DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

A mixed-frequency Bayesian vector autoregression with a steady-state prior

Sebastian Ankargren, Måns Unosson and Yukai Yang

CREATES Research Paper 2018-32

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

A mixed-frequency Bayesian vector autoregression with a steady-state prior*

Sebastian Ankargren†, Måns Unosson‡, and Yukai Yang†,§

†*Department of Statistics, Uppsala University, P.O. Box 513, 751 20 Uppsala, Sweden (e-mail: sebastian.ankargren@statistics.uu.se, yukai.yang@statistics.uu.se)*

‡*Department of Statistics, University of Warwick, Coventry, CV4 7AL, United Kingdom (e-mail: m.unosson@warwick.ac.uk)*

§*Center for Economic Statistics, Stockholm School of Economics, P.O. Box 6501, 113 83 Stockholm, Sweden*

Abstract

We consider a Bayesian vector autoregressive (VAR) model allowing for an explicit prior specification for the included variables' 'steady states' (unconditional means) for data measured at different frequencies. We propose a Gibbs sampler to sample from the posterior distribution derived from a normal prior for the steady state and a normal-inverse-Wishart prior for the dynamics and error covariance. Moreover, we suggest a numerical algorithm

*Special thanks to Johan Lyhagen for helpful comments on an earlier draft. Yang acknowledges support from Jan Wallander's and Tom Hedelius's Foundation, grants No. P2016-0293:1. We also wish to thank the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC 2015/6-117 for providing the necessary computational resources.

for computing the marginal data density that is useful for finding appropriate values for the necessary hyperparameters. We evaluate the proposed model by applying it to a real-time data set where we forecast Swedish GDP growth. The results indicate that the inclusion of high-frequency data improves the accuracy of low-frequency forecasts, in particular for shorter time horizons. The proposed model thus facilitates a simple and helpful way of incorporating information about the long run through the steady-state prior as well as about the near future through its ability to cope with mixed frequencies of the data.

JEL Classification numbers: C11, C32, C52, C53

Keywords: VAR, state space models, macroeconometrics, marginal data density, forecasting, nowcasting, hyperparameters.

I. Introduction

The vector autoregressive model (VAR) is a commonly used tool in applied macroeconometrics, in part motivated by its simplicity. Over the years, VAR models have developed in many different directions under both frequentist and Bayesian paradigms. The Bayesian approach offers the attractive ability to easily incorporate soft restrictions and shrinkage, which ameliorates the issue of overparametrization. Within the Bayesian framework itself, a large number of papers have developed prior distributions for the parameters in VAR models. Many of these are, in one way or another, variations of the Minnesota prior proposed by Litterman (1986) (see for example the book chapters Del Negro and Schorfheide, 2011; Karlsson, 2013). Gains in computational power have led to further alternatives in the choice of prior distribution as intractable posteriors can efficiently be sampled using Markov Chain Monte Carlo (MCMC) methods such as the Gibbs sampler (Gelfand and Smith, 1990; Kadiyala and Karlsson, 1997).

One of the Bayesian developments of the VAR model is the steady-state prior proposed by Villani (2009). It is based on a mean-adjusted form of the VAR where the unconditional

mean is explicitly parameterized. This seemingly innocuous reparametrization is motivated by the fact that practitioners and analysts often have prior information regarding the steady-state (or unconditional mean) readily available, e.g. inflation targeting by central banks. In the standard parametrization a prior on the unconditional mean is only implicit as a function of the other parameters' priors. Since the forecast in a stationary VAR converges to the unconditional mean, a prior for this parameter can help retaining the long run forecasts in the direction implied by theory, even if the model is estimated during a period of divergence.

In empirical macroeconomics, VARs have typically been hypothesized and estimated on a quarterly basis, see e.g. Adolfson et al. (2007); Stock and Watson (2001), which is related to the fact that many variables of interest are unavailable at higher frequencies. In the cases when some variables included are available at different frequencies, such as quarterly for macroeconomic and daily for financial data, the variables at higher frequency have traditionally been aggregated to the lowest frequency present.

The data aggregation incurs a loss of information for variables measured throughout the quarter: the aggregated quarterly values are typically sums or means of the constituent months, and any information carried by a within-quarter trend or pattern will be disregarded by the data aggregation. From a forecasting perspective an analyst will be unconsciously forced to disregard part of the information set when constructing a forecast from within a quarter as the most recent realizations are only available for the high-frequency variables. Another motivation for utilizing higher frequencies of the data is that the number of observations is increased. A VAR estimated on data collected over, say, ten years makes use of 120 observations of the monthly variables instead of being limited to the 40 aggregated quarterly observations.

Multiple approaches to dealing with the problem of mixed frequencies are available in the literature. Mixed data sampling (MIDAS) regression and MIDAS VAR proposed by Ghysels et al. (2007) and Ghysels (2016), respectively, use fractional lag polynomials to

regress a low-frequency variable on lags of itself as well as high-frequency lags of other variables. This approach is predominantly frequentist, although Bayesian versions are available (Ghysels, 2016; Rodriguez and Puggioni, 2010). A second approach, which is the focus of this work, is to exploit the ability of state-space modelling to handle missing observations (Harvey and Pierse, 1984). Eraker et al. (2015), concerned with Bayesian estimation, used this very idea to treat intra-quarterly values of quarterly variables as missing data and proposed measurement and state-transition equations for the monthly VAR. Schorfheide and Song (2015) considered forecasting using a construction along the lines of Carter and Kohn (1994) and provided empirical evidence that the mixed-frequency VAR (MF-VAR) improved forecasts of eleven US macroeconomic variables as compared to a quarterly VAR.

The main contribution of this paper is the proposal of a mixed-frequency steady-state Bayesian VAR, which effectively combines the steady-state parametrization of Villani (2009) with the state-space representation and filtering for mixed-frequency data of Schorfheide and Song (2015). The proposed model accommodates explicit modelling of the unconditional mean with data measured at different frequencies. In order to employ the model in a realistic forecasting situation, we construct a real-time data set consisting of Swedish macroeconomic data, which we use to forecast Swedish GDP growth. The combination of a steady-state prior and mixed-frequency data is found to be helpful as we see improved forecasting accuracy as compared to quarterly models as well as a mixed-frequency VAR without the steady-state prior. Moreover, we investigate the role of the hyperparameters and the empirical Bayes strategy for selection defined by maximizing the marginal data density at every forecast origin. The set of selected hyperparameters is relatively stable throughout the forecast evaluation period, whereby we can corroborate previous findings that a maximization approach is relatively close to an adequately fixed selection.

The structure of the paper is as follows. Section II describes the main methodology,

Section III develops an estimator for the marginal data density and Section IV gives an illustrative application forecasting Swedish GDP growth. Section V concludes.

II. Combining a mixed-frequency vector autoregression with steady-state beliefs

The mixed-frequency method adopted in this work is a state space-based model which follows the work by Eraker et al. (2015); Mariano and Murasawa (2010); Schorfheide and Song (2015). There are several modelling approaches available for handling mixed-frequency data, including MIDAS (Ghysels et al., 2007), bridge equations (Baffigi et al., 2004) and factor models (Giannone et al., 2008; Mariano and Murasawa, 2003). We do not review these further here, but instead refer the reader to the survey by Foroni and Marcellino (2013) and the comparison by Kuzin et al. (2011).

State space representation of the mixed-frequency model

To cope with mixed observed frequencies of the data, we assume the system to be evolving at the highest available frequency, which implies that many high-frequency observations for low-frequency variables are simply missing data. By doing so, the approach naturally lends itself to a state-space representation of the system, in which the underlying monthly series of the quarterly variables become the latent states of the system.

Let $z_t = (z'_{m,t}, z'_{q,t})'$ denote the underlying high-frequency vector in the system, consisting of n_m monthly and n_q quarterly variables. Note that the time t here takes the highest frequency, i.e. monthly. Furthermore, we denote by y_t what is observed at time t. The empirical problem that is often present is that what is observed varies over time such that the dimension n_t of y_t is not always equal to $n = n_m + n_q$.

The observed data y_t is generally supposed to be some linear aggregate of $Z_t = (z'_t, \dots, z'_{t-p+1})'$

such that

$$y_t = \begin{pmatrix} y_{m,t} \\ y_{q,t} \end{pmatrix} = \begin{pmatrix} I_{n_m} & 0 \\ 0 & M_{q,t} \end{pmatrix} \begin{pmatrix} I_{n_m} & 0 \\ 0 & \Lambda_q \end{pmatrix} Z_t = M_t \Lambda Z_t, \tag{1}$$

where $M_{q,t}$ and Λ_q are deterministic selection and aggregation matrices, respectively.

We let $M_{q,t}$ be the n_q identity matrix I_{n_q} if all quarterly variables are observed at time t so that $y_{q,t} = \Lambda_q Z_t$. In the remaining periods, $M_{q,t}$ is an empty matrix such that $y_t = y_{m,t}$. More complicated observational structures can easily be accomodated using the very same approach; instead of being empty or a full I_n matrix, M_t can have rows deleted which correspond to unobserved variables. This idea is briefly revisited later in Section II when discussing ragged edges.

The aggregation matrix Λ_q represents the assumed aggregation scheme of unobserved high-frequency latent observations $z_{q,t}$ into occasionally-observed low-frequency observations $y_{q,t}$. We employ a quarterly average such that if t is the final month of a quarter, then $y_{q,t} = \frac{1}{3}(y_{q,t} + y_{q,t-1} + y_{q,t-2})$. It is, however, possible to use other schemes (see e.g. Mariano and Murasawa, 2010).

To enable modelling despite the variation in the observational structure, a model is assumed for the underlying high-frequency variable. More specifcally, a VAR(p) for z_t is employed such that

$$\Pi(L) z_t = \Phi d_t + u_t, \quad u_t \sim N_n(0, \Sigma), \tag{2}$$

where $\Pi(L) = (I_n - \Pi_1 L - \Pi_2 L^2 - \cdots - \Pi_p L^p)$ is a p-th order invertible lag polynomial, d_t is an $m \times 1$ vector of deterministic components and Φ is an $n \times m$ matrix of parameters.

