, Y., & Fan, Q. (2014). MODWT-ARMA model for time series prediction. *Applied Mathematical Modelling, 38*(5-6), 1859-1865.
Zhu, Y., & Chen, H. (2017). The asymmetry of US monetary policy: Evidence from a threshold Taylor rule with time-varying threshold values. *Physica A: Statistical Mechanics and its Applications, 473*, 522-535.
Zhu, Y., Chen, H., & Lin, M. (2019). Threshold models with time-varying threshold values and their application in estimating regime-sensitive Taylor rules. *Studies in Nonlinear Dynamics and Econometrics, 23*(5).

Appendix

Appendix A (Essay 1)

A.1 Technical appendix

A.1.1 Particle swarm optimisation

Particle swarm optimisation is a derivative-free, stochastic optimisation technique which iteratively tries to improve a candidate solution for the optimisation of an objective function. PSO (Kennedy & Eberhart, 1995; Kennedy, 1997; Shi & Eberhart, 1998) mimics social behaviour in order to reach an optimal solution for a problem. The rationale of the algorithm can be briefly described as follows. A swarm of particles explores the *K*-dimensional space defined for the objective function that needs to be optimised in a stochastic way which is a mixture of the swarm's best solution and each particle's best own solution. The particles are initialised uniformly in the *K*-dimensional space, and after each iteration, they converge slowly to the global optimum by updating the swarm's best position and their own best position every time a better candidate solution is found. The algorithm stops when the particles cannot reach a better candidate solution within a tolerance limit set ex-ante. Obviously, PSO can be easily parallelised since, at each iteration, each particle's new position requires an evaluation of the objective function, which in this case is the time-expensive approximation of the marginal likelihood following Chib (1995) can be estimated independently.

More formally, the swam is comprised of $n = 1,\cdots, N$ particles. Each particle at iteration t has a position defined as x_t^n where the dimension of the vector x_t^n is equal to K for the K-dimensional space defining the objective function f. At iteration i particle's n own best candidate solution as estimated over all previous iterations is p_i^n and the swarm's best candidate solution is p_t^s. Each particle's position at the following iteration $i + 1$ is updated as $x_{i+1}^n = x_i^n + v_{i+1}^n$, where v_{i+1}^n is called velocity and is defined as $v_{i+1}^n = wv_i^n + c_1 r_1 \left(p_i^n - x_i^n \right) + c_2 r_2 \left(p_i^s - x_i^n \right)$ where c_1, c_2 are constant parameters, r_1, r_2 are draws from a uniform distribution defined in [0, 1], and w is inertia weight. Inertia weight can be either fixed ex-ante or be a function of time, and its impact is that it controls PSO trade-off between exploration of new areas versus exploitation of already best candidate solutions.

The basic PSO algorithm is described as follows.

Initialise the N particles uniformly in the K-dimensional space and set some random initial velocities

repeat
    for each particle $n = 1,\cdots, N$ do
        Update the velocity: v_{i+1}^n
        Update the position: x_{i+1}^n
        Evaluate the objective function: $f\left(x_{i+1}^n\right)$
    end

for each particle $n = 1,\cdots, N$ do

if $f\left(x_{i+1}^n\right) < f\left(p_i^n\right)$ then

$p_{i+1}^n = x_{i+1}^n$

else

$p_{i+1}^n = p_i^n$

end

if $f\left(x_{i+1}^n\right) < f\left(p_i^s\right)$ then

$p_{i+1}^s = x_{i+1}^n$

else

$p_{i+1}^s = p_i^s$

end

end

until a stopping criterion is met

Regarding the implementation details, the swarm size was set equal to 18 and as a stopping criterion the objective function's tolerance was set to 0.1. Since, as shown previously in the sensitivity analysis of the log marginal likelihood, the values which maximise it lie close to conventional values in the literature, 2/3 of the particles were initialized in all possible combinations (12 in total) of hyperparameters $\lambda_1 \in \{0.1, 0.2\}$, $\lambda_2 \in \{0.25, 1\}$, $\lambda_3 \in \{1, 2, 4\}$, while the remaining six were initialized randomly. The lower and upper bound for λ_1, λ_2, λ_3 were [0.05, 0.5], [0.1, 5], and [0.1, 5], respectively.

A.1.2 Kalman filtering and smoothing

For the state-space model defined according to equations (1.22) and (1.23), the Kalman filter with missing observations is described by the following recursion for $t = T + 1,\cdots, T + H$. The prediction step consists of the following equations.

$$S_{t|t-1} = FS_{t-1|t-1}, \quad P_{t|t-1} = FP_{t-1|t-1}F' + Q_t$$

The update step, in order to account for the missing observations, i.e. the conditional forecasts that will be estimated, should adjust the standard equations by removing the corresponding rows and columns for the case of R referring to missing observations. In short, the update step consists of the following two equations:

$$S_{t|t} = S_{t|t-1} + K_t\xi_{t|t-1}, \quad P_{t|t} = P_{t|t-1}F' - K_tGP_{t|t-1}$$

Where $\xi_{t|t-1} = y_t - GS_{t|t-1}$ is the prediction error, $f_{t|t-1} = GP_{t|t-1}G + R_t$ is the prediction error's variance and $K_t = P_{t|t-1}G'f_{t|t-1}^{-1}$ is the Kalman gain.

The de Jong (1988) fixed interval Kalman smoother, as further clarified in Koopman (1997), is described by the following recursion for $t = T + H,\cdots, T + 1$.

$$r_{t-1} = G'F^{-1}\xi_{t|t-1} + L_t'r_t, \quad S_{t|t+H} = S_{t|t-1} + P_{t|t-1}r_{t-1}$$

The recursion is initialised with $r_{T+H} = 0$, and the quantity L_t is estimated during the Kalman filter recursion for $t = T + 1, \cdots, T + H$ with $L_t = F - F K_t G$.

A.1.3 Stochastic volatility estimation

Following Kim, Shephard, and Chib (1998), stochastic volatility as a vector is estimated using the following state-space model:

$$\tilde{y}_t = \tilde{h}_t + \tilde{e}_t, \quad \tilde{e}_{n,t} \left| s_{n,t} = j \sim \mathcal{N}\left(\mu_j, \sigma_j^2\right) \right. \tag{A.1}$$

$$\tilde{h}_t = \tilde{h}_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}\left(0, \Phi\right) \tag{A.2}$$

Where $\tilde{y}_t = \log\left(\left(A\varepsilon_t\right)^2 + c\right)$ with A and ε_t referring to equations (1.17) and (1.16), respectively. Further $\tilde{h}_t = \ln h_t$ corresponds to equation (1.20) in a vector notation. In order to avoid numerical issues for $\varepsilon_{n,t} \to 0$, c is set to 10^{-6}. As shown by Kim, Shephard, and Chib (1998) $\tilde{e}_{n,t}$ does not follow a normal distribution anymore, but its log-chi-square distribution can be approximated by a seven-component mixture of normal distributions:

$$f\left(\tilde{e}_{i,t}\right) = \sum_{j=1}^{7} p_j \varphi\left(\tilde{e}_{i,t} \left| \mu_j - 1.2704, \sigma_j^2\right.\right)$$

Where $\varphi(x \,|\, \mu, \sigma^2)$is the probability density function of a normal distribution. The exact values of μ_j and σ_j^2 are available in Kim, Shephard, and Chib (1998), Table 4. The state-space model defined by equations (A.1) and (A.2) conditional on the state indicator $s_{n,t} \in \{1, \cdots, 7\}$, is a standard linear Gaussian state-space model where the state vector $S_t = \tilde{h}_t$ is estimated by a forward pass of the Kalman filter as described in Appendix A.1.2 Kalman filtering and smoothing and the Carter and Kohn (1994) smoothing algorithm with draws from the following distributions:

$$\tilde{S}_T \sim \mathcal{N}\left(S_{T|T}, P_{T|T}^{ss}\right), \quad t = T$$

$$\tilde{S}_t \sim \mathcal{N}\left(S_{t|T}^*, P_{t|T}^*\right), \quad t = T - 1, \cdots, 1$$

Where $S_{t|T}^* = S_{t|t} + P_{t|t}^{ss}\left(P_{t|t}^{ss} + Q\right)^{-1}\left(\tilde{S}_{t+1} - S_{t|t}\right)$ and $P_{t|T}^* = P_{t|t}^{ss} - P_{t|t}^{ss}\left(P_{t|t}^{ss} + Q\right)^{-1} P_{t|t}^{ss}$.

A.2 Additional results – Log predictive score

Log predictive score (LPS) for variable n and horizon h is defined as:

$$\log S_{n,h} = \frac{1}{T-H-t_0-1}\sum_{t=t_0}^{T-H} \log f\left(x = Y_{n,t+h}\right)$$

Where f denotes the probability density function of a t location-scale distribution fitted to the data: $f(x|\mu,\sigma,v)$, where the location, scale and shape parameters are estimated using the predictive density of the MCMC draws. For the rest of this section, when referring to LPS, I actually refer to minus LPS, i.e. the lower the value, the better the performance.

Table A.1: Steady-state prior conditional versus unconditional forecasts evaluated with LPS

	GDP	CPI	FFR	10-Y	Unemployment	Employment
			SS unconditional forecasts			
$h = 1$	2.327	2.254	0.716	0.632	0.184	1.585
$h = 4$	2.526	2.178	1.824	1.684	2.212	2.429
$h = 8$	2.505	2.295	2.415	2.134	3.218	2.432
$h = 12$	2.419	2.379	2.768	2.249	3.264	2.200
			SS conditional forecasts 2-step estimation			
$h = 1$	-0.050*	-0.058	-	-0.108***	0.067	0.020
$h = 4$	-0.147***	0.093	-	-0.220*	-0.252*	-0.210*
$h = 8$	-0.116*	-0.035	-	-0.578**	-0.527*	-0.149
$h = 12$	0.035	-0.135**	-	-0.912**	-0.142	0.158
			SS conditional forecasts joint estimation			
$h = 1$	-0.055**	0.011	-	-0.119***	-0.017	-0.124*
$h = 4$	-0.162***	0.060	-	-0.232*	0.182	-0.322***
$h = 8$	-0.117*	0.016	-	-0.556**	-0.570*	-0.118
$h = 12$	0.042	-0.160**	-	-0.907**	-0.162***	0.131
			SS-SV unconditional forecasts			
$h = 1$	2.371	2.045	0.303	0.623	-0.021	1.501
$h = 4$	2.497	2.158	1.905	1.448	1.818	2.128
$h = 8$	2.506	2.153	2.432	1.809	2.505	2.299
$h = 12$	2.488	2.242	2.678	1.994	2.816	2.283
			SS-SV conditional forecasts 2-step estimation			
$h = 1$	0.047	0.043	-	-0.049*	0.127	0.053
$h = 4$	-0.153**	-0.126**	-	-0.162*	-0.072	-0.058
$h = 8$	-0.176***	-0.030	-	-0.422***	-0.387**	-0.254**
$h = 12$	-0.119***	-0.056	-	-0.595***	-0.426**	-0.243***
			SS-SV conditional forecasts joint estimation			
$h = 1$	0.021	-0.002	-	-0.084***	0.101	-0.012
$h = 4$	-0.161***	-0.109***	-	-0.207**	-0.276**	-0.079
$h = 8$	-0.210***	-0.042	-	-0.450***	-0.460**	-0.257**
$h = 12$	-0.159***	-0.101**	-	-0.604***	-0.537**	-0.304***

Note. Panels with unconditional forecasts show the raw LPS. Conditional forecasts for each specification show the difference in LPS with respect to the unconditional forecasts of the same specification. Values in parentheses show p-values of the Diebold and Mariano (1995) one-sided test with finite sample adjustment of Harvey, Leybourne, and Newbold (1997) with respect to the unconditional forecasts of the same specification. *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively. 2-step and joint estimation conditional forecasts refer to parameters' uncertainty accounting for the 'in-sample' data only and the 'in-sample' data extended with the conditional forecasts respectively. The evaluation period is 1997Q1-2014Q1.