The model in (2) is a conventional VAR specification, but, in the spirit of Villani (2009),

we instead employ the mean-adjusted form as

$$\Pi(L)(z_t - \Psi d_t) = u_t, \quad u_t \sim N_n(0, \Sigma), \tag{3}$$

where $\Psi = [\Pi(L)]^{-1}\Phi$, if it is stationary. It can be readily confirmed that $\mathrm{E}(z_t) = \Psi d_t := \mu_t$, and thus μ_t is the unconditional mean—*steady state*—of the process. The steady-state representation (3) requires an explicit prior on the steady state parameters. However, common practice applies a loose prior on Φ in (2), which implicitly defines an intricate (but loose) prior on Ψ and, subsequently, μ_t. We argue that in many applications, the parametrization in (3) is more convenient as it allows for a more natural elicitation of prior beliefs. In what follows, we will extend the work of Villani (2009) such that (3) may still constitute a viable option in the presence of mixed frequencies.

We build on the work by Schorfheide and Song (2015) to set up a Gibbs sampling procedure in conjunction with simulation smoothing in a state-space framework which makes it possible to sample from the posterior distribution of the parameters. The approach rests on the previously established notion that low-frequency series are aggregates of unobservable high-frequency series. The aggregation equation in (1) and the high-frequency model in (3) constitute the measurement and state equations, respectively, summarized as

$$y_t = M_t \Lambda Z_t, \tag{4}$$

$$Z_{t+1} = W_{t+1}\psi + F(\Pi)(Z_t - W_t\psi) + \varepsilon_t, \tag{5}$$

$$\varepsilon_t \sim N(0, \Omega(\Sigma)),$$

where $W_t = [(d_t \otimes I_p), \ldots, (d_{t-p+1} \otimes I_p)]'$ and $\psi = \mathrm{vec}(\Psi)$. The model is now written in

companion form, where

$$\Pi = (\Pi_1, \ldots, \Pi_p), \quad F(\Pi) = \begin{pmatrix} \Pi & \\ I_{n(p-1)} & 0_{n(p-1)\times p} \end{pmatrix}, \quad \Omega(\Sigma) = \begin{pmatrix} \Sigma & 0_{n\times n(p-1)} \\ 0_{n(p-1)\times n} & 0_{n(p-1)} \end{pmatrix}.$$

We assume here that the aggregation requires no more than p lags. If the aggregation scheme at time t depends on lags beyond $t-p$ it is possible to simply append blocks of zeros to $F(\Pi)$ without changing the model itself (with corresponding changes to $\Omega(\Sigma)$).

As an example, consider a bivariate VAR model with three lags and one monthly and one quarterly variable in which the quarterly variable is observed at the last month of each quarter. Using the intra-quarter average as the aggregation scheme,

$$y_t = \begin{pmatrix} z_{m,t} \\ \frac{1}{3}(z_{q,t} + z_{q,t-1} + z_{q,t-2}) \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{M_t} \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \end{pmatrix}}_{\Lambda} \begin{pmatrix} z_t \\ z_{t-1} \\ z_{t-2} \end{pmatrix},$$

if $t \in \{\text{Mar, Jun, Sep, Dec}\}$. Thus, whenever t corresponds to an end-of-quarter month, $M_t\Lambda$ relates the monthly variables in Z_t to the observables y_t appropriately. When t does not correspond to an end-of-quarter month, M_t in the above display is instead $M_t = (1, 0)$ and thus simply selects the monthly variable.

Incorporating prior beliefs

We consider a normal prior for the parameters in Ψ and a normal-inverse Wishart as a joint prior for the VAR coefficients and error covariance. Thus, the prior used is

$$(\Pi, \Sigma) \sim MNIW(\underline{\Pi}, \underline{\Omega}_\Pi, \underline{S}, \underline{\nu}), \tag{6}$$

such that

$$\Sigma \sim IW(\underline{S}, \underline{\nu}), \quad \text{vec}(\Pi')|\Sigma \sim N_{n^2p}(\text{vec}(\underline{\Pi}'), \Sigma \otimes \underline{\Omega}_{\Pi}).$$

The main diagonal of $\underline{\Omega}_{\Pi}$ is set to be

$$\underline{\omega}_{ii} = \frac{\lambda_1^2}{(l^{\lambda_2} s_r)^2} \text{ for lag } l \text{ of variable } r,\ i = (l-1)p + r$$

where λ_1 is the overall tightness and λ_2 determines the lag decay rate; the inclusion of s_r adjusts for differences in measurement scale of the variables. A more thorough exposition of the normal inverse Wishart prior is given by Karlsson (2013). In Section III we discuss how to estimate the marginal data density, which is used in Section E to select λ_1 and λ_2 by maximization of the marginal data density.

Finally, we follow Villani (2009) and let the prior for the unconditional mean be normal,

$$\psi = \text{vec}(\Psi) \sim N_{nm}(\underline{\psi}, \underline{\Omega}_{\Psi}).$$

Sampling from the posterior distribution

In order to sample from the intractable posterior distribution of latent variables and parameters given the data, $p(\Pi, \Sigma, \psi, Z|Y)$, a Gibbs sampler is applied here which decomposes the posterior into three blocks of full conditional densities which is easy to sample from. Mathematical details concerning the posterior distributions can be found in the Supplementary material, Appendix C, whereas additional information regarding the simulation smoothing technique used is available in Appendix D.

The three blocks that compose the Gibbs sampler are

$$p(Z|\Pi, \Sigma, \psi, Y), \quad p(\Pi, \Sigma|\psi, Z), \quad \text{and} \quad p(\psi|\Pi, \Sigma, \psi, Z),$$

where it can be observed that the parameters (Π, Σ) and ψ are independent of Y given

Z. Conditional on the parameters, the unobservables can be sampled using a simulation smoother (Durbin and Koopman, 2002, 2012). The Kalman filter is initialized by conditioning on the first p observations, where any missing observations are replaced by the most recent observation. Given this initialization, the simulation smoother can be applied using the mean-adjusted processes $y_t^* = y_t - M_t \Lambda W_t \psi$ and $Z_t^* = Z_t - W_t \psi$ in (4)–(5) and then adding $W_t \psi$ to the resulting draws of Z_t^*.

The MNIW prior for (Π, Σ) is conjugate for the Gaussian likelihood, and thus the conditional posterior is in the same family of distributions by standard results (Karlsson, 2013). Similarly, the conditional posterior of ψ derived by Villani (2009) appears in the same fashion while also conditioning on the unobservables. Thus, the conditional posterior of ψ is normal.

Forecasting with ragged edges

In real-time forecasting, publication delays generally cause the available information sets to possess ragged edges for both single- and mixed-frequency data sets. The simplest way to handle these ragged edges is to use as final period in the sample the most recent time point at which all variables are observed, denoted by T^*, effectively discarding observations from time periods with incomplete data. This, however, is inefficient as it does not make use of all the available information. A second approach is to forecast conditional on the observations that do exist at $t > T^*$, which can be done in numerous ways. Within our framework, this is easily accomplished by simply treating the missing observations at $t = T^* + 1, \ldots, T$ as regular missing data, as also suggested by Bańbura et al. (2015); Schorfheide and Song (2015). Thus, by adjusting the selection matrix M_t accordingly at the ragged-edge time points, we can also make draws from the posterior distribution of the missing high-frequency variables. More specifically, if $z_{m,t}$ is missing at time t, the procedure simply amounts to dropping the row of M_t that corresponds to this variable.

III. Estimation of the marginal data density

Since the various high-dimensional prior distributions that are popular in the literature are usually parameterized by a low-dimensional vector of hyperparameters, it is of great importance to choose these auxiliary parameters appropriately. A crude way is to rely on what has become default values. In fact, many authors resort to an overall tightness of $\lambda_1 = 0.2$ and a lag decay of $\lambda_2 = 1$ (for examples, see Canova, 2007; Carriero et al., 2015a; Villani, 2009). As applications vary, it is natural to believe that also the hyperparameters may need to change.

Multiple approaches that aid in the selection of hyperparameters exist, among which some of the more prominent methods include using hierarchical prior distributions or by maximization of the marginal data density (MDD). The former is e.g. studied by Giannone et al. (2015), who treat the vector λ of hyperparameters as additional parameters and specify a prior for these parameters, yielding a hierarchical prior $p(\theta|\lambda)p(\lambda)$. As remarked by the authors, if a flat prior for λ is specified, then the posterior distribution of the hyperparameters, $p(\lambda|y)$, is proportional to the marginal data density. Thus, the second approach entails selecting values of the hyperparameters that maximize the MDD, as these also maximize the posterior of the hyperparameters under a flat hyperprior. This route—an empirical Bayes approach—is the one we choose, and was also taken by e.g. Carriero et al. (2012); Schorfheide and Song (2015).

An estimator of the marginal data density

The MDD is not analytically tractable under the modelling situation described in Section II, but can be estimated using the improved Chib (1995) estimator proposed by Fuentes-Albero and Melosi (2013).

The quantity of interest to estimate is the MDD, which is

$$p(Y|\lambda) = \int p(Y, \Pi, \Sigma, \psi, Z|\lambda) d(\Pi, \Sigma, \psi, Z).$$

In slight abuse of notation, in what follows we omit the dependence on the hyperparameters.