Table A.2: Steady-state versus Minnesota prior for conditional and unconditional forecasts evaluated with LPS

	GDP	CPI	FFR	10-Y	Unemployment	Employment
			SS unconditional forecasts			
$h = 1$	0.003	-0.129	0.031	-0.006	-0.054	-0.101*
$h = 4$	0.051	-0.093*	-0.057	-0.021	-0.020	0.225
$h = 8$	0.022	-0.069**	0.031	0.002	0.129	0.225
$h = 12$	0.022	-0.010	0.131	0.018	-0.074	0.077
			SS conditional forecasts 2-step estimation			
$h = 1$	-0.046*	-0.216	-	-0.115***	-0.047	-0.028
$h = 4$	0.008	-0.072	-	-0.053	-0.340*	0.145
$h = 8$	0.007	-0.048	-	-0.133**	-0.215**	0.289
$h = 12$	0.042	-0.054	-	-0.192**	-0.311*	0.329
			SS conditional forecasts joint estimation			
$h = 1$	-0.051**	-0.146	-	-0.127***	-0.132*	-0.172**
$h = 4$	-0.008	-0.105	-	-0.064	0.094	0.034
$h = 8$	0.007	0.003	-	-0.111*	-0.258**	0.320
$h = 12$	0.049	-0.079	-	-0.187***	-0.331*	0.302
			SS-SV unconditional forecasts			
$h = 1$	0.047	-0.339	-0.382***	-0.015	-0.258*	-0.185
$h = 4$	0.023	-0.114	0.024	-0.257*	-0.414	-0.075
$h = 8$	0.022	-0.211	0.047	-0.323	-0.585	0.091
$h = 12$	0.090	-0.147	0.041	-0.237	-0.522	0.159
			SS-SV conditional forecasts 2-step estimation			
$h = 1$	0.095	-0.324	-	-0.064	-0.191	-0.080
$h = 4$	-0.027	-0.311	-	-0.230*	-0.554	-0.002
$h = 8$	-0.051	-0.185	-	-0.302	-0.788	0.050
$h = 12$	-0.043**	-0.112	-	-0.131	-1.044	0.011
			SS-SV conditional forecasts joint estimation			
$h = 1$	0.069	-0.369	-	-0.099**	-0.218	-0.145
$h = 4$	-0.035	-0.294	-	-0.275*	-0.758	-0.024
$h = 8$	-0.085*	-0.196	-	-0.330*	-0.861	0.048
$h = 12$	-0.084**	-0.157	-	-0.139	-1.155	-0.050

Note. Panels show the LPS difference of SS minus the Minnesota prior (benchmark). Values in parentheses show p-values of the Diebold and Mariano (1995) one-sided test with finite sample adjustment of Harvey, Leybourne, and Newbold (1997). *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively. 2-step and joint estimation conditional forecasts refer to parameters' uncertainty accounting for the 'in-sample' data only and the 'in-sample' data extended with the conditional forecasts respectively. The evaluation period is 1997Q1-2014Q1.

Appendix B (Essay 2)

B.1 Technical Appendix

B.1.1 Particle Gibbs with ancestor sampling

Following Lindsten, Jordan, and Schon (2014), particle Gibbs with ancestor sampling consists of the following steps. For a state-space model with parameters θ and latent states $S_{1:T,}$ the (r) iteration of the algorithm has the following two blocks:

1. $p\left(\theta \middle| S_{1:T}^{(r-1)}, Y\right)$
2. $p\left(S_{1:T}^{(r)} \middle| \theta^{(r)}, Y\right)$

In particular, the second block concerning the estimation of the latent states uses $n = 1,\cdots, N$ particles, and for the case of TV-LSTAR model is implemented as follows:

1. For $t = 1$;

   a. Draw $S_1^{(n)} \middle| S_0^{(n)}, \theta$ for $n = 1,\cdots, N-1$ using the state equation. Fix $S_1^{(N)} = S_1^{(r-1)}$.

   b. Estimate the normalised weights $W_{t=1}^{(n)} = \dfrac{w_1^{(n)}}{\sum_{n=1}^{N} w_1^{(n)}}$, where $w_1^{(n)}$ is the model's conditional likelihood: $w_1^{(n)} = \left(2\pi R\right)^{-0.5} \exp\left(-0.5 e_1^{2(n)} R^{-1}\right)$ where:

$$e_1 = y_1 - S_{1,1}^{(n)} + \left(1 - G\left(y_{t-d}, \gamma, S_{2,1}^{(n)}\right)\right)\left(\sum_{p=1}^{P} \beta_{1,p} y_{t-p}\right) + G\left(y_{t-d}, \gamma, S_{2,1}^{(n)}\right)\left(\sum_{p=1}^{P} \beta_{2,p} y_{t-p}\right)$$

2. For $t = 2,\cdots, T$

   a. Resample $S_{t-1}^{(n)}$ for $n = 1,\cdots, N-1$ using indices $\alpha_t^{(n)}$ with $\Pr\left(\alpha_t^{(n)} = n\right) \propto W_{t-1}^{(n)}$

   b. Draw $S_t^{(n)} \middle| S_{t-1}^{\alpha_t^{(n)}}, \theta$ using the state equation where $S_{t-1}^{\alpha_t^{(n)}}$ denoted the previously resampled particles.

   c. Fix $S_t^{(N)} = S_t^{(r-1)}$

   d. Sample $\alpha_t^{(N)}$ with $\Pr\left(\alpha_t^{(N)} = n\right) \propto W_{t-1}^{(n)} \Pr\left(S_t^{(r-1)} \middle| S_{t-1}^{(n)}, \theta\right)$ where the density $\Pr\left(S_t^{(r-1)} \middle| S_t^{(n)}, \theta\right)$ is equal to $\left(2\pi Q\right)^{-0.5} \exp\left(-0.5 e_1^{2(n)} R^{-1}\right)$

   e. Update the normalised weights $W_t^{(n)} = \dfrac{w_t^{(n)}}{\sum_{n=1}^{N} w_t^{(n)}}$, where $w_t^{(n)}$ is the model's conditional likelihood:

$$w_t^{(n)} = (2\pi)^{-2/2} |Q|^{-0.5} \exp\left(-0.5\left(S_t^{(n)} - S_t^{(r-1)}\right)' Q^{-1} \left(S_t^{(n)} - S_t^{(r-1)}\right) \right)$$

f. Update the states' trajectories with the ancestor indices: $\tilde{S}_{1:t}^{(n)} = \left\{ \tilde{S}_{1:t-1}^{\alpha_t^{(n)}}, S_t^{(n)} \right\}$

3. Draw k with $\Pr(k = n) \propto W_T^{(n)}$ and return $S_{1:T}^{(r)} = \tilde{S}_{1:T}^{(k)}$.

B.1.2 Particle filtering

Assuming that the parameters θ of a non-linear state-space model described below are known, then the unknown posterior distribution of the states $S_{1:T}$ can be estimated using the standard bootstrap particle filtering method of Gordon, Salmond, and Smith (1993).

$$y_t \sim p(y_t | S_t) \tag{B.1}$$

$$S_t \sim p(S_t | S_{t-1}) \tag{B.2}$$

Assuming $n = 1,\cdots, N$ particles, the algorithm is initialised by drawing N samples from the prior:

$$S_0^{(n)} \sim p(S_0), \quad n = 1,\cdots, N$$

For periods $t = 1,\cdots, T$ perform a prediction and an update step:

1. Prediction step: Using the state equation, draw N samples:

$$S_t^{(n)} \sim p\left(S_t | S_{t-1}^{(n)}\right), \quad n = 1,\cdots, N$$

In particular, since the states in the generic case are assumed to follow an autoregressive process, this step is defined as:

$$S_t^{(n)} = C + FS_{t-1}^{(n)} + v_t, \quad v_t \sim \mathcal{N}(0, Q), \quad n = 1,\cdots, N$$

2. Update step: Calculate the weights for each particle n based on the likelihood of the model given by equation (2.34) conditional on the predicted state at period t:

$$w_t^{(n)} \propto p\left(y_t | S_t^{(n)}\right), \quad n = 1,\cdots, N$$

The weights are normalised to sum to unity: $\tilde{w}_t^{(n)} = w_t^{(n)} \Big/ \sum_{n=1}^{N_p} w_t^{(n)}$

3. Resample step $S_t^{(n)}$ with probability $\Pr(i = n) \propto \tilde{w}_t^{(n)}$ Resampling step addresses the fundamental degeneracy problem, which is present in sequential importance resampling algorithms. In particular, the standard multinomial resampling is employed. A discrete distribution over particles $n = 1,\cdots, N$ is defined using the probability mass of the

normalised weights $\tilde{w}_t^{(n)}$. Then N draws are performed from this multinomial distribution and substitute the old N particles with the resampled ones resulting in the filtered distribution:

The integrated likelihood, which has been estimated once the latent states $S_{1:T}$ have been integrated out $\mathcal{L}(\theta|Y) = p(Y|\theta) = \int p(Y,S|\theta)dS_{1:T}$, following Andrieu, Doucet, and Holenstein (2010), can be approximated as:

$$\hat{\mathcal{L}}(\theta|Y) = \hat{p}(Y|\theta) = \prod_{t=1}^{T} \frac{1}{N} \sum_{n=1}^{N} w_t^{(n)} \tag{B.3}$$

B.1.3 Stochastic volatility estimation

Following Kim, Shephard, and Chib (1998), stochastic volatility as a scalar is estimated using the following state-space model:

$$\tilde{y}_t = \tilde{h}_t + \xi_t, \quad \xi_t | s_t = i \sim \mathcal{N}(\mu_i, \sigma_i^2) \tag{B.4}$$

$$\tilde{h}_t = \tilde{h}_{t-1} + \eta_t, \quad \eta_t \sim iid\ \mathcal{N}(0, \sigma_\eta^2) \tag{B.5}$$

Where $\tilde{y}_t = \log(e_t^2 + c)$ with e_t referring to equation (2.19). Further $\tilde{h}_t = \ln h_t$ corresponds to equation (2.41). In order to avoid numerical issues for $e_t \to 0$, c is set to 10^{-6}. As shown by Kim, Shephard, and Chib (1998), ξ_t does not follow a normal distribution anymore, but its log-chi-square distribution can be approximated by a seven-component mixture of normal distributions:

$$f(\xi_t) = \sum_{i=1}^{7} p_i \varphi(\xi_t | \mu_i - 1.2704, \sigma_i^2)$$

Where $\varphi(x|\mu,\sigma^2)$ is the probability density function of a normal distribution. The exact values of μ_i and σ_i^2 are available in the Kim, Shephard, and Chib (1998) Table 4. The state-space model defined by equations (B.4) and (B.5) conditional on the state indicator $s_t \in \{1,\cdots,7\}$ is a standard linear Gaussian state-space model and is estimated by the Carter and Kohn (1994) algorithm.

Appendix C (Essay 3)

C.1 Technical appendix

C.1.1 Stochastic volatility estimation

Following Kim, Shephard, and Chib (1998), stochastic volatility as a vector is estimated using the following state-space model:

$$\tilde{y}_t = \tilde{h}_t + \tilde{e}_t, \quad \tilde{e}_{n,t} \left| s_{n,t} = j \sim \mathcal{N}\left(\mu_j, \sigma_j^2\right) \right. \tag{C.1}$$

$$\tilde{h}_t = \tilde{h}_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}\left(0, \Phi\right) \tag{C.2}$$

Where $\tilde{y}_t = \log\left(\left(Ae_t\right)^2 + c\right)$ with A and e_t referring to equations (3.85) and (3.84), respectively. Further $\tilde{h}_t = \ln h_t$ corresponds to equation (3.88) in a vector notation. In order to avoid numerical issues for $e_{n,t} \to 0$, c is set to 10^{-6}. As shown by Kim, Shephard, and Chib (1998) $\tilde{e}_{n,t}$ does not follow a normal distribution anymore, but its log-chi-square distribution can be approximated by a seven-component mixture of normal distributions:

$$f\left(\tilde{e}_{n,t}\right) = \sum_{j=1}^{7} p_j \varphi\left(\tilde{e}_{n,t} \left| \mu_j - 1.2704, \sigma_j^2\right.\right)$$

Where $\varphi(x \mid \mu, \sigma^2)$ is the probability density function of a normal distribution. The exact values of μ_j and σ_j^2 are available in Kim, Shephard, and Chib (1998), Table 4. The state-space model defined by equations (C.1) and (C.2) conditional on the state indicator $s_{n,t} \in \{1, \cdots, 7\}$ is a standard linear Gaussian state-space model and is estimated by Carter and Kohn (1994) algorithm.

C.1.2 Marginal likelihood estimation

Below, it follows the estimation of the log marginal likelihood for a homoscedastic MBVAR using the approximation method of Chib (1995). Beginning with Bayes' rule:

$$p\left(\theta \left| Y, \mathcal{M}\right.\right) = \frac{p\left(Y \left| \theta, \mathcal{M}\right.\right) p\left(\theta, \mathcal{M}\right)}{p\left(Y \left| \mathcal{M}\right.\right)}$$

by simple re-ordering, the log marginal likelihood is equal to:

$$\ln p\left(Y \left| \mathcal{M}\right.\right) = \ln p\left(Y \left| \theta, \mathcal{M}\right.\right) + \ln p\left(\theta, \mathcal{M}\right) - \ln p\left(\theta \left| Y, \mathcal{M}\right.\right)$$

This formula is valid at any point of the posterior distribution of parameters θ; however, for the approximation to be as accurate as possible, a point of support $\tilde{\theta}$ with high density is chosen, where $\tilde{\theta}$ is the mean of the posterior's distribution. In the following equations, the index $\mathcal{M}$ representing the model hyperparameters is dropped for notational simplicity.