The method is a refinement of Chib (1995) insofar as the existence of an analytical expression for $p(\Pi, \Sigma|\psi, Z, Y)$ is exploited, which reduces the need for two reduced Gibbs steps to only one. The idea is to decompose the MDD as

$$p(Y) = \frac{p(Y|\Pi, \Sigma, \psi)p(\Pi, \Sigma)}{p(\Pi, \Sigma|\psi, Y)} \frac{p(\psi)}{p(\psi|Y)}.$$

Fuentes-Albero and Melosi (2013) suggest to evaluate the terms analytically—if possible—at some measure of centrality (i.e. posterior mode, median or mean); when not possible, numerical approximations are necessary. Let p denote a known density and $\hat{p}$ one which is estimated in a sense that will be made precise, and let $\tilde{A}$ denote a matrix with elements being the posterior means of the respective elements of A. The MDD is estimated by

$$\hat{p}(Y) = \frac{p(Y|\tilde{\Pi}, \tilde{\Sigma}, \tilde{\psi})p(\tilde{\Pi}, \tilde{\Sigma})}{\hat{p}(\tilde{\Pi}, \tilde{\Sigma}|\tilde{\psi}, Y)} \frac{p(\tilde{\psi})}{\hat{p}(\tilde{\psi}|Y)},$$

where $p(Y|\tilde{\Pi}, \tilde{\Sigma}, \tilde{\psi})$ is the data likelihood, $p(\tilde{\Pi}, \tilde{\Sigma})$ is the prior for (Π, Σ), and $p(\tilde{\psi})$ is the prior for ψ, with all three terms evaluated at the posterior centers. The two denominator terms require numerical approximations, which is accomplished by a reduced Gibbs step and the Rao-Blackwellization technique (Gelfand et al., 1992), respectively. More specifically, we let

$$\hat{p}(\tilde{\Pi}, \tilde{\Sigma}|\tilde{\psi}, Y) = \frac{1}{R} \sum_{i=1}^{R} p(\tilde{\Pi}, \tilde{\Sigma}|\tilde{\psi}, Z^{(i)}, Y), \tag{7}$$

where $Z^{(i)}$ are draws from $p(Z|\tilde{\psi}, Y)$. The marginal posterior $p(\psi|Y)$ is estimated using

draws from the original Gibbs sampler as

$$\hat{p}(\tilde{\psi}|Y) = \frac{1}{R}\sum_{i=1}^{R} p(\tilde{\psi}|\Pi^{(i)}, \Sigma^{(i)}, Z^{(i)}, Y).$$

IV. Using real-time data to forecast Swedish GDP growth

In this section, we assess the forecasting ability of the model that we propose. The assessment is carried out by checking its out-of-sample predictive accuracy based on the Swedish quarterly GDP growth data. The quarterly steady-state Bayesian VAR model has been applied in several previous studies, see for example, Adolfson et al. (2007); Clark (2011); Iversen et al. (2016); Österholm (2008); Villani (2009). The model is a small-scale macroeconomic VAR model for Swedish data including GDP growth, unemployment rate, CPI inflation, industrial production index and the economic tendency indicator. The economic tendency indicator is the main indicator published in the National Institute of Economic Research's (NIER) Economic Tendencies Survey. All series, except the forecasting target GDP growth, are available monthly.

Data

We construct a real-time data set by combining available data from Statistics Sweden, OECD and the National Institute of Economic Research (NIER). From Statistics Sweden we collect real-time vintages of real GDP, of which we take log-differences to obtain GDP growth. The OECD's main economic indicators archive contains real-time data on the harmonized unemployment rate, the consumer price index (CPI) as well as an index of in-

TABLE 1
Summary of the real-time data set

Series	Transformation	Source	Frequency
GDP growth	$\ln \Delta$	Statistics Sweden*	Quarterly
Harmonized unemployment rate	None	OECD MEI†	Monthly
Consumer price index	$\ln \Delta$	OECD MEI†	Monthly
Index of industrial production	$\ln \Delta$	OECD MEI†	Monthly
Economic tendency indicator	(0, 1)	NIER‡	Monthly

Sources:
* Working-day and seasonally adjusted GDP in constant prices
† OECD's Main economic indicators (MEI) revisions analysis database
‡ The (quasi-)real-time data made available by Billstam et al. (2016)

dustrial production (IP).[1] We leave the unemployment rate as it is, but take log-differences of also CPI and IP. Finally, we retrieve the economic tendency indicator (ETI) from the National Institute of Economic Research, which recently published a (quasi-)real-time data set that includes the ETI. We standardize the series to have mean and variance $(0, 1)$ instead of $(100, 100)$. Table 1 contains a summary of the data used.

Real-time data

In constructing a real-time forecasting scenario, the goal is to have data which mirror exactly what the forecaster had available in the corresponding time period. The publication of the monthly vintages by Statistics Sweden of GDP and OECD of its main economic indicators and the attempt by Billstam et al. (2016) to create a real-time dataset for the NIER's Economic Tendencies Survey make it possible to create a situation which resembles the reality to a high degree. In the application, we focus on end-of-month forecasting and thus do not treat mid-month publications any differently from publications on the final day of the month.

The ETI is constructed based on surveys to households and business in Sweden and is

[1]OECD also provides data for Swedish GDP using both constant and current prices. However, for the series using constant prices, the reported series was not seasonally adjusted over the period 2000M10–2007M02. For this reason, we instead turn to Statistics Sweden to obtain a GDP series which is seasonally adjusted over the entire time span.

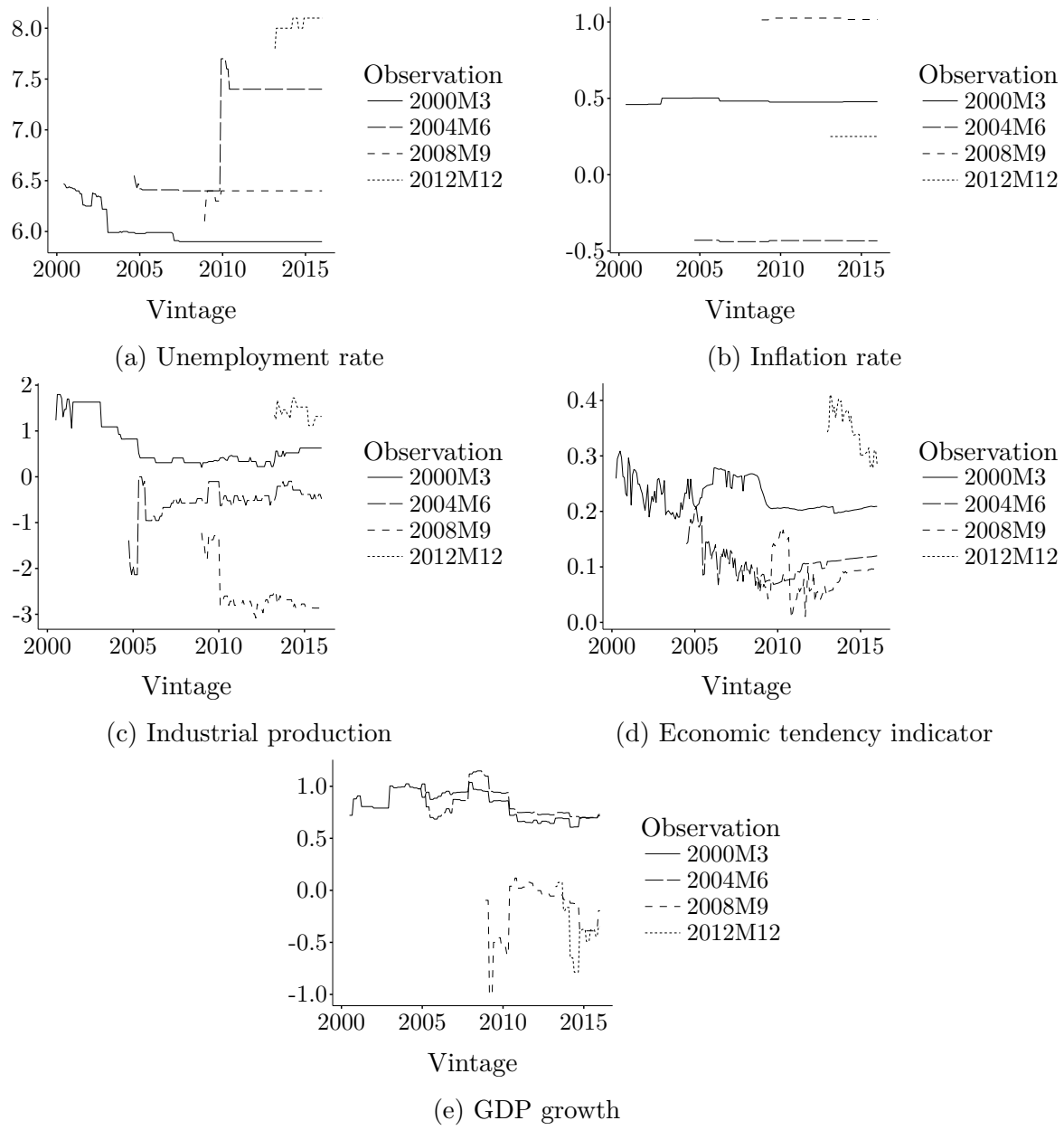
published as an index with mean and variance standardized to be equal to 100. The raw data underlying the ETI is typically not revised, with the exception of correcting apparent errors. In order to construct a quasi-real-time dataset, Billstam et al. (2016) note that it involves taking the necessary raw data seasonally adjusted and standardized, with the appropriate series being weighted altogether and then re-standardized. The dataset is thus referred to as 'quasi' for mainly two reasons: first, it is based on today's methods for standardization and weighting, and second, it may contain corrections of errors. However, Billstam et al. (2016) argue that for evaluating out-of-sample forecast performance, 'the quasi-real-time data should ... be close to a perfect substitute to actual real-time data'.

Figure 1 displays the revision tendencies for four arbitrary observations from March, June, September and December in 2000, 2004, 2008 and 2012, respectively. As the figure illustrates, some of the series are subject to larger revisions than others, occasionally exhibiting large jumps.

Publication delays

Figure 2 displays the structure of publication delays for the five series throughout the sample period. For the monthly variables, the delay is in general consistent over time, with unemployment and inflation generally being published within two months, industrial production within three and ETI in the concurrent month. The delay for GDP growth varies between 2 and 5 months.

The missing cells in the publication delay for the unemployment rate is caused by a lack of vintage data during this period in the OECD database. As a proxy in our data set we take the first new publication and use this to impute the missing vintages by assuming a two-month publication delay throughout the period with missing data.



(a) Unemployment rate

(b) Inflation rate

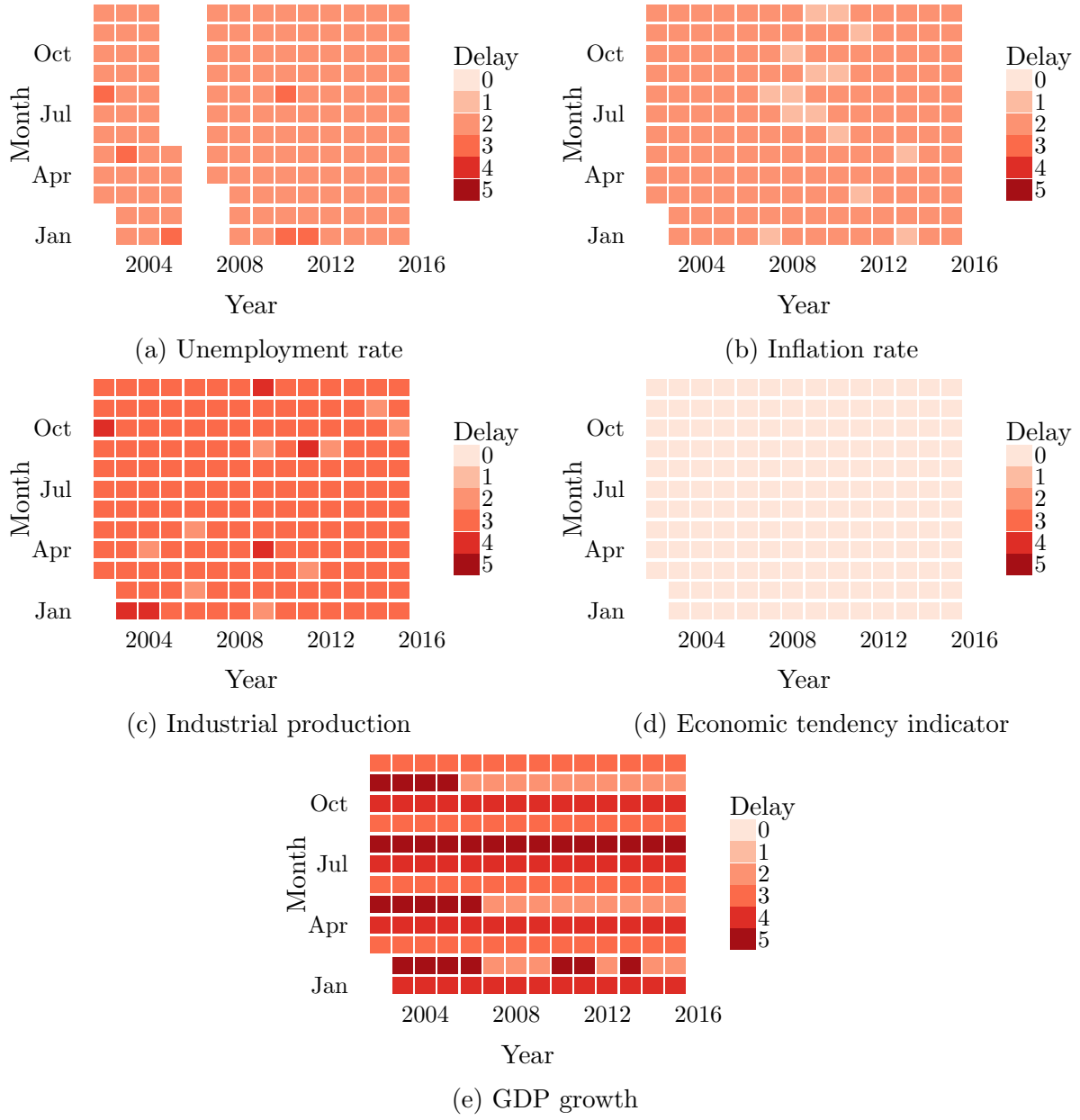
(c) Industrial production

(d) Economic tendency indicator

(e) GDP growth

Figure 1. Revision tendencies

Notes: The figures display how observations change across vintages for four fixed time points.



(a) Unemployment rate

(b) Inflation rate

(c) Industrial production

(d) Economic tendency indicator

(e) GDP growth

Figure 2. Publication delays

Notes: Each cell represents one month and its color corresponds to the number of months since the most recent observation was published. The delay is computed end of month; a zero-period delay means that the observation is available at the end of the current month. The missing cells in (a) stem from temporary non-publication of vintages (see text for more information).

Forecasting setup and evaluation

We consider a forecasting situation similar to that studied by Schorfheide and Song (2015). We forecast GDP growth 0–8 quarters ahead at the end of every month, where the 0-step forecast denotes the forecast of the current quarter. Because of publication lags and mixed frequencies of the data, the available information varies and most notably so depending on the relative position of the month within the quarter.

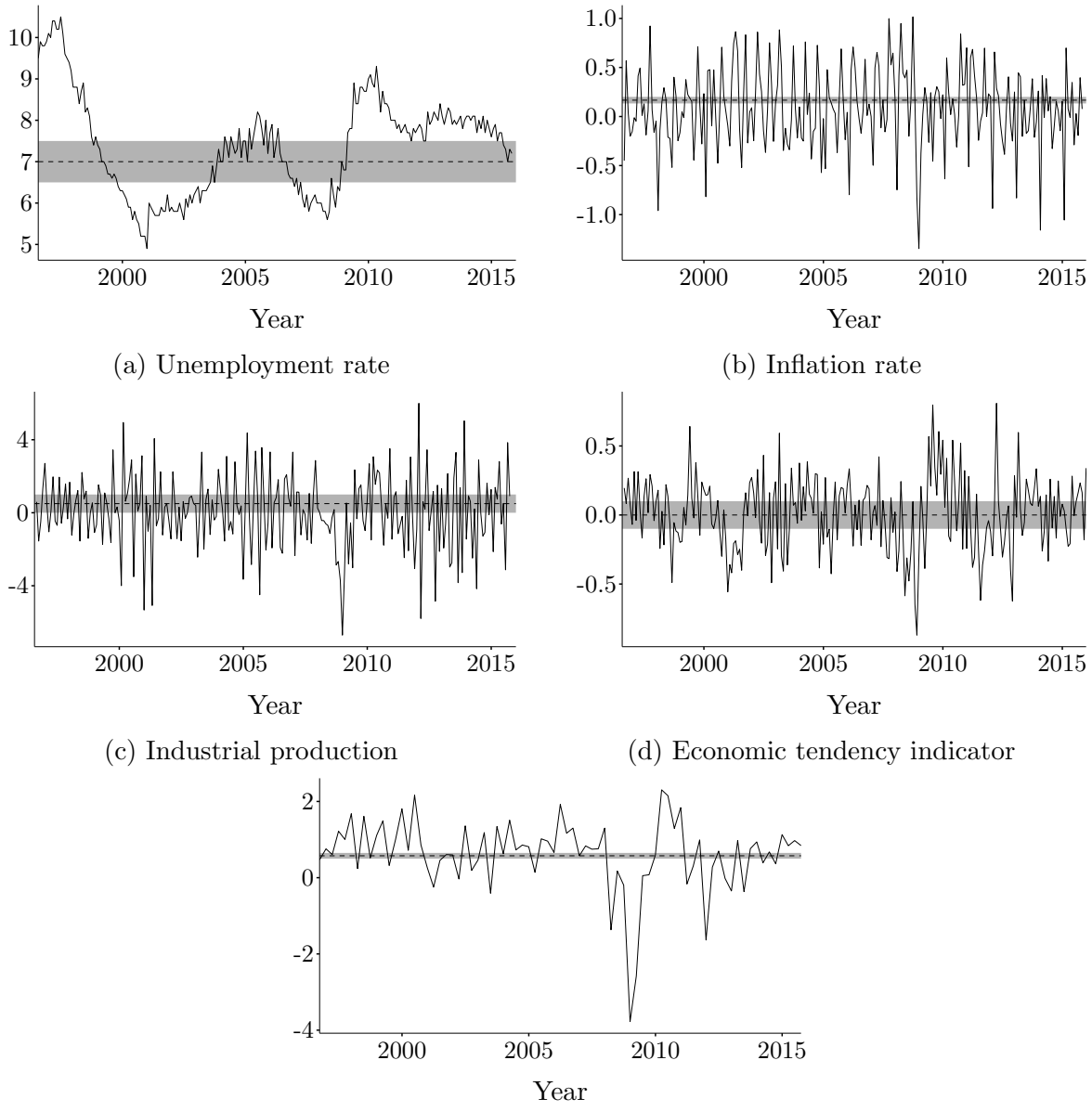
To be able to gauge the relative performance of the MF-BVAR with a steady-state prior (abbreviated by MF-SS), we also include the MF-BVAR with a Minnesota prior (MF-Minn), as well as quarterly-frequency versions with both priors (QF-SS and QF-Minn, respectively).[2] For the mixed-frequency models, we employ the ragged-edge forecasting approach discussed in Section II, whereas the quarterly models are estimated and forecasted using complete quarters. All models use a lag length of $p = 4$.

In the application of the Gibbs sampler to numerically approximate the posterior distribution, we make 20,000 draws for each run and keep the final 15,000. We do so for a recursively expanding estimation window, where the first forecast is made in January 2004 and the final in November 2015. We select the hyperparameters using adaptive grid search; see Appendix E for more information.

Steady-state prior

As for the steady-state prior, these are presented visually in Figure 3. Where possible, we keep largely in line with previous studies (see e.g. Ankargren et al., 2017; Österholm, 2010; Villani, 2009).

[2]The implementation of the Minnesota prior is a standard implementation whose prior for dynamic coefficients and the error covariance is the same as described in Section II and Equation (6) in particular.



(a) Unemployment rate

(b) Inflation rate

(c) Industrial production

(d) Economic tendency indicator

(e) GDP growth

Figure 3. Steady-state priors

Notes: The shaded areas in the figures correspond to 95 % prior probability intervals of the variables, with the dashed line showing the prior mean.