The term representing the MBVAR's likelihood is:

$$p\left(Y\middle|\tilde{\theta}\right)=\left(2\pi\right)^{-0.5TN}\left|\tilde{\Sigma}\right|^{-0.5T}\exp\left(-0.5vec\left(E\right)\left(\tilde{\Sigma}\otimes I_T\right)^{-1}vec\left(E\right)\right)$$

The parameters' vector θ prior is comprised of the two terms corresponding to the priors of B and Σ, respectively:

$$\ln p\left(\tilde{\theta}\right)=\ln p\left(\tilde{B},\tilde{\Sigma}\right)=\ln p\left(\tilde{B}\right)+\ln p\left(\tilde{\Sigma}\right)$$

In the same manner, the posterior distribution is comprised of the following two parts:

$$\ln p\left(\tilde{\theta}\middle|Y\right)=\ln p\left(\tilde{B},\tilde{\Sigma}\middle|Y\right)=\ln p\left(\tilde{B}\middle|\tilde{\Sigma},Y\right)+\ln p\left(\tilde{\Sigma}\middle|Y\right)$$

Chib (1995) showed that the first term $p\left(\tilde{B}\middle|\tilde{\Sigma},Y\right)$ can be estimated by evaluating the pdf of the posterior of B as defined for $\tilde{\Sigma}$ at the point $\tilde{B}$:

$$p_{\mathcal{N}}\left(\tilde{B}\middle|\overline{\mathrm{M}}_B,\overline{\Omega}_B\right)$$

Where: $\overline{\Omega}_{\mathbf{B}}=\left(\tilde{\Sigma}^{-1}\otimes\left(\tilde{\mathrm{X}}'\tilde{\mathrm{X}}\right)+\underline{\Omega}_{\mathrm{B}}^{-1}\right)^{-1}$, $\overline{\mathrm{M}}_B=\overline{\Omega}_B\left(\underline{\Omega}_B^{-1}\underline{\mathrm{M}}_B+\left(\tilde{\Sigma}^{-1}\otimes\tilde{\mathrm{X}}'\right)vec\left(\tilde{Y}\right)\right)$

The second term can be approximated following the 'Rao-Blackwellization' technique as follows:

$$p\left(\tilde{\Sigma}\middle|Y\right)\simeq\frac{1}{R}\sum_{r=1}^{R}p\left(\tilde{\Sigma}\middle|B^{(r)},Y\right)$$

Where $r = 1,\cdots, R$ represent Gibbs sampler retained draws.

Each term of the summation $p\left(\tilde{\Sigma}\middle|B^{(r)},Y\right)$ is the pdf of the inverse Wishart with degrees of freedom $\overline{d}=\underline{d}+T$ and scale matrix: $\overline{S}=\underline{S}+\left(\tilde{Y}-\tilde{X}B^{(r)}\right)'\left(\tilde{Y}-\tilde{X}B^{(r)}\right)$ evaluated at $\tilde{\Sigma}$: $p_{\mathcal{IW}}\left(\tilde{\Sigma}\middle|\overline{S},\overline{d}\right)$

C.1.3 Fast sampling algorithm

Bhattacharya, Chakraborty, and Mallick (2016) suggested the following computationally efficient algorithm for high-dimension univariate regressions when the number of regressors k is much larger than the dimension of the regressor $k \gg T$. The original algorithm was originally suggested for the case of global-local prior on the coefficients β, but here is presented for the more generic case:

$$Y=X\beta+E,\quad E\sim\mathcal{N}\left(0,\sigma^2I_T\right)$$

Where Y is $T \times 1$, X is $T \times k$, β is $k \times 1$, and E is $T \times 1$. Then assuming a normal prior on β, $\beta\sim\mathcal{N}\left(\underline{\mathrm{M}}_\beta,\underline{\Omega}_\beta\right)$ its posterior should be $\beta\mid\sigma^2,Y\sim\mathcal{N}\left(\overline{\mathrm{M}}_\beta,\overline{\Omega}_\beta\right)$ where $\overline{\Omega}_\beta=\left(\underline{\Omega}_\beta^{-1}+\sigma^{-2}\left(X'X\right)\right)^{-1}$, and $\overline{\mathrm{M}}_\beta=\overline{\Omega}_\beta\left(\underline{\Omega}_\beta^{-1}\underline{\mathrm{M}}_\beta+\sigma^{-2}X'Y\right)$. Bhattacharya, Chakraborty, and Mallick (2016) suggested that when the posterior is patterned in the given manner instead of using the Cholesky decomposition to make draws, it is more efficient to follow the algorithm below:

Set $\Phi = X/\sigma$, $D = \underline{\Omega}_\beta$, $\alpha = Y/\sigma$

1. Sample $u \sim \mathcal{N}\left(\underline{\mathrm{M}}_\beta, D\right)$
2. Sample $\delta \sim \mathcal{N}\left(0, \mathrm{I}_T\right)$
3. Set $v = \Phi u + \delta$
4. Solve $\left(\Phi D \Phi' + I_T\right) w = \alpha - v$, and obtain w
5. Set $\theta = u + D\Phi' w$

The value of θ obtained in step 5 is equivalent to a draw from the posterior: $\beta \mid \sigma^2, Y \sim \mathcal{N}\left(\overline{\mathrm{M}}_\beta, \overline{\Omega}_\beta\right)$.

C.1.4 Kalman filtering and smoothing

For the state-space model defined by equations (3.112) and (3.113), I follow algorithm '2.a' of the Durbin and Koopman (2002) simulation smoother, as clarified explicitly in Jarociński (2015). The algorithm is comprised of the following three steps:

1. Perform the recursion of the state-space model equations (3.112) and (3.113) by making draws for the state-space model's innovations e_t and v_t from their respective distributions. For $t = 1, \cdots, T$ estimate S_t^+, Z_t^+ and finally save $Z^+ = \left[Z_1^+, \cdots, Z_T^+\right]'$ and $S^+ = \left[S_1^+, \cdots, S_T^+\right]'$.
2. Apply the Kalman filter and de Jong (1988) Kalman smoother to $Z^* = Z - \mathrm{Z}^+$ and estimate the states $\hat{S}^*$. The intercept of the state equation (3.113) C is set to zero, as suggested in Jarociński (2015). More information on this step is provided below.
3. Finally, estimate a draw of the state as $\tilde{S} = \hat{S}^* + S^+$.

The Kalman filter applied in the second step of the Durbin and Koopman (2002) simulation smoother mentioned above is described by the following recursion for $t = 1, \cdots, T$. Firstly, the prediction step is performed, which consists of the following two equations.

$$S_{t|t-1} = FS_{t-1|t-1}, \quad P_{t|t-1} = FP_{t-1|t-1}F' + Q$$

In continuation, the update step is performed. In order to account for the missing observations in the observations vector Z_t in the observation equation (3.112), the rows with missing observations are removed and additionally, for the case of the covariance matrix R, the relevant columns are as well. In short, the update step consists of the following two equations.

$$S_{t|t} = S_{t|t-1} + K_t \xi_{t|t-1}, \quad P_{t|t} = P_{t|t-1}F' - K_t G P_{t|t-1}$$

Where $\xi_{t|t-1} = Z_t - GS_{t|t-1}$ is the prediction error, $f_{t|t-1} = GP_{t|t-1}G + R_t$ is the prediction error's variance and $K_t = P_{t|t-1}G' f_{t|t-1}^{-1}$ is the Kalman gain.

The de Jong (1988) Kalman smoother, as further clarified in Koopman (1997), is described by the following recursion for $t = T,\cdots,1$.

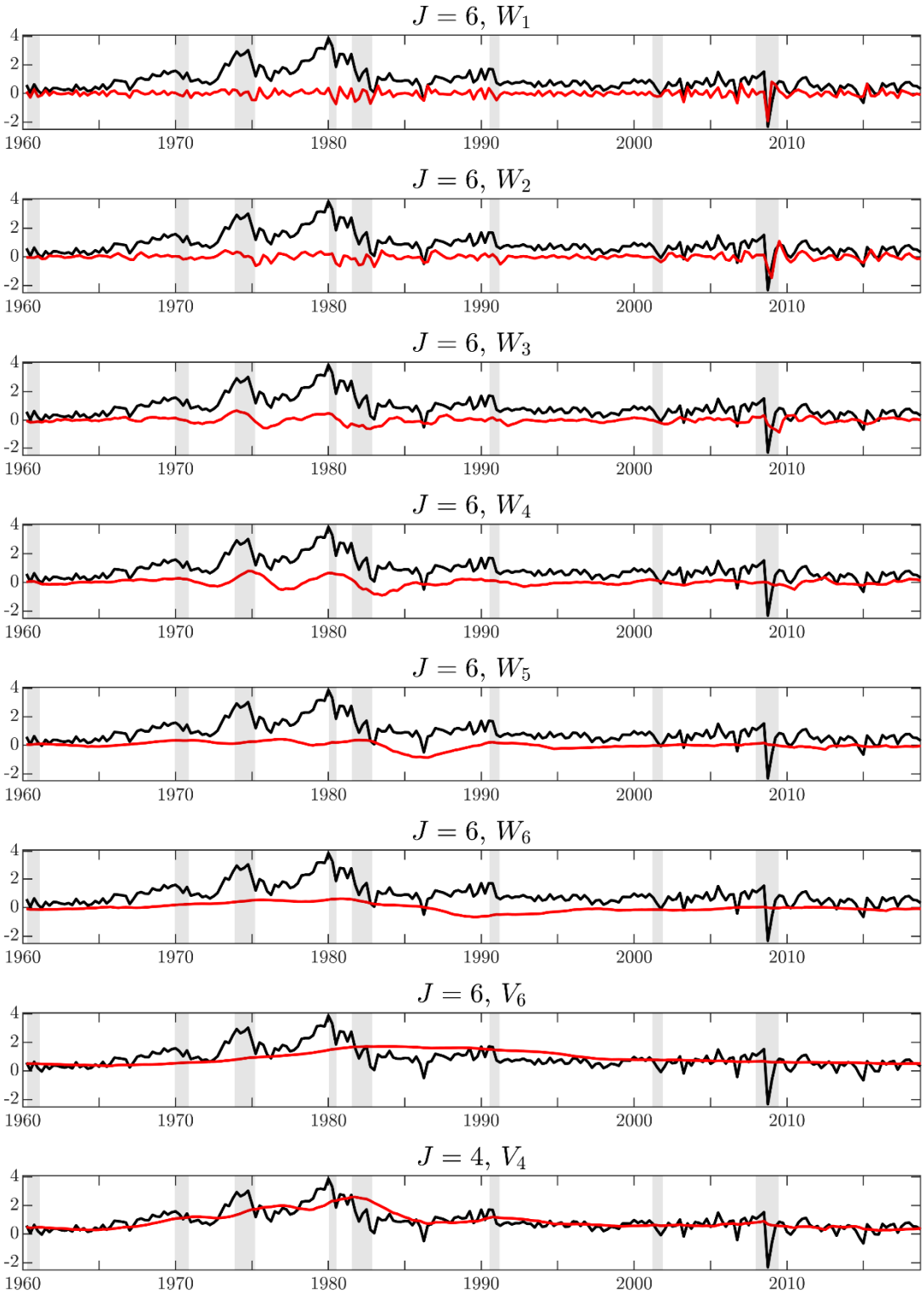
$$r_{t-1} = G'F^{-1}\xi_{t|t-1} + L_t'r_t\,, \quad S_{t|T} = S_{t|t-1} + P_{t|t-1}r_{t-1}$$

The recursion is initialised with $r_T = 0$, and the quantity L_t is estimated during the Kalman filter recursion for $t = 1,\cdots, T$ with $L_t = F - F\,K_t\,G$.

C.2 Additional single-frequency time series results

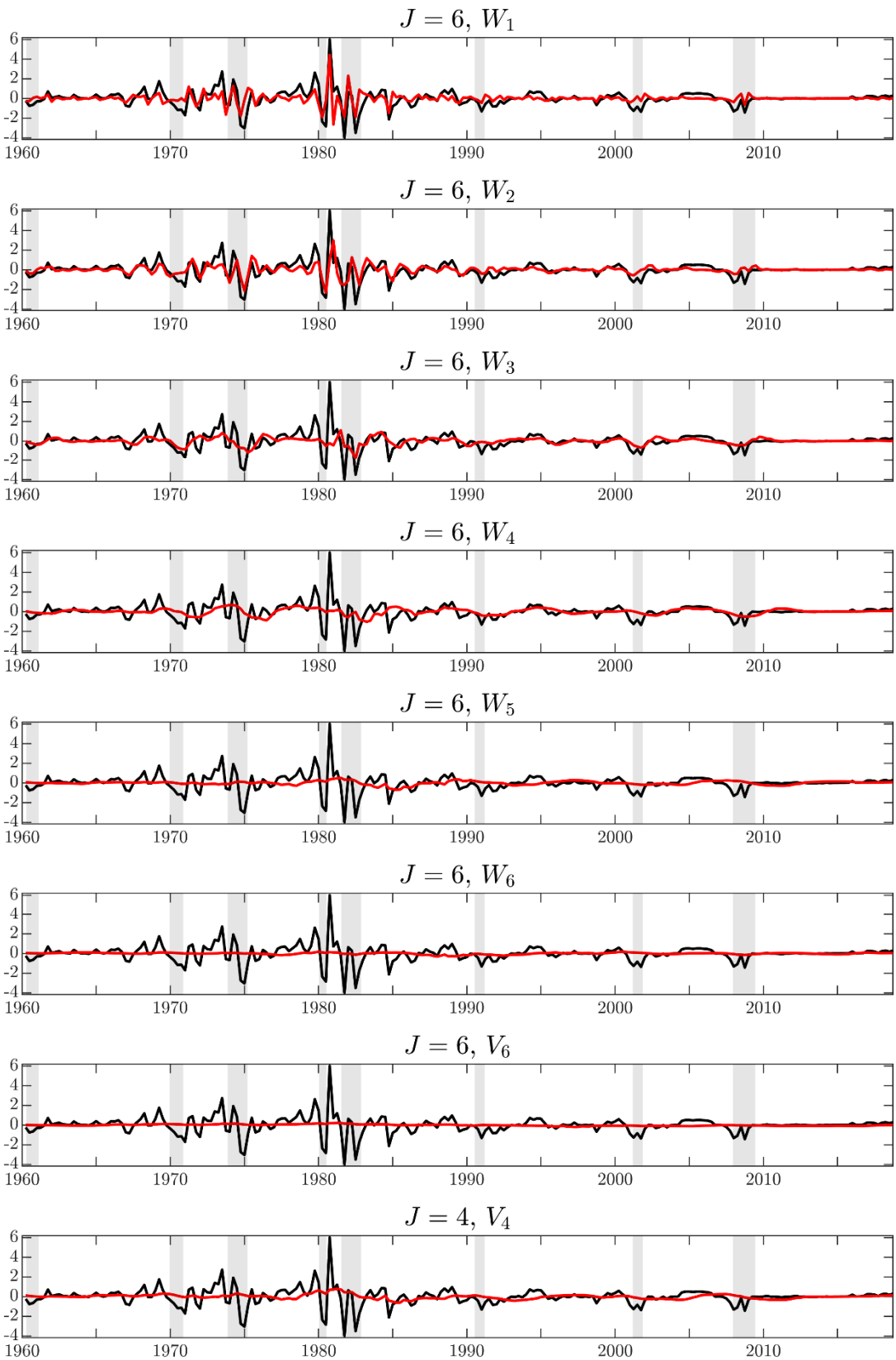
C.2.1 Wavelet and scaling coefficients

Figure C.1: CPI $J = 6$ levels Haar MODWT



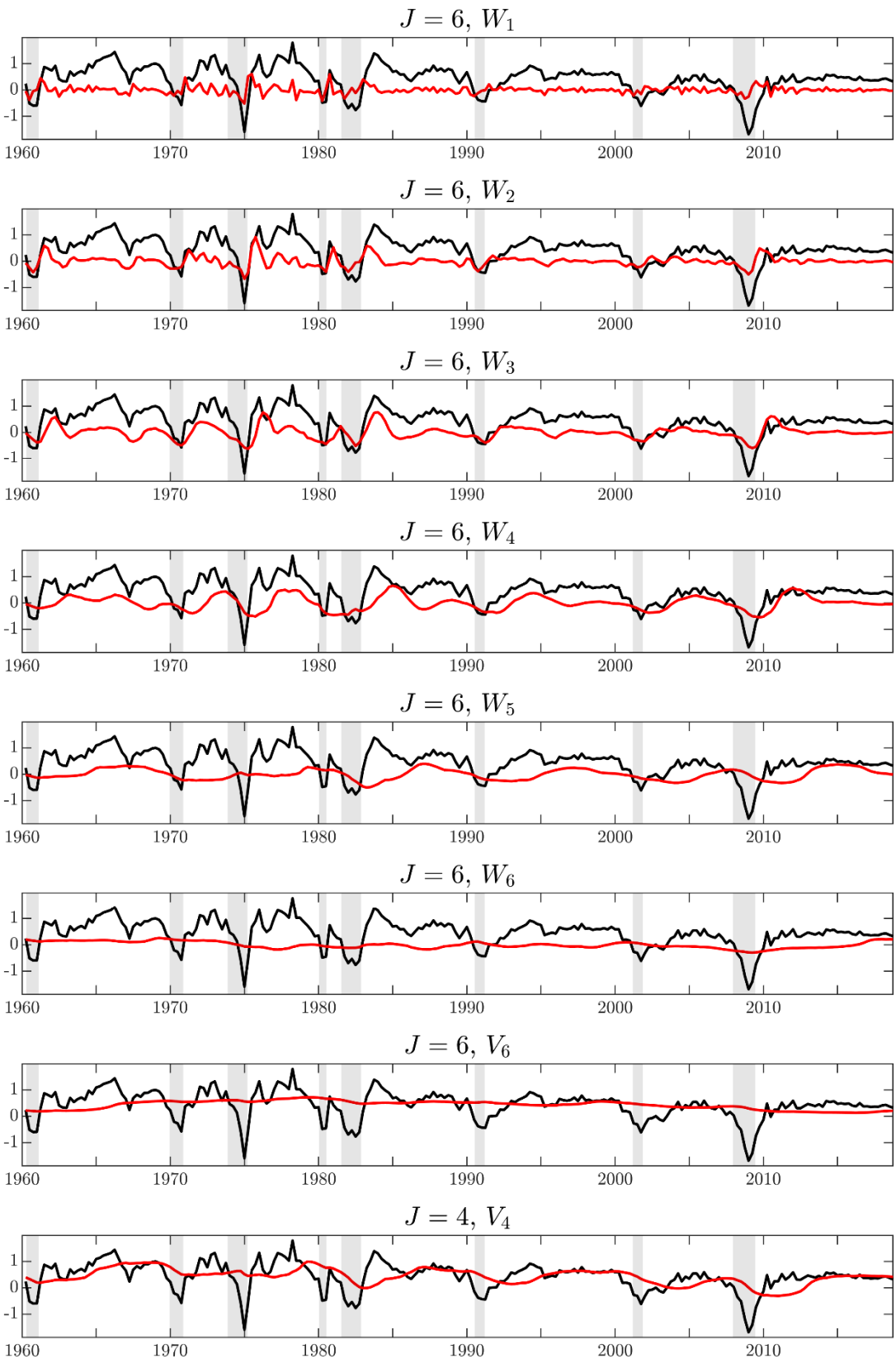
Note. Black lines show actual time series. Red lines show MODWT $J = 6$ wavelet and scaling coefficients and $J = 4$ scaling coefficients.

Figure C.2: FFR $J = 6$ levels Haar MODWT



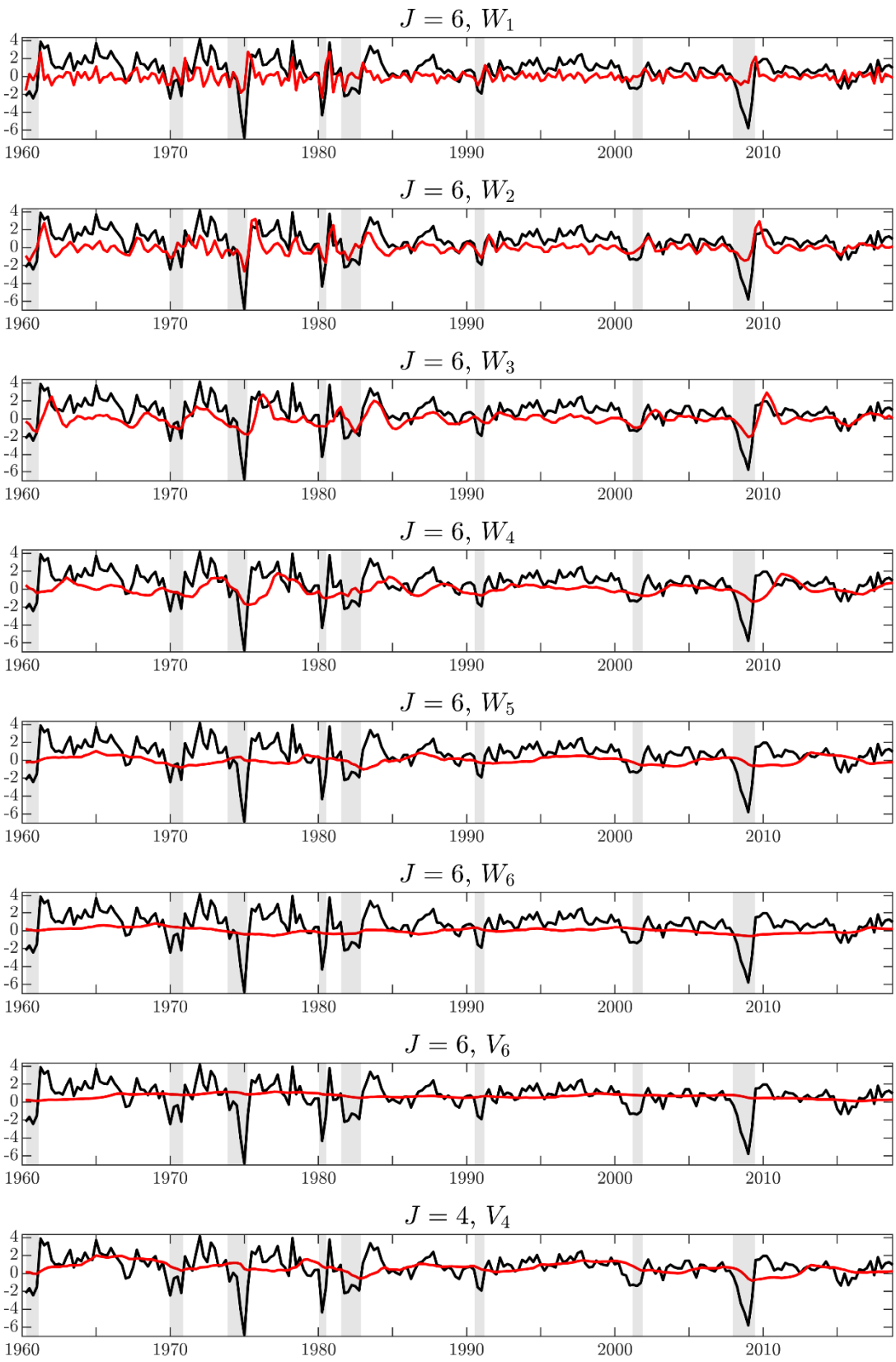
Note. See Figure C.1 note details.

Figure C.3: Employment $J = 6$ levels Haar MODWT



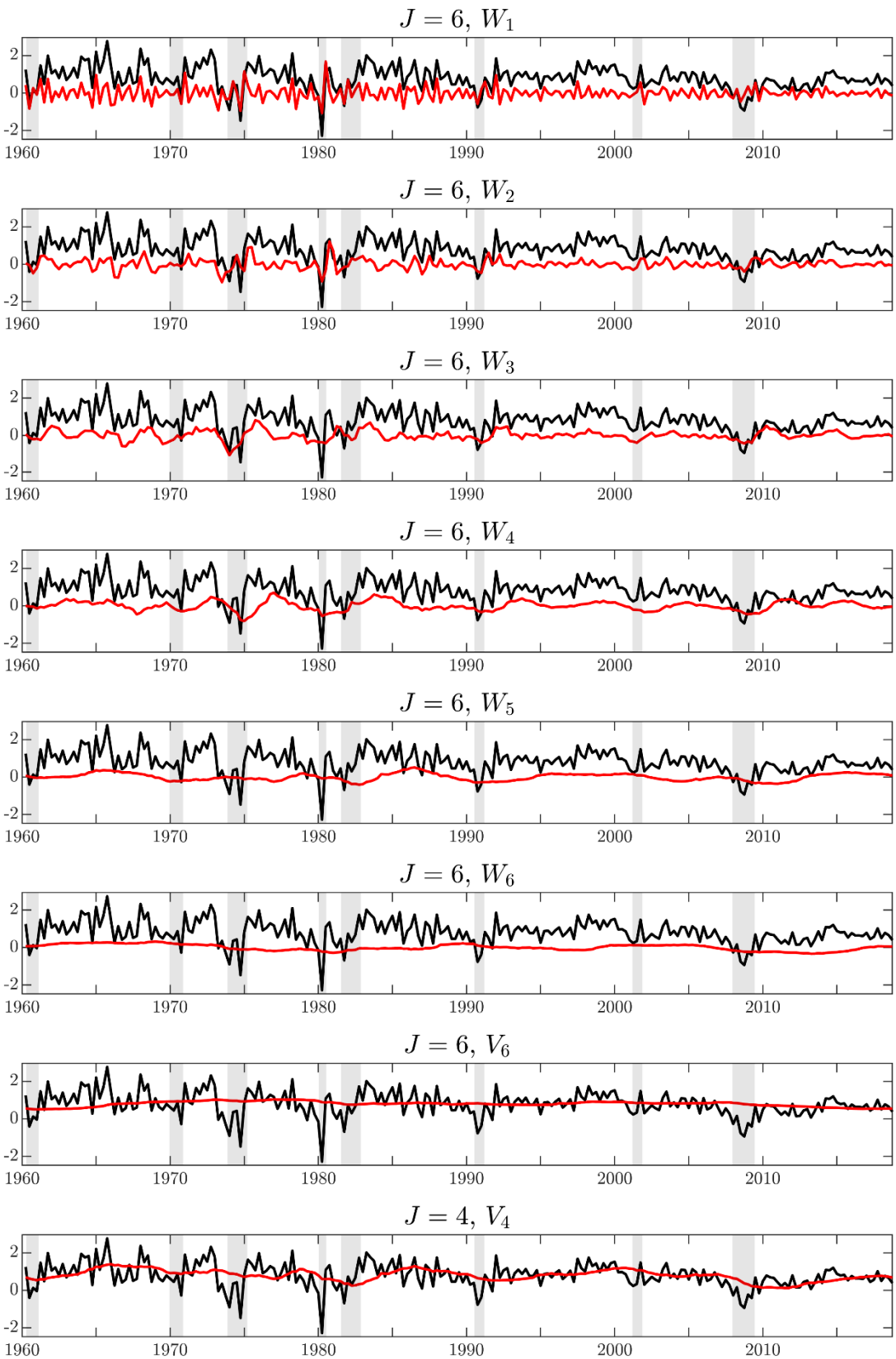
Note. See Figure C.1 note details.