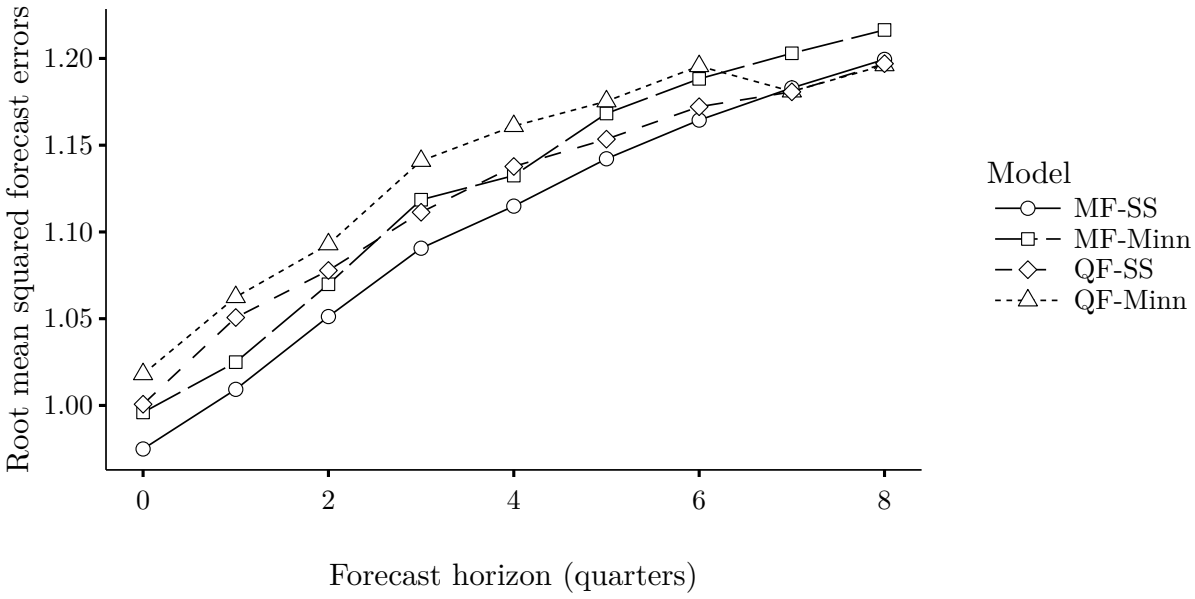


Figure 4. Root mean squared forecast errors by forecast horizon and model

Forecasting performance

To evaluate the forecasting ability, we consider both point and density forecasts.

Point forecasts

We start by comparing the point forecasts with respect to the root mean squared forecast error (RMSFE) in Figure 4.

The figure clearly shows that the MF-SS model performs better in the short to middle horizons and is caught up with by QF-SS and QF-Minn in the long horizon at two years. Interestingly, both of the MF models perform well for short horizons, but after that MF-Minn is closer to its quarterly counterpart. It is worth noting that the results display the same relative ordering as previous studies: Villani (2009) finding QF-SS to outperform QF-Minn, and Schorfheide and Song (2015) demonstrating that MF-Minn performed better than QF-Minn in the short run. Thus, the results indicate that there is additional merit in combining the mixed-frequency model with a steady-state prior. Overall, although the differences are moderate, the results suggest that MF-SS is to be preferred.

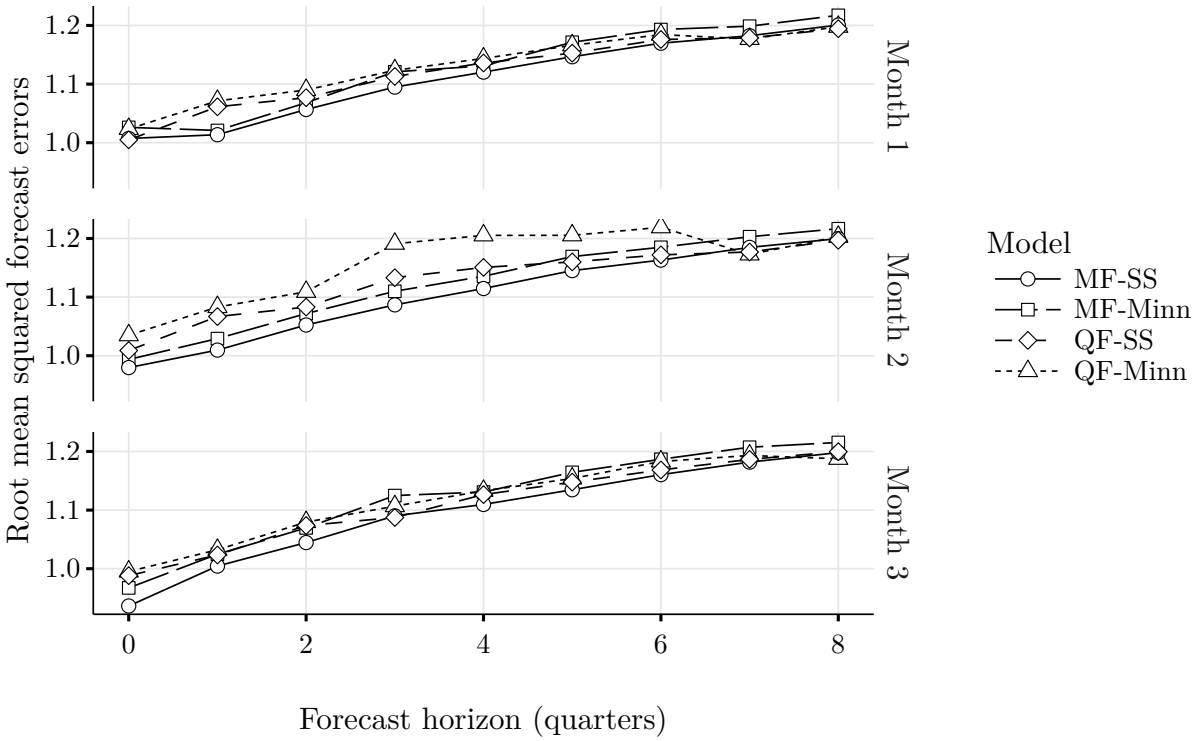


Figure 5. Root mean squared forecast errors by forecasting horizon, within-quarter origin and model

Breaking the results down by each forecast origin's within-quarter position, the picture remains largely unchanged in pattern, as is shown in Figure 5.

MF-SS is dominating in each group, but the difference compared to MF-Minn in particular is often negligible. Overall, no drastic differences are present between the within-quarter forecast origins. However, the value of recent publications can be seen by the fact that the nowcasting ability improves with the month of origin within the quarter.

In order to see how the relative performance has evolved over time, Figure 6 shows the cumulative RMSFE. Interestingly enough, in the pre-crisis period the Minnesota-based models exhibits smaller RMSFEs than the steady-state models, while in the post-crisis period the mixed-frequency models start to outperform the quarterly ones.

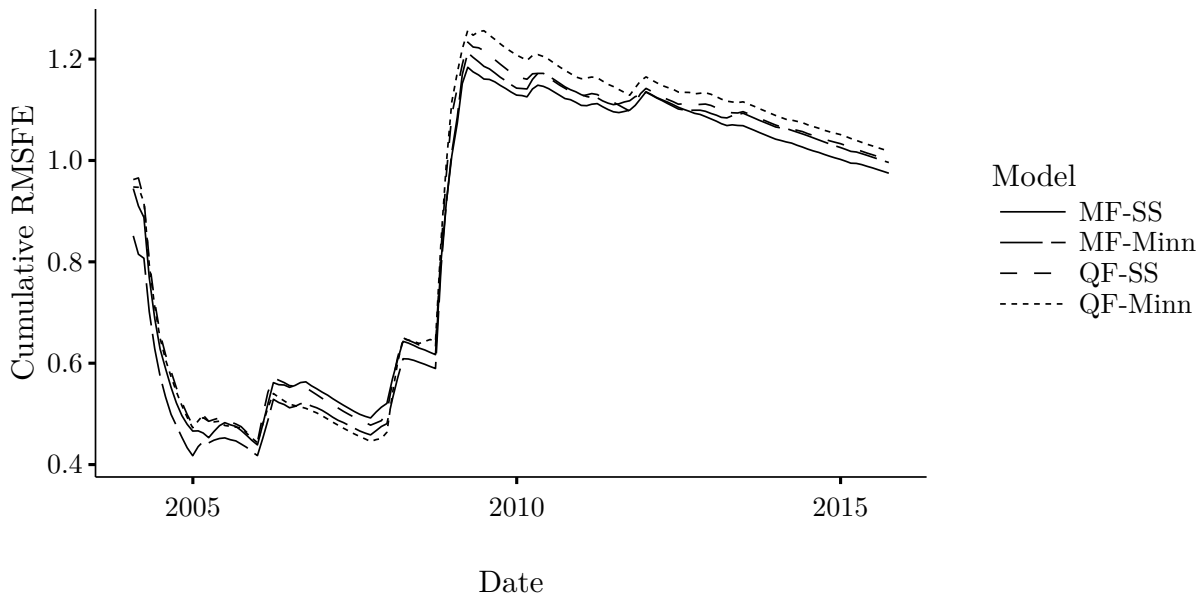


Figure 6. Evolvement of nowcast (0-step) root mean squared forecast errors over the evaluation period by model

Density forecasts

For density forecasts, we compute the probability integral transform $z_t = \int_{-\infty}^{y_t} p_t(u)du$, where p_t is the predictive density and y_t the outcome of GDP growth (Diebold et al., 1998). Using the MCMC draws, we approximate the transform by $z_{t+h} = R^{-1} \sum_{r=1}^{R} I(y_{t+h} < \hat{y}_{t+h|t}^{(r)})$, where $\hat{y}_{t+h|t}^{(r)}$ denotes the h-step ahead forecast of GDP growth at time t in iteration r. If the predictive density coincides with the true, the sequence $\{z_{t+h}\}$ is a dependent sequence of variates with marginal distribution $U(0, 1)$.

Histograms for z_{t+h} are provided in Figure 7, where the horizontal line corresponds to the bin height that would be if the transforms were indeed $U(0, 1)$ variables. None of the models perform strikingly well with the performance deteriorating with h. MF-SS and MF-Minn appear to do a decent job for $h = 0$ and less so for $h = 1$ and discouragingly worse for the long-run forecasts.