Figure C.4: Industrial production $J = 6$ levels Haar MODWT



Note. See Figure C.1 note details.

Figure C.5: Personal consumption expenditure $J = 6$ levels Haar MODWT



Note. See Figure C.1 note details.

C.2.2 Log predictive score

Log predictive score (LPS) for variable n and horizon h is defined as:

$$\log S_{n,h} = \frac{1}{T-h-t_0-1}\sum_{t=t_0}^{T-h}\log f\left(x = Y_{n,t+h}\right)$$

Where f denotes the probability density function of a t location-scale distribution fitted to the data: $f(x|\mu,\sigma,\nu)$, where the location, scale and shape parameters are estimated using the predictive density of the MCMC draws. For the rest of this section, when referring to LPS, I actually refer to minus LPS, i.e. the lower the value, the better the performance.

Table C.1: Wavelet-based forecasting methods density forecast evaluation with LPS

	GDP	CPI	FFR	Employment	IP	PCE
			BVAR			
$h = 1$	0.897	0.817	0.957	-0.037	1.481	0.787
$h = 1$	1.036	0.827	1.022	0.638	1.735	0.896
$h = 4$	1.052	0.897	1.029	0.843	1.731	0.888
$h = 8$	1.037	0.978	1.036	0.827	1.725	0.879
			Universal hard threshold denoising BVAR			
$h = 1$	-0.076**	0.183	-0.032	0.088	-0.039*	-0.121***
$h = 1$	-0.069***	-0.025*	-0.052***	0.042	0.029**	-0.086***
$h = 4$	-0.074***	-0.050	-0.055***	-0.032	-0.005**	-0.031**
$h = 8$	-0.023**	-0.093***	-0.066***	-0.026	0.008**	-0.033*
			SURE soft threshold denoising BVAR			
$h = 1$	-0.032***	0.166	-0.096***	-0.050	-0.098***	-0.058***
$h = 1$	-0.034**	0.042	-0.106***	0.024	-0.032*	0.016**
$h = 4$	0.055*	-0.041	-0.103***	0.079	0.068	0.044
$h = 8$	-0.028	-0.060**	-0.110***	0.088	-0.031	0.018
			MODWT separate scale BVAR			
$h = 1$	-0.105*	0.455	-0.142***	-0.039	-0.022	-0.127**
$h = 1$	-0.130***	-0.001**	-0.058	0.039	-0.011	-0.161***
$h = 4$	-0.094**	-0.120***	-0.056	-0.119	-0.042	-0.101*
$h = 8$	-0.038*	-0.279***	-0.056	-0.176	-0.036	-0.045
			MODWT separate scale BVAR-SV ($j \leq 4$) / BVAR($j > 4$)			
$h = 1$	-0.072	-0.152	-0.570***	-0.201***	-0.086	-0.094*
$h = 1$	-0.150***	-0.137***	-0.282***	0.163	0.179	-0.174***
$h = 4$	-0.115***	-0.198***	-0.313***	-0.021	0.059	-0.102**
$h = 8$	-0.071**	-0.259***	-0.329***	-0.122	-0.020**	-0.042
			Multiscale BVAR-Minnesota			
$h = 1$	0.136	0.015	0.064	0.284	0.159	0.040
$h = 1$	0.008	-0.025**	0.007	-0.132***	-0.044	-0.051**
$h = 4$	0.006	-0.156***	0.008	-0.175***	-0.032	-0.027***
$h = 8$	0.019	-0.191***	0.007	-0.149***	-0.028**	-0.012**
			Multiscale BVAR-Minnesota-SV			
$h = 1$	-0.079	-0.256	-0.830***	-0.105	-0.075	-0.173***
$h = 1$	-0.118***	-0.171***	-0.564***	-0.126***	0.036	-0.140***
$h = 4$	-0.074**	-0.209***	-0.423***	0.292	0.132	-0.031**
$h = 8$	-0.025	-0.297***	-0.393***	0.562	0.028	-0.014
			Multiscale BVAR-SSVS			
$h = 1$	0.000	0.005	0.011	0.129	0.037	-0.019
$h = 4$	-0.056***	-0.022**	0.028	-0.142**	-0.047	-0.083***
$h = 8$	-0.003	-0.091**	0.036	-0.213**	-0.018	-0.033*
$h = 12$	0.034	-0.159***	0.044	-0.169	-0.003	0.014
			Multiscale BVAR-SSVS-SV			
$h = 1$	-0.138**	-0.209	-0.792***	-0.074	-0.072	-0.121*
$h = 4$	-0.141***	-0.177***	-0.535***	-0.146**	0.102	-0.148***
$h = 8$	-0.086**	-0.140**	-0.407***	0.024	0.055	-0.047**
$h = 12$	-0.042	-0.145	-0.381***	0.043	0.001	0.002

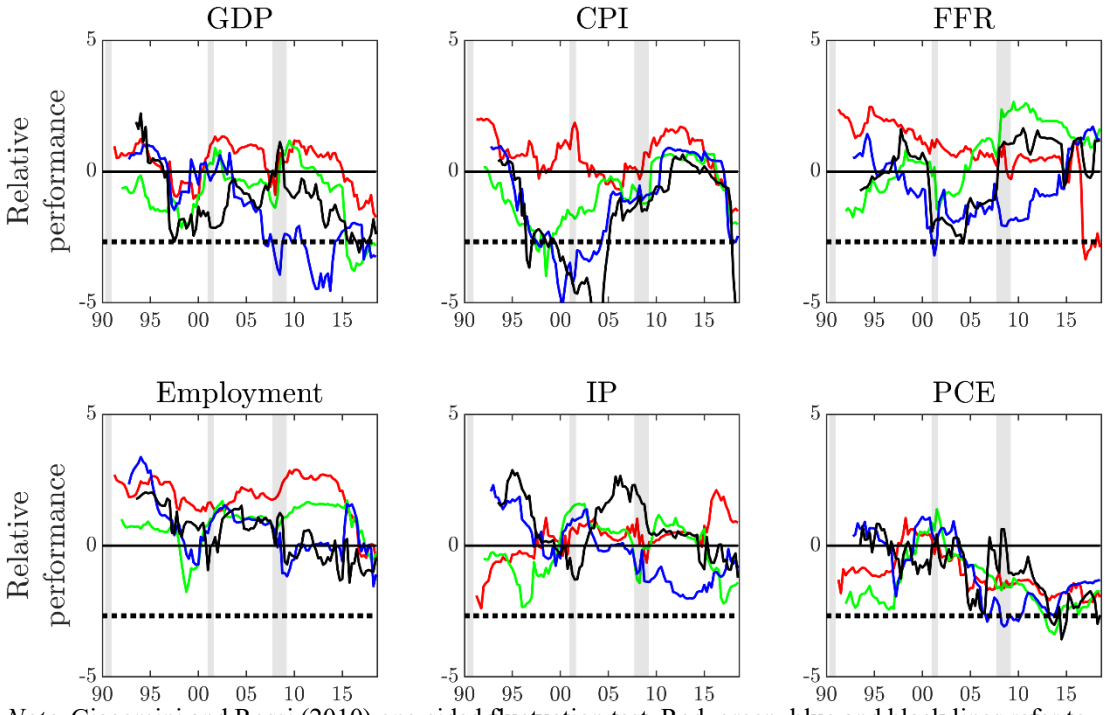
Note. First panel shows benchmark's raw LPS. Remaining panels show LPS differences of the proposed models with respect to the benchmark. Stars refer to the p-values of the Diebold and Mariano (1995) one-sided test with respect to the benchmark. *, **, *** indicate rejection of the null at 10%, 5% and 1% significance level respectively. The evaluation period is 1985-2018.

C.2.3 Time variation of results

This section presents the time variation of the forecast accuracy results. Point forecasts from Figure C.6 to Figure C.13 are evaluated according to the Giacomini and Rossi (2010) fluctuation test, based on a rolling window of 25 quarterly observations[74], where a value smaller than the critical value shown with the dash lines implies the significance of the proposed models with respect to the benchmark model.

Evaluation of the density forecast from Figure C.14 to Figure C.21 is performed by plotting the cumulative difference between the proposed models and the benchmark model, where values below zero imply the superiority of the proposed models.

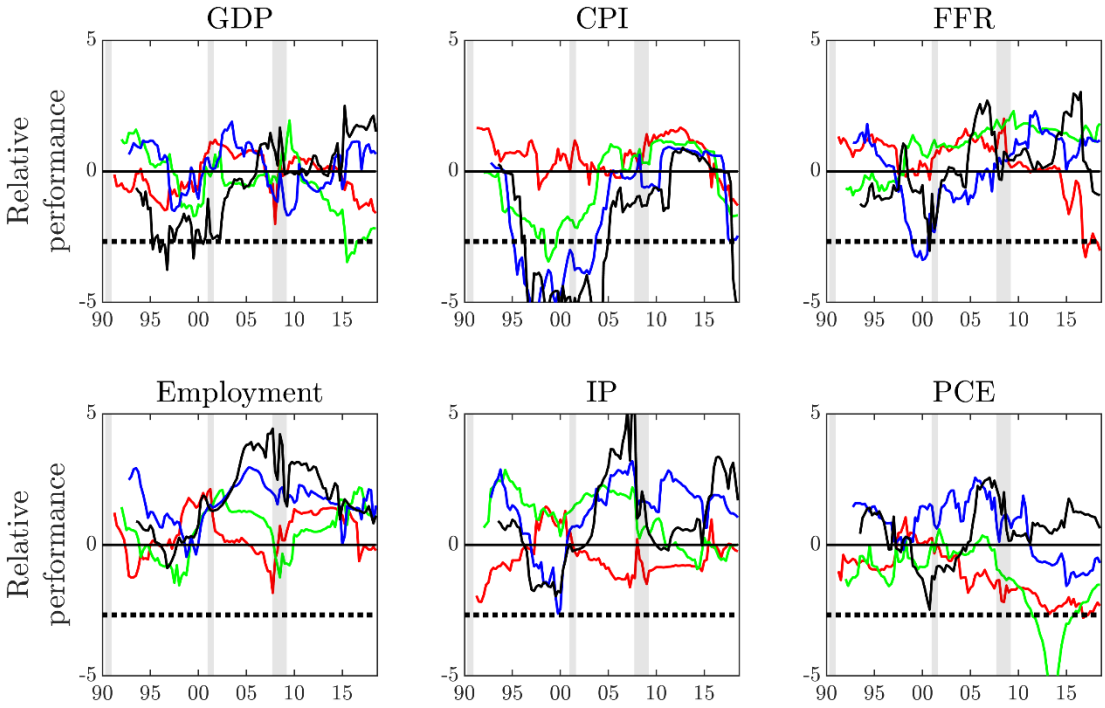
Figure C.6: Fluctuation test of the universal hard threshold denoising BVAR minus the benchmark



Note. Giacomini and Rossi (2010) one-sided fluctuation test. Red, green, blue and black lines refer to 1, 4, 8, and 12-quarters ahead forecasts, respectively. Dash lines show critical values at 5% significance level. Window size is 40 quarters.