Finally, the interval forecasts are evaluated by computing the coverage rates of the predictive intervals. That is, for a nominal coverage of $100(1 - \alpha)\%$ the corresponding

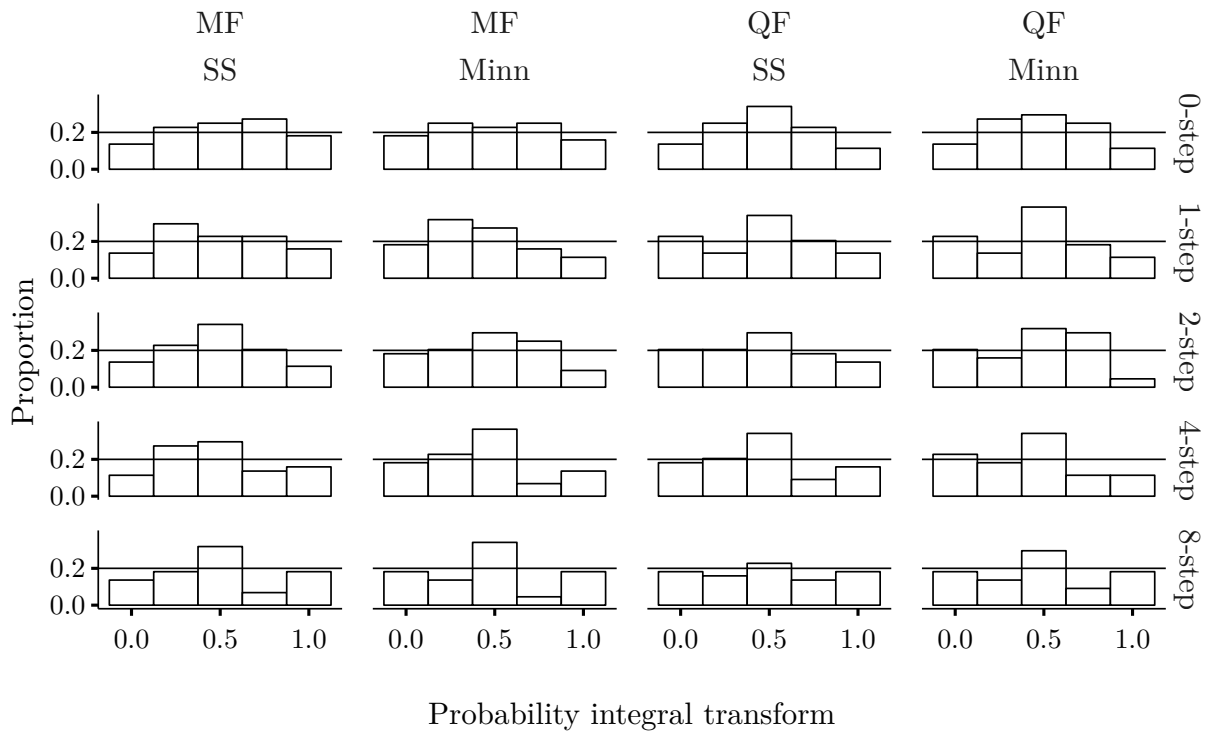


Figure 7. Histograms of probability integral transformations for forecasts of GDP growth
Notes: The solid line represents the expected bin height under a uniform distribution.

interval is computed and we then average over hits and misses in the evaluation period to obtain the empirical coverage rate. Figure 8 plots the nominal rates against the empirical.

The mixed-frequency models again show somewhat better results for short horizons, as they tend to be closer to the diagonal line. For the nowcast, there is some distortion for intervals with higher nominal coverage, but this disappears for the 1-step forecast. The QF-SS model tends to have too high empirical coverage for lower nominal coverage levels, but too low empirical coverage for higher nominal. For the 4-step and 8-step forecasts, all models exhibit this pattern to some degree.

The role of hyperparameter selection

The previous section relies on an empirical Bayes strategy for selecting hyperparameters, and at this point it is warranted to ask: what is the role of the hyperparameters for the models' forecasting performance? Previous studies in this regard include Carriero et al. (2015a, 2012); Giannone et al. (2015). Carriero et al. (2012) conduct a grid search for the overall tightness λ_1 in a large Bayesian VAR used for forecasting bond yields, whereas Carriero et al. (2015a) systematically study specification choices in Bayesian VARs, including hyperparameter selection by maximizing the marginal data density. Giannone et al. (2015), on the other hand, conduct a fully Bayesian analysis and employ a hierarchical model in which priors are assigned to the hyperparameters. Both Carriero et al. (2012) and Carriero et al. (2015a) find that the selection tends to be stable over time and that the advantages compared with using a fixed set of hyperparameters is minimal; similarly, Schorfheide and Song (2015) found the selection to be stable and resorted to using fixed values. However, the main advantage of maximizing the marginal data density lies in the approach being a principled and transparent way. Additionally, the specific set of hyperparameters which yields a good forecasting performance may not be obvious; the optimal level of shrinkage is intimately tied to the dimension of the model, as discussed by Bańbura et al. (2010); Carriero et al. (2012); De Mol et al. (2008). Marginal data density

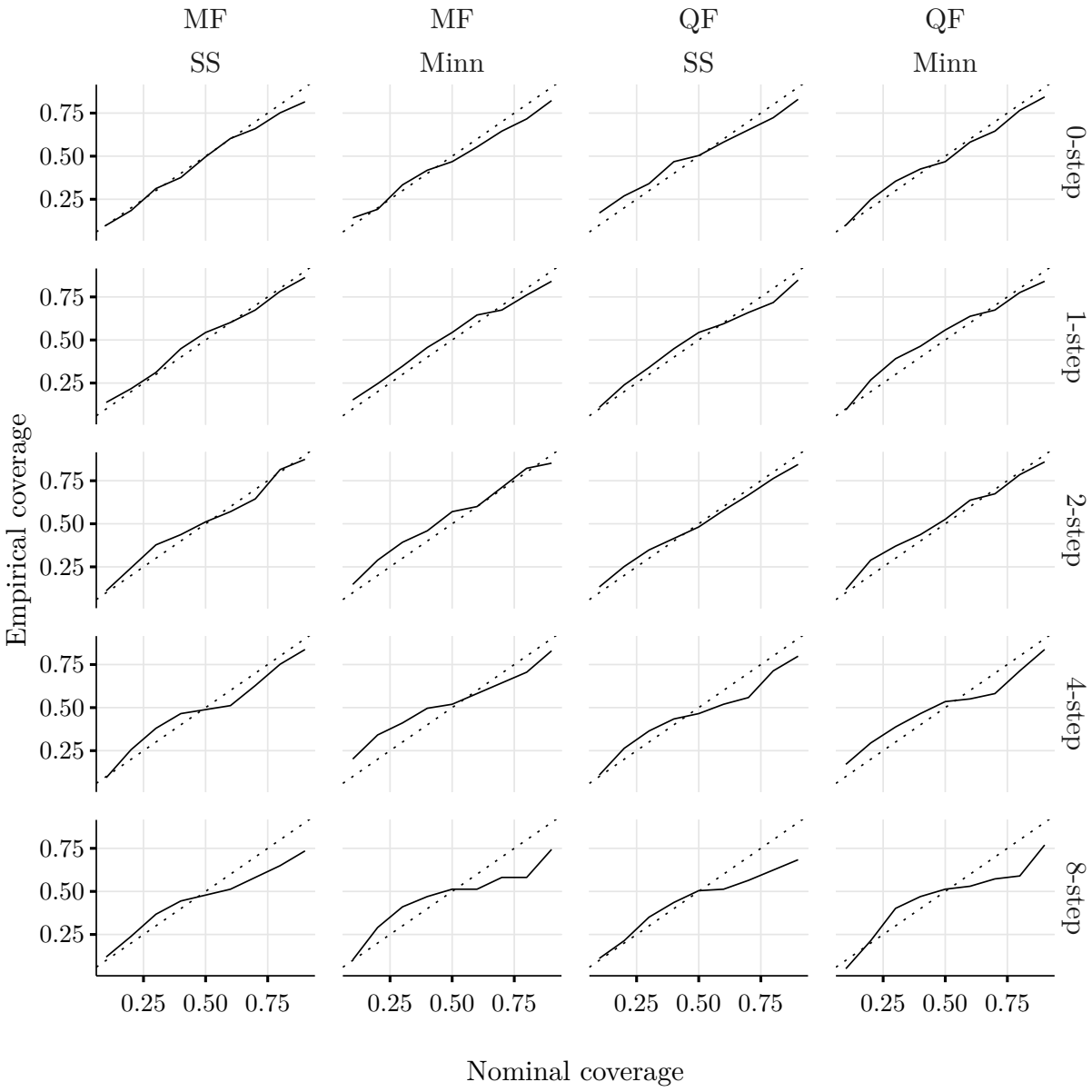


Figure 8. Coverage rates of prediction intervals
Notes: The dashed line represents an empirical rate equal to the nominal.

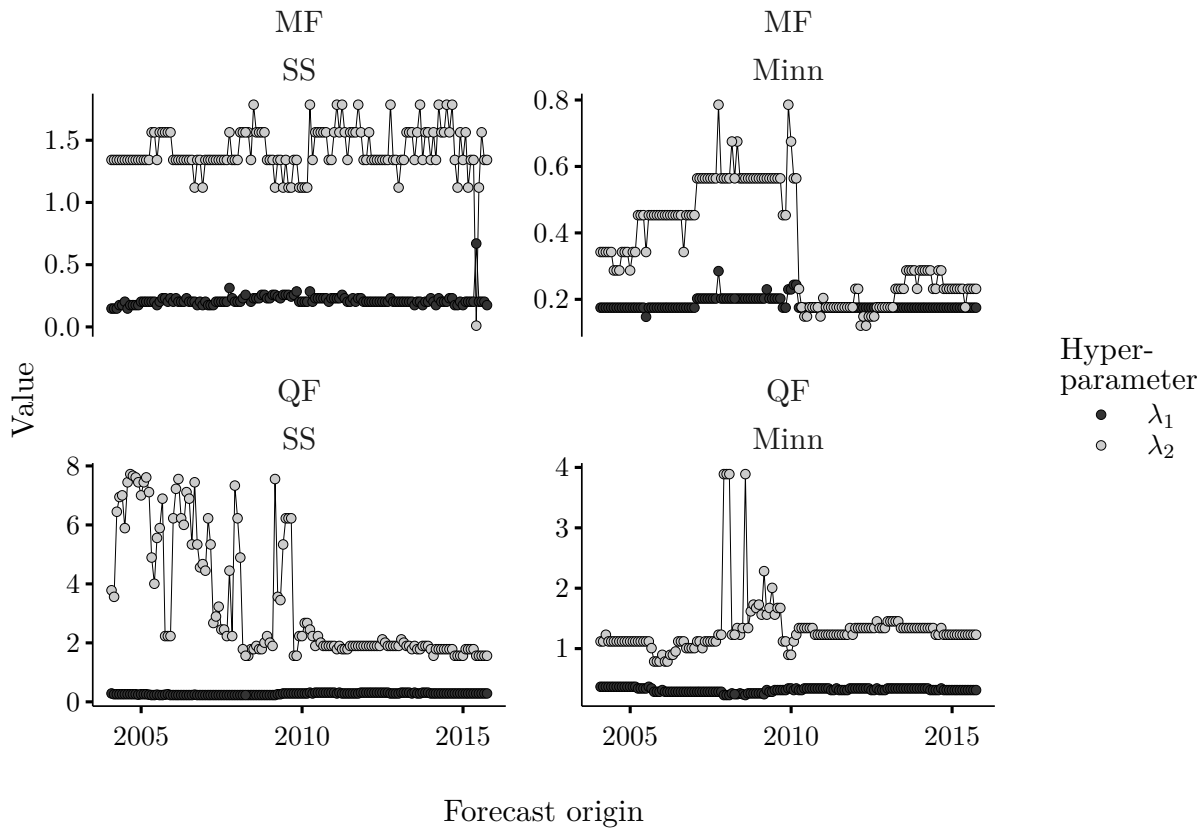


Figure 9. Time series plots of selected values of the hyperparameters
Notes: λ_1 controls the overall tightness and λ_2 the lag decay. The selected value for overall tightness is stable over time, while the chosen lag decay is more variable. The selection stabilizes in the second half of the evaluation sample.

maximization can be used as a means of identifying appropriate hyperparameter values.

Figure 9 illustrates the trajectories of selected hyperparameters throughout the evaluation sample for all four models considered. It is somewhat striking that the selected value of the overall tightness parameter hovers around 0.2–0.3 showing little variability in all four panels. The value of the selected lag decay parameter varies to a larger extent, yet seems to stabilize when the sample period extends beyond 2009–10. The hyperparameter values the quarterly models stabilize around—$\lambda_1 = 0.2$ and $\lambda_2 = 1$ or $\lambda_2 = 2$—are exactly the values discussed by Canova (2007) as default values that generally work well. In the case of the mixed-frequency models, a less tight prior is selected for the Minnesota-based model, whereas the model with a steady-state prior selects a lag decay around 1.5. Thus, fixing the

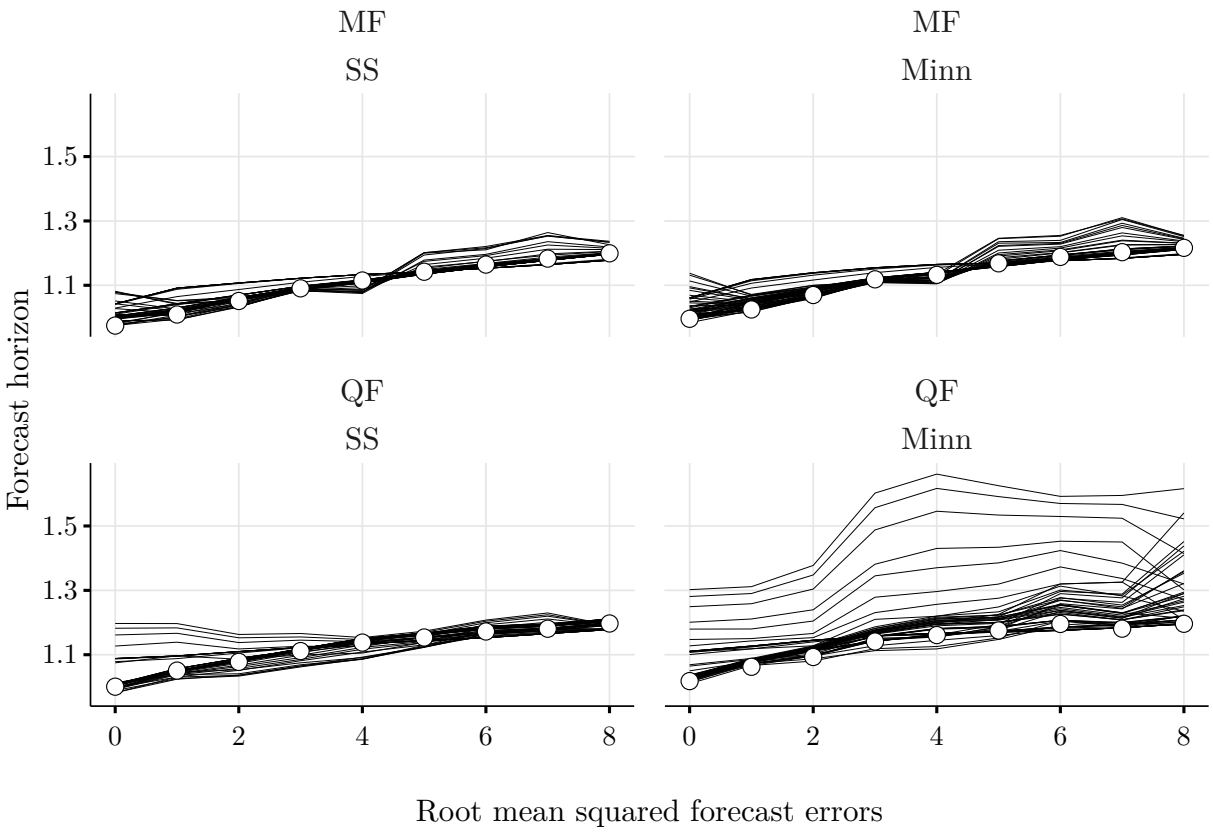


Figure 10. Forecasting performance for different combinations of hyperparameter values
Notes: Each line represents a unique combination of (λ_1, λ_2) from the first step of the adaptive grid search (see Section E). The points represent the corresponding root mean squared forecast errors using the maximizers of the marginal data density. The maximizing approach generally performs well, but not necessarily the best at each horizon.

hyperparameters to the values the selection approach stabilizes at will likely yield a similar performance. However, the figure shows that what these values are varies across model and prior configurations. Finally, for larger models additional shrinkage is anticipated to be warranted, as demonstrated by Bańbura et al. (2010); Carriero et al. (2012); De Mol et al. (2008).

The differences in forecasting performance with respect to the choice of hyperparameter values is shown in Figure 10, where lines correspond to one of the 49 combinations of

(λ_1, λ_2) used in the first step of the adaptive grid search (see Section E).[3] The forecasting accuracy among the hyperparameter combinations included vary greatly between configurations. For the quarterly Minnesota model, some hyperparameter values result in very poor performance, whereas for the mixed-frequency model with a steady-state prior the differences are relatively small for all horizons. In general, selecting hyperparameters based on maximization of the marginal data density appears to, in some sense, be a robust strategy. Poor hyperparameter values are avoided in all cases, but the best performance at each horizon is not achieved. Nevertheless, the maximizing pair appears to offer a decent balance and overall forecasting ability, as some of the fixed hyperparameter combinations forecast well for some horizons but relatively worse for others (e.g. the lines initially below the circles in the MF-SS pane eventually cross the circled line indicating poorer performance).

V. Conclusion

In this paper we present a mixed-frequency vector autoregressive model estimated using Bayesian methods, which incorporates prior beliefs about the steady states—the unconditional means—of the included variables. Previous literature has already established that there is value in using mixed-frequency data and avoiding temporal aggregation for forecasting purposes and this finding is also presented in our results for forecasting Swedish GDP growth in a real-time data set. Additionally, Villani (2009) demonstrated the virtue of a steady-state prior in the single-frequency case and we find that this improves forecasts also in the mixed-frequency model.

We also revisit the question of the role of hyperparameter selection. In our application,

[3]The figure only includes the root mean squared forecast errors (RMSFE) from the first step of the grid search since some of the values present in the second or third steps only occur once or a couple of times. Thus, their RMSFE values would be based on a single or a handful of forecasts and as such would be associated with a large amount of uncertainty. The included lines are all based on the same number of forecasts.

we take an empirical Bayes approach and select hyperparameters which maximize the marginal data density. The main conclusion is that the set of selected hyperparameters shows a limited degree of variability over the evaluation period, thus indicating that a fixed set of hyperparameters will perform similarly. However, maximizing the marginal data density is a transparent and principled way and can, at the very least, be used to find appropriate values to fix the hyperparameters at in the sequel.

On the downside, none of the evaluated models—quarterly and mixed-frequency VARs with Minnesota or steady-state priors—demonstrate adequate density forecasting abilities for horizons beyond the very short term. Studies such as Clark (2011), Carriero et al. (2015b) and Carriero et al. (2016) suggest that incorporating stochastic volatility can be helpful for density forecasting. As a result, developing mixed-frequency Bayesian VAR models which allow for more flexibility of the innovation variance is on our current research agenda.

References

Adolfson, M., Andersson, M. K., Lindé, J., Villani, M., and Vredin, A. (2007). 'Modern forecasting models in action: Improving macroeconomic analyses at central banks'. *International Journal of Central Banking*, 3, pp. 111–144.