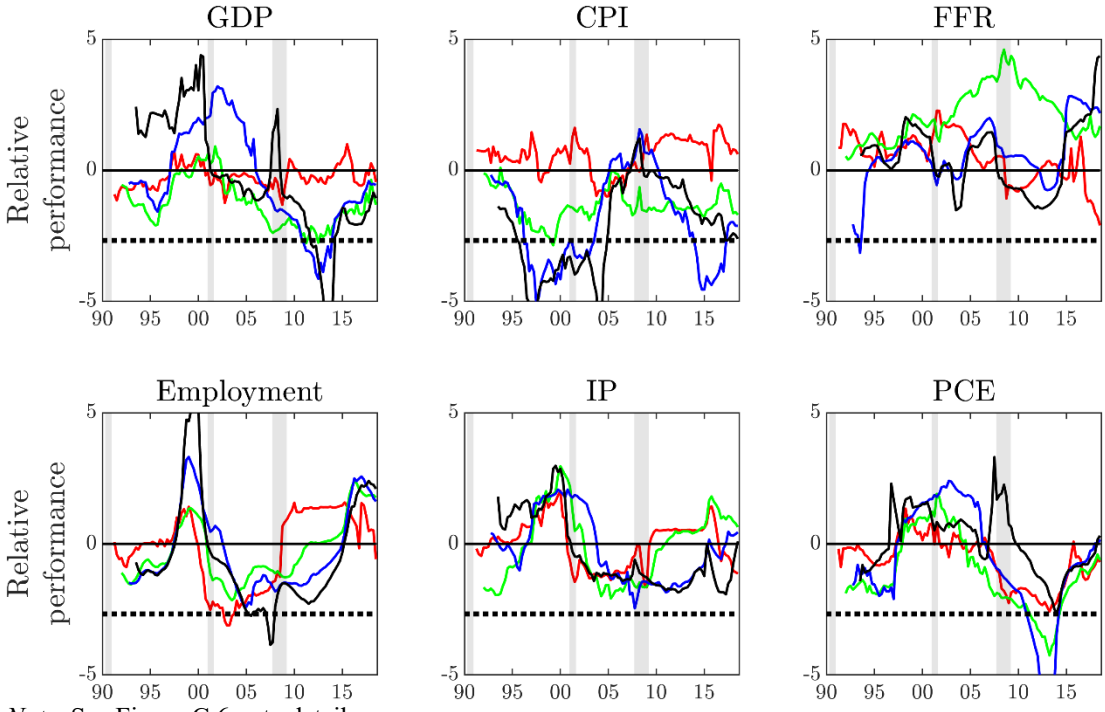
[74] The critical value used is 2.6760 and can be found in Giacomini and Rossi (2010), Table 1.

Figure C.7: Fluctuation test of the SURE soft threshold denoising BVAR minus the benchmark



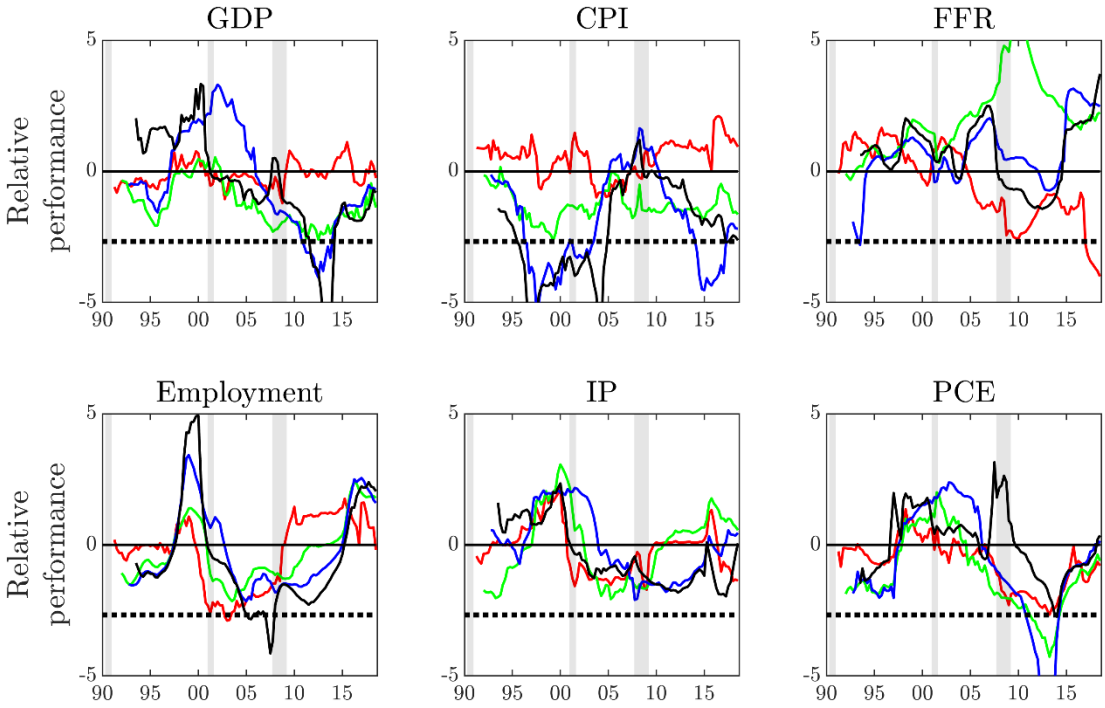
Note. See Figure C.6 note details.

Figure C.8: Fluctuation test of MODWT separate scale BVAR minus the benchmark



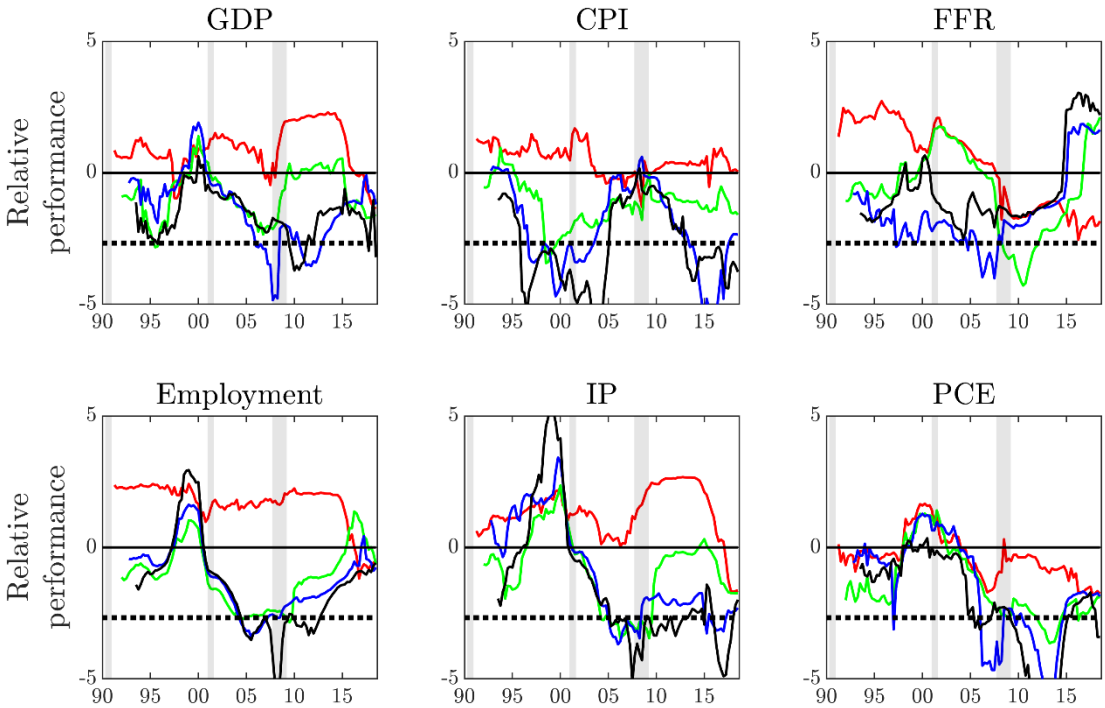
Note. See Figure C.6 note details.

Figure C.9: Fluctuation test of MODWT separate scale BVAR/BVAR-SV minus the benchmark



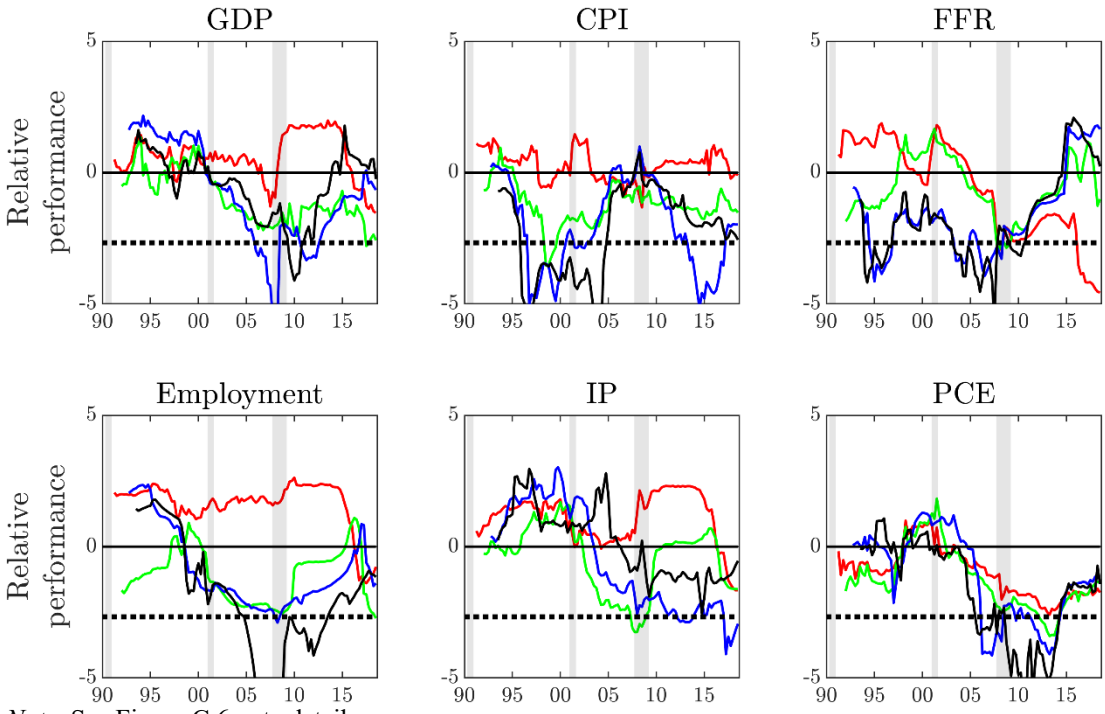
Note. See Figure C.6 note details.

Figure C.10: Fluctuation test of the multiscale BVAR-Minnesota minus the benchmark



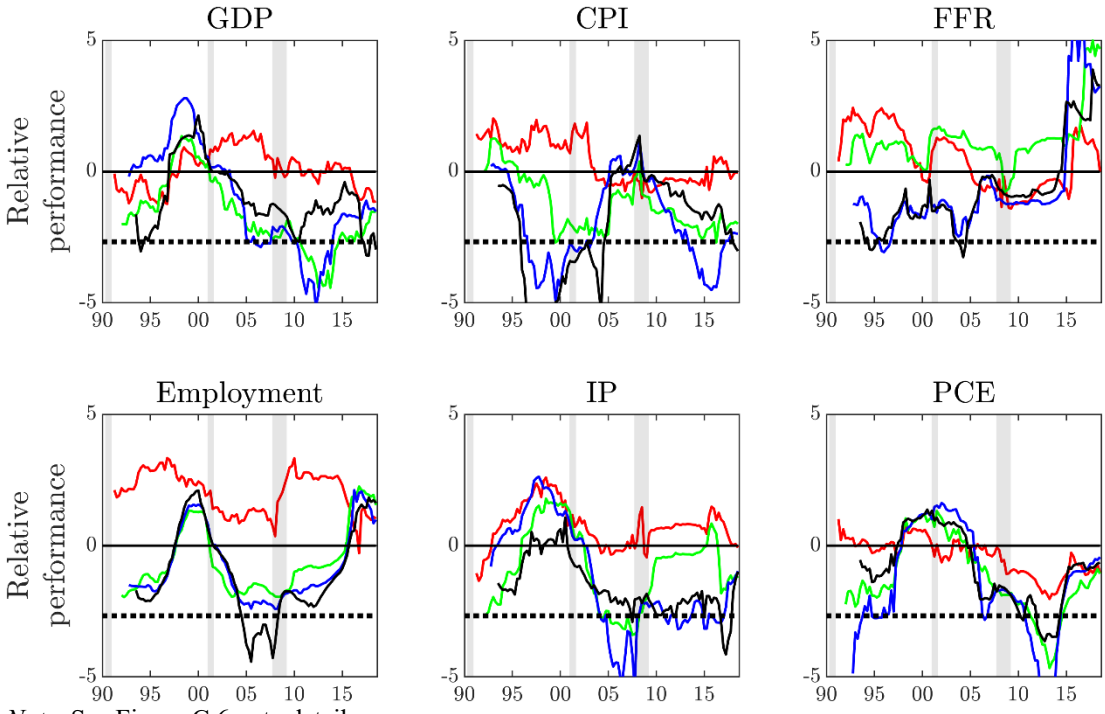
Note. See Figure C.6 note details.

Figure C.11: Fluctuation test of the multiscale BVAR-Minnesota-SV minus the benchmark



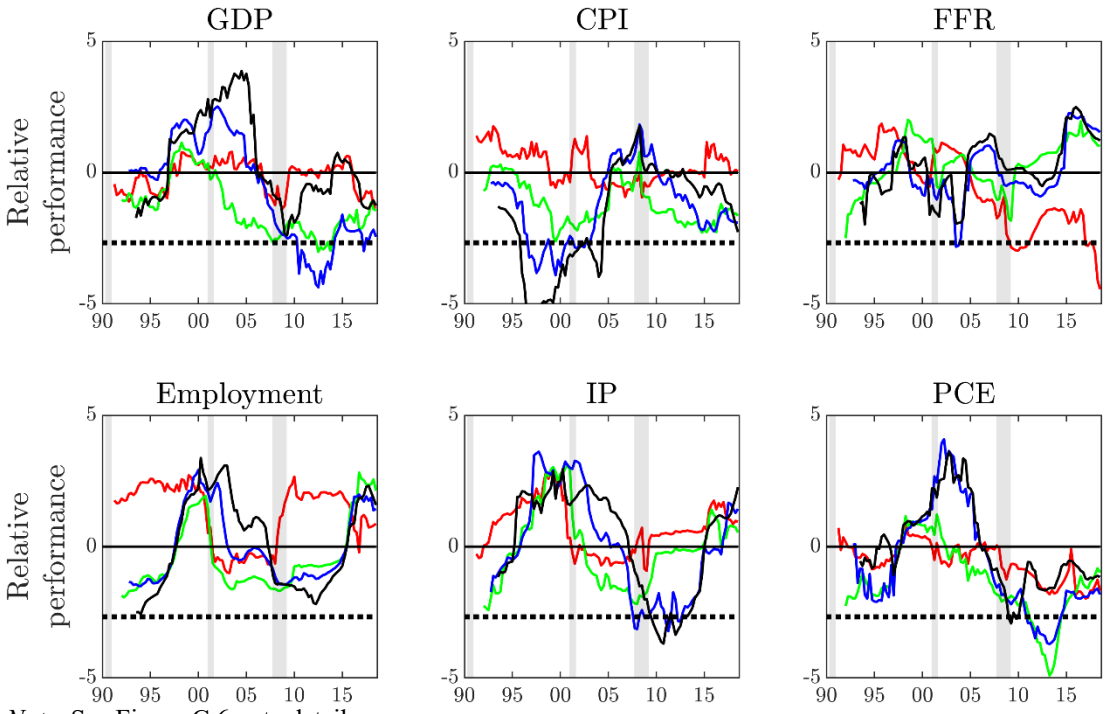
Note. See Figure C.6 note details.

Figure C.12: Fluctuation test of the multiscale BVAR-SSVS minus the benchmark



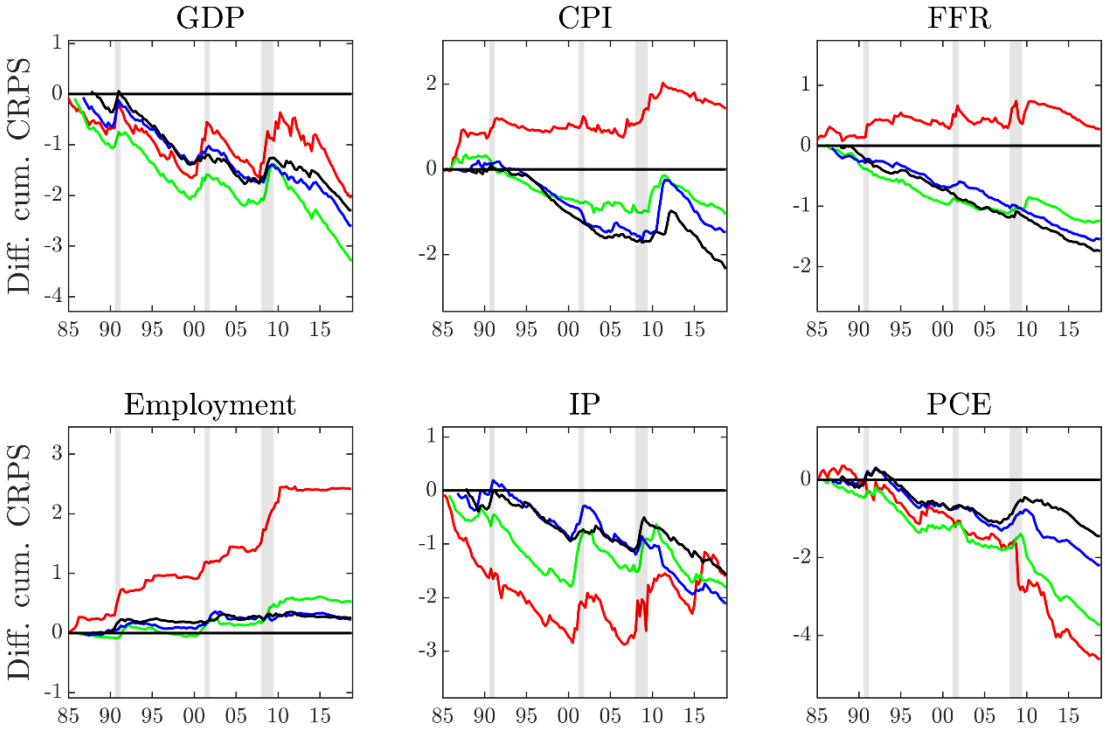
Note. See Figure C.6 note details.

Figure C.13: Fluctuation test of the multiscale BVAR-SSVS-SV minus the benchmark



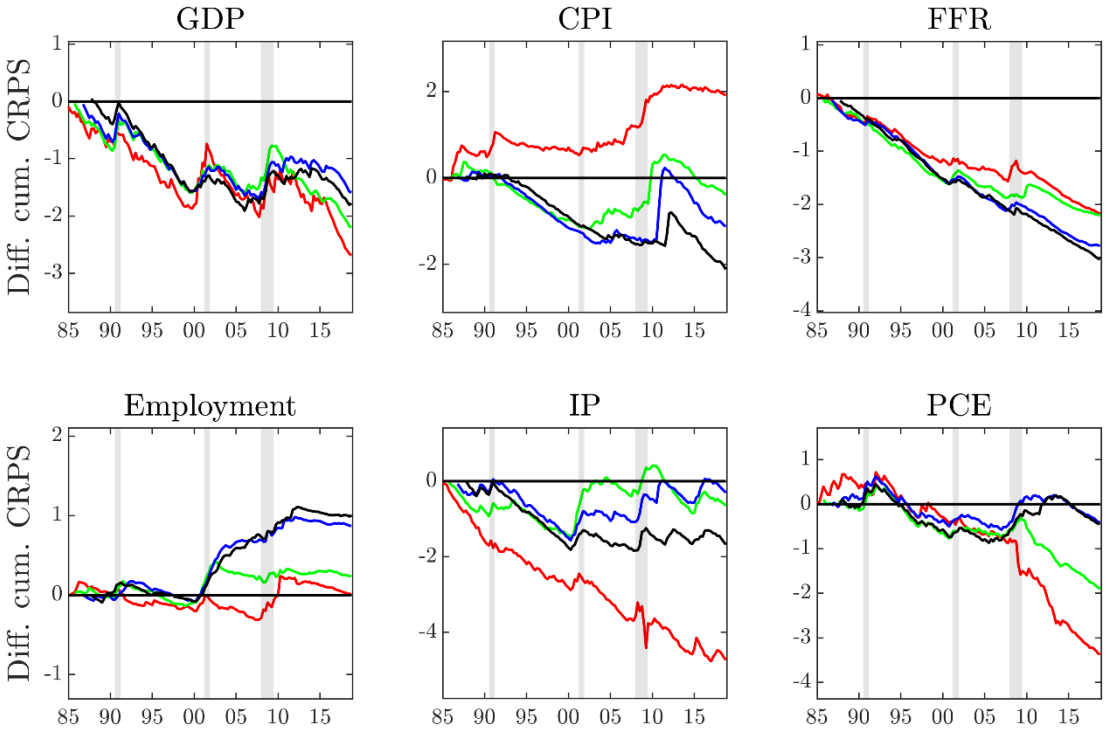
Note. See Figure C.6 note details.

Figure C.14: Difference of the cumulative CRPS of the universal hard threshold denoising BVAR minus the benchmark



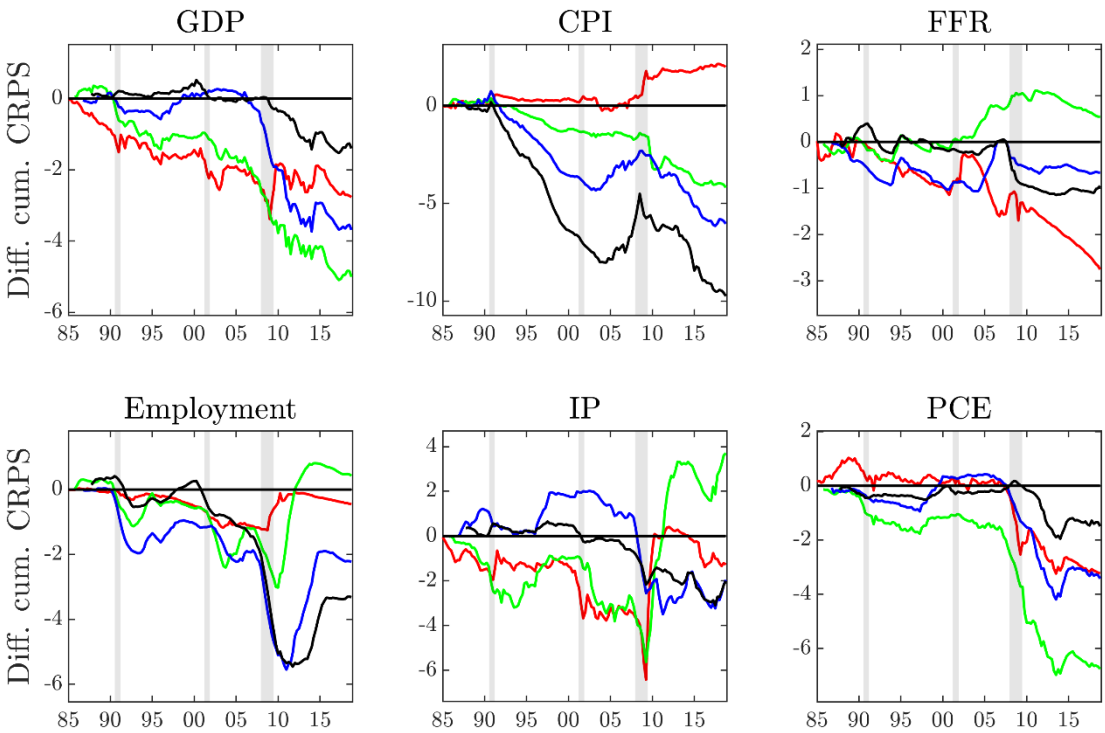
Note. Red, green, blue and black lines refer to 1, 4, 8, and 12-quarters ahead forecasts, respectively.

Figure C.15: Difference of the cumulative CRPS of the SURE soft threshold denoising BVAR minus the benchmark



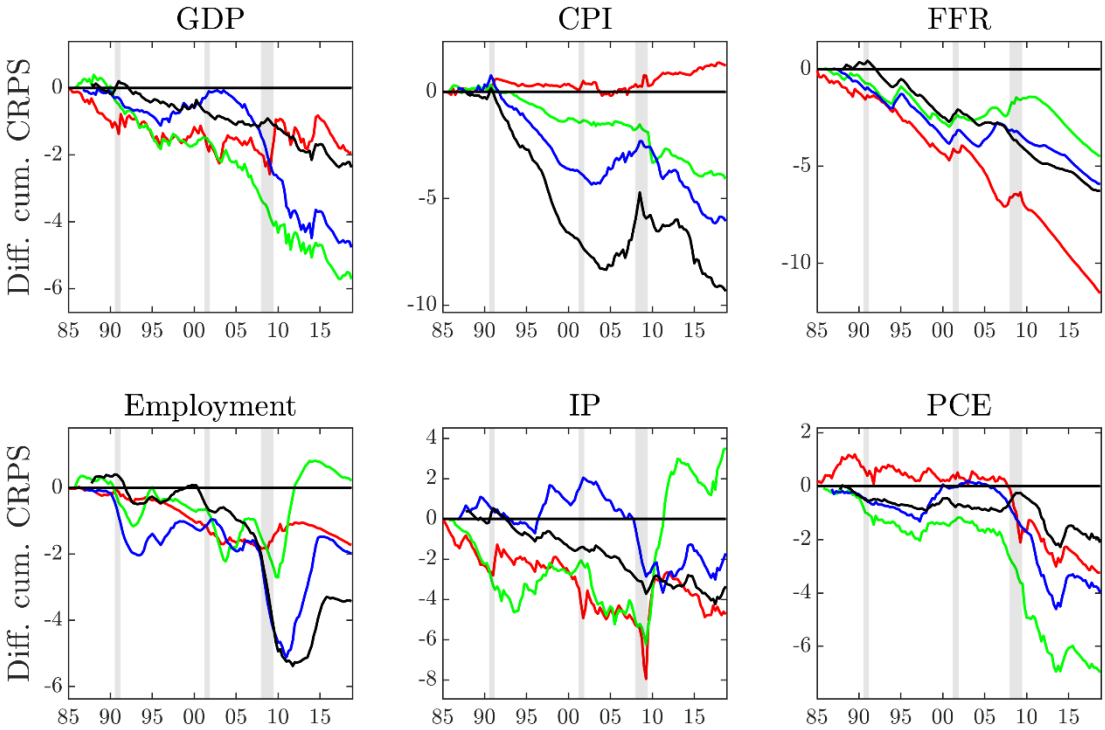
Note. See Figure C.14 note details.

Figure C.16: Difference of the cumulative CRPS of the MODWT separate scale BVAR minus the benchmark



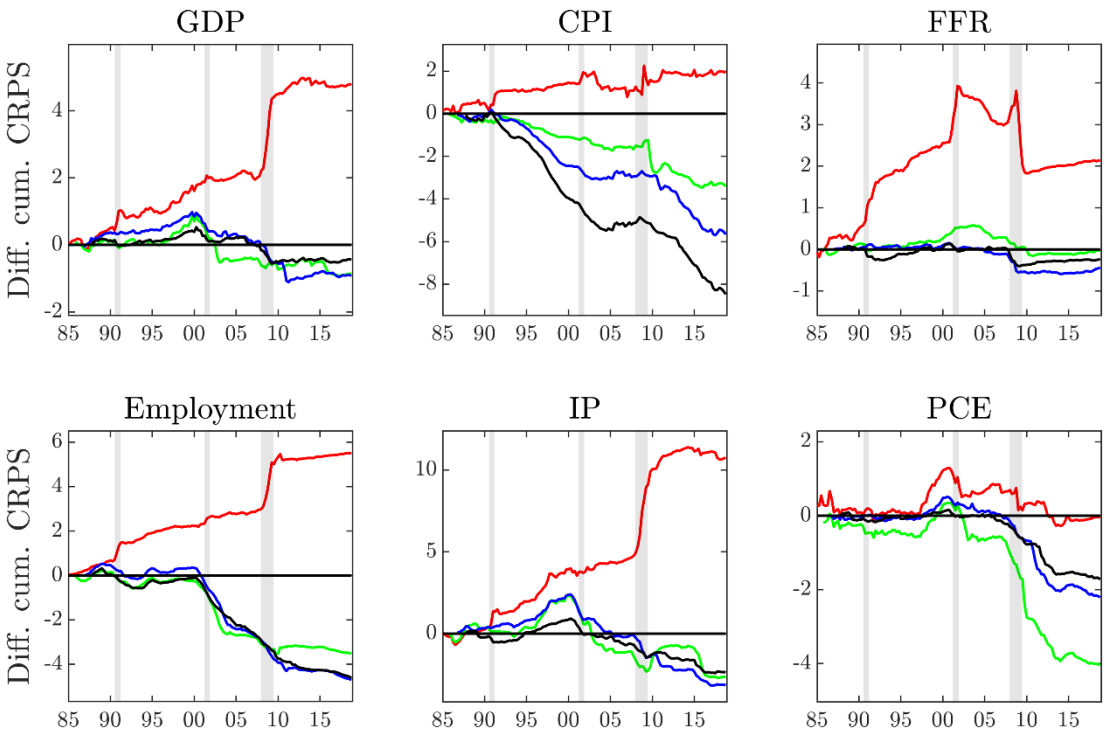
Note. See Figure C.14 note details.

Figure C.17: Difference of the cumulative CRPS of the MODWT separate scale BVAR/BVAR-SV minus the benchmark



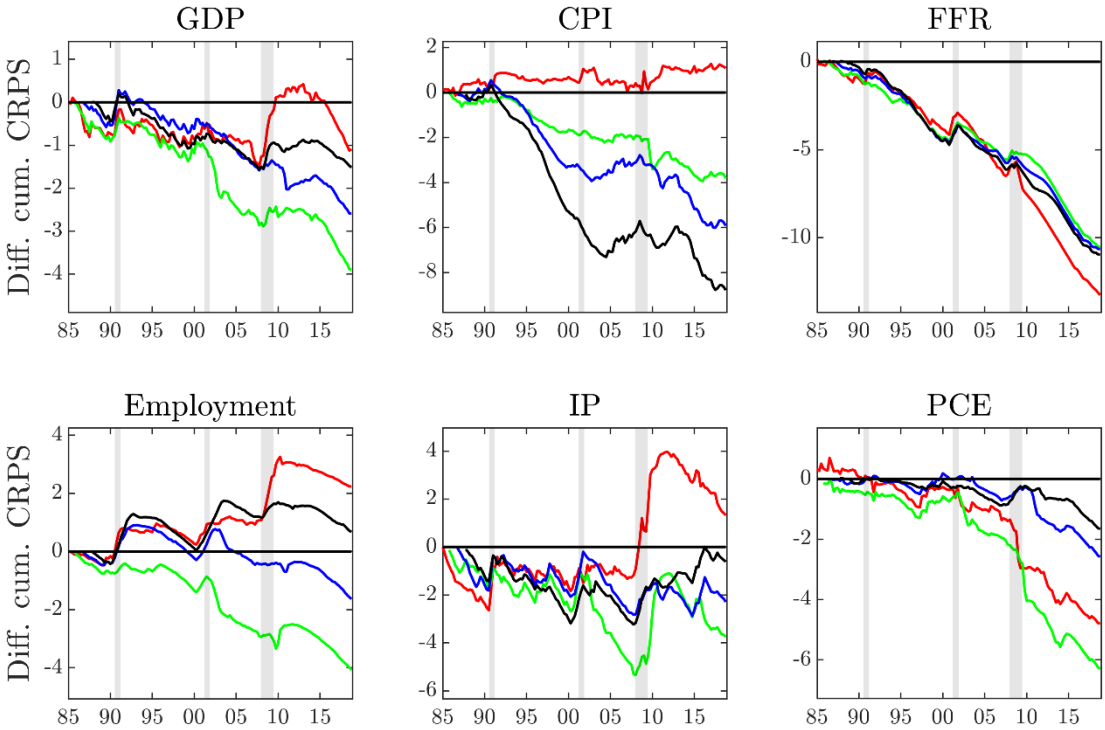
Note. See Figure C.14 note details.

Figure C.18: Difference of the cumulative CRPS of the multiscale BVAR-Minnesota minus the benchmark



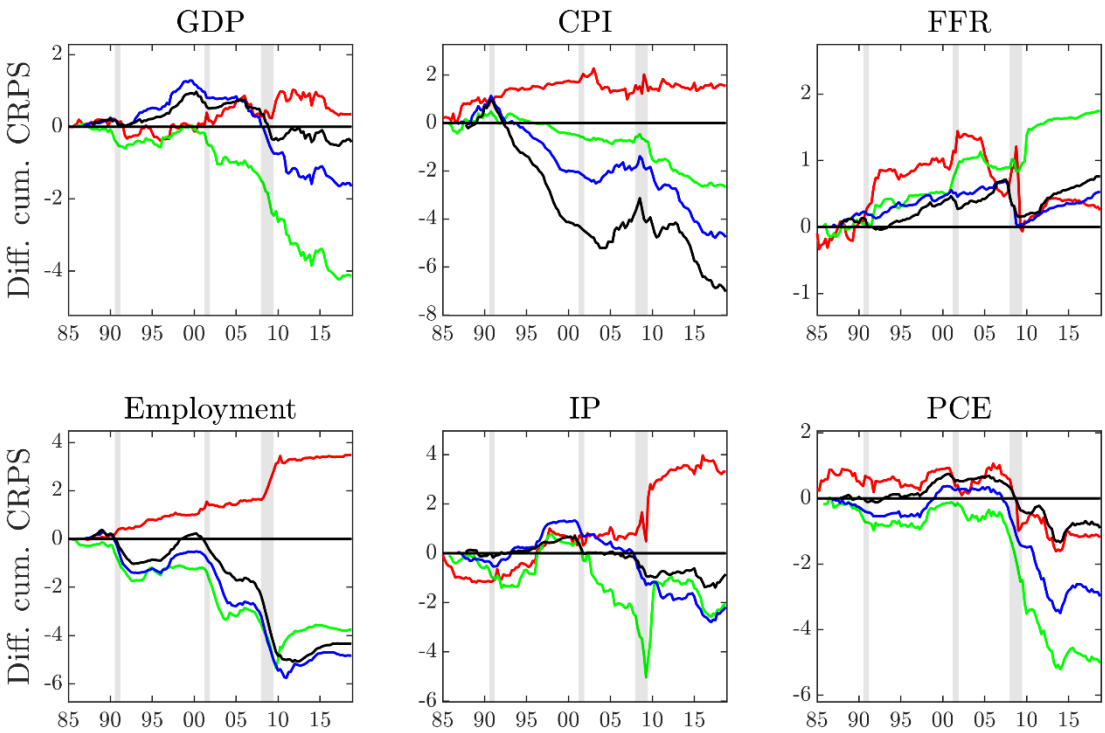
Note. See Figure C.14 note details.

Figure C.19: Difference of the cumulative CRPS of the multiscale BVAR-Minnesota-SV minus the benchmark



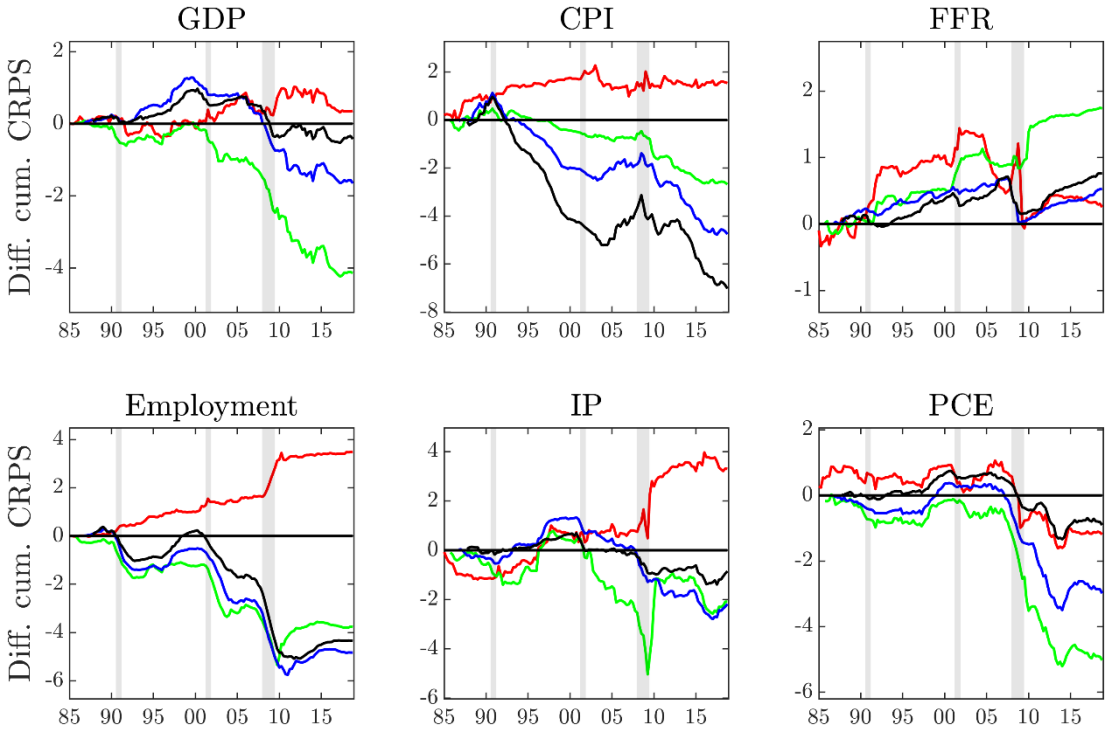
Note. See Figure C.14 note details.

Figure C.20: Difference of the cumulative CRPS of the multiscale BVAR-SSVS minus the benchmark.



Note. See Figure C.14 note details.

Figure C.21: Difference of the cumulative CRPS of the multiscale BVAR-SSVS-SV minus the benchmark



Note. See Figure C.14 note details.