Ankargren, S., Bjellerup, M., and Shahnazarian, H. (2017). 'The importance of the financial system for the real economy'. *Empirical Economics*, 53, pp. 1553–1586.

Baffigi, A., Golinelli, R., and Parigi, G. (2004). 'Bridge models to forecast the Euro area GDP'. *International Journal of Forecasting*, 20, pp. 447–460.

Bańbura, M., Giannone, D., and Lenza, M. (2015). 'Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections'. *International Journal of Forecasting*, 31, pp. 739–756.

Bańbura, M., Giannone, D., and Reichlin, L. (2010). 'Large Bayesian vector auto regressions'. *Journal of Applied Econometrics*, 25, pp. 71–92.

Billstam, M., Frändén, K., Samuelsson, J., and Österholm, P. (2016). 'Quasi-real-time data of the Economic Tendency Survey'. Working Paper No. 14, National Institute of Economic Research.

Canova, F. (2007). 'Bayesian VARs'. In *Methods for Applied Macroeconomic Research*, chapter 10, pp. 351–394. Princeton University Press, Princeton.

Carriero, A., Clark, T. E., and Marcellino, M. (2015a). 'Bayesian VARs: Specification choices and forecast accuracy'. *Journal of Applied Econometrics*, 30, pp. 46–73.

Carriero, A., Clark, T. E., and 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, pp. 837–862.

Carriero, A., Clark, T. E., and Marcellino, M. (2016). 'Large vector autoregressions with stochastic volatility and flexible priors', Federal Reserve Bank of Cleveland.

Carriero, A., Kapetanios, G., and Marcellino, M. (2012). 'Forecasting government bond yields with large Bayesian vector autoregressions'. *Journal of Banking and Finance*, 36, pp. 2026–2047.

Carter, C. K. and Kohn, R. (1994). 'On gibbs sampling for state space models'. *Biometrika*, 81, pp. 541–553.

Chib, S. (1995). 'Marginal likelihood from the Gibbs output'. *Journal of the American Statistical Association*, 90, pp. 1313–1321.

Clark, T. E. (2011). 'Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility'. *Journal of Business & Economic Statistics*, 29, pp. 327–341.

De Mol, C., Giannone, D., and Reichlin, L. (2008). 'Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?'. *Journal of Econometrics*, 146, pp. 318–328.

Del Negro, M. and Schorfheide, F. (2011). 'Bayesian macroeconometrics'. In Geweke, J., Koop, G., and van Dijk, H., editors, *The Oxford Handbook of Bayesian Econometrics*, pp. 293–389. Oxford University Press, Oxford.

Diebold, F. X., Gunther, T., and Tay, A. (1998). 'Evaluating density forecasts with applications to financial risk management'. *International Economic Review*, 39, pp. 863–883.

Durbin, J. and Koopman, S. J. (2002). 'A simple and efficient simulation smoother for state space time series analysis'. *Biometrika*, 89, pp. 603–615.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, UK, second edition.

Eraker, B., Chiu, C. W., Foerster, A. T., Kim, T. B., and Seoane, H. D. (2015). 'Bayesian mixed frequency VARs'. *Journal of Financial Econometrics*, 13, pp. 698–721.

Foroni, C. and Marcellino, M. (2013). 'A survey of econometric methods for mixed-frequency data'. Working Paper No. 6, Norges Bank.

Fuentes-Albero, C. and Melosi, L. (2013). 'Methods for computing marginal data densities from the Gibbs output'. *Journal of Econometrics*, 175, pp. 132–141.

Gelfand, A. E. and Smith, A. F. M. (1990). 'Sampling-based approaches to calculating marginal densities'. *Journal of the American Statistical Association*, 85, pp. 398–409.

Gelfand, A. E., Smith, A. F. M., and Lee, T.-m. (1992). 'Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling'. *Journal of the American Statistical Association*, 87, pp. 523–532.

Ghysels, E. (2016). 'Macroeconomics and the reality of mixed frequency data'. *Journal of Econometrics*, 193, pp. 294–314.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). 'MIDAS regressions: Further results and new directions'. *Econometric Reviews*, 26, pp. 53–90.

Giannone, D., Lenza, M., and Primiceri, G. E. (2015). 'Prior selection for vector autoregressions'. *The Review of Economics and Statistics*, 97, pp. 436–451.

Giannone, D., Reichlin, L., and Small, D. (2008). 'Nowcasting: The real-time informational content of macroeconomic data'. *Journal of Monetary Economics*, 55, pp. 665–676.

Harvey, A. C. and Pierse, R. G. (1984). 'Estimating missing observations in economic time series'. *Journal of the American Statistical Association*, 79, pp. 125–131.

Iversen, J., Laséen, S., Lundvall, H., and Söderström, U. (2016). 'Real-time forecasting for monetary policy analysis: The case of Sveriges Riksbank'. Working Paper No. 318, Sveriges Riksbank.

Kadiyala, K. R. and Karlsson, S. (1997). 'Numerical methods for estimation and inference in Bayesian VAR-models'. *Journal of Applied Econometrics*, 12, pp. 99–132.

Karlsson, S. (2013). 'Forecasting with Bayesian vector autoregression'. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2, chapter 15, pp. 791–897. Elsevier B.V.

Kuzin, V., Marcellino, M., and Schumacher, C. (2011). 'MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro area'. *International Journal of Forecasting*, 27, pp. 529–542.

Litterman, R. B. (1986). 'A statistical approach to economic forecasting'. *Journal of Business & Economic Statistics*, 4, pp. 1–4.

Mariano, R. S. and Murasawa, Y. (2003). 'A new coincident index of business cycles based on monthly and quarterly series'. *Journal of Applied Econometrics*, 18, pp. 427–443.

Mariano, R. S. and Murasawa, Y. (2010). 'A coincident index, common factors, and monthly real GDP'. *Oxford Bulletin of Economics and Statistics*, 72, pp. 27–46.

Österholm, P. (2008). 'Can forecasting performance be improved by considering the steady state? an application to Swedish inflation and interest rate'. *Journal of Forecasting*, 27, pp. 41–51.

Österholm, P. (2010). 'The effect on the Swedish real economy of the financial crisis'. *Applied Financial Economics*, 20, pp. 265–274.

Rodriguez, A. and Puggioni, G. (2010). 'Mixed frequency models: Bayesian approaches to estimation and prediction'. *International Journal of Forecasting*, 26, pp. 293–311.

Schorfheide, F. and Song, D. (2015). 'Real-time forecasting with a mixed-frequency VAR'. *Journal of Business & Economic Statistics*, 33, pp. 366–380.

Stock, J. H. and Watson, M. W. (2001). 'Vector Autoregressions'. *Journal of Economic Perspectives*, 15, pp. 101–115.

Villani, M. (2009). 'Steady state priors for vector autoregressions'. *Journal of Applied Econometrics*, 24, pp. 630–650.

Supplementary materials to 'A mixed-frequency Bayesian vector autoregression with a steady-state prior'

A. Replication files

The data set and all files used for producing the results in the paper are available at `https://doi.org/10.5281/zenodo.1145828`.

B. MCMC algorithms

Algorithm 1 presents the main steps of the Gibbs sampler that can be employed to sample from the posterior distribution.

Algorithm 1 Gibbs sampler for mixed-frequency steady-state Bayesian VAR

1: **for** $j = 1, \ldots, R$ **do**
2: Draw $Z^{(j)}$ from $p(Z|\Sigma^{(j-1)}, \Pi^{(j-1)}, \psi^{(j-1)})$
3: Draw $(\Pi, \Sigma)^{(j)}$ from $p(\Pi, \Sigma|\psi^{(j-1)}, Z^{(j)})$
4: Draw $\psi^{(j)}$ from $p(\psi|\Pi^{(j)}, \Sigma^{(j)}, Z^{(j)})$
5: **end for**

As discussed in the main text, the first step is carried out by use of the simulation smoother, which is described in more detail in Appendix D. The second and third steps amount to draws from the normal-inverse-Wishart and multivariate normal distributions, respectively, for which the posterior moments are given in Appendix C.

Section III mentions a reduced Gibbs step for estimating the marginal data density. Such a step entails estimating the full model as usual and computing the posterior mean $\tilde{\psi}$. Next, draws $\{Z^{(j)}\}$ from $p(Z|\tilde{\psi}, Y)$ are obtained using Algorithm 2, which is the main MCMC algorithm with the alteration that ψ is fixed at $\tilde{\psi}$.

The draws $\{Z^{(j)}\}$ obtained from Algorithm 2 are used to compute (7), after which the marginal data density estimate can be computed.

Algorithm 2 Reduced Gibbs step

1: **for** $j = 1, \ldots, R$ **do**
2: Draw $Z^{(j)}$ from $p(Z|\Pi^{(j-1)}, \Sigma^{(j-1)}, \tilde{\psi}, Y)$
3: Draw $(\Pi, \Sigma)^{(j)}$ from $p(\Pi, \Sigma|\tilde{\psi}, Z^{(j)}, Y)$
4: **end for**

C. Posterior distributions

When conditioning on latent states and unconditional mean, the model is a standard VAR for $(z_t - \Psi d_t)$ and the resulting posteriors follow standard results, available in for example Karlsson (2013). Thus, the posterior distribution for the dynamic coefficients is

$$\text{vec}(\Pi')|Z, \Sigma, \psi \sim N(\text{vec}(\overline{\Pi}'), \Sigma \otimes \overline{\Omega}_\Pi),$$

or, equivalently

$$\Pi'|Z, \Sigma, \psi \sim MN(\overline{\Pi}', \Sigma, \overline{\Omega}_\Pi)$$

where

$$\overline{\Omega}_\Pi^{-1} = \underline{\Omega}_\Pi^{-1} + \tilde{Z}'_{1:T-1}\tilde{Z}_{1:T-1}, \quad \overline{\Pi} = \overline{\Omega}_\Pi(\underline{\Omega}_\Pi^{-1}\underline{\Pi} + \tilde{Z}'_{1:T-1}\tilde{z}_{2:T}).$$

The demeaned z_t (in non-companion form) can be written as

$$\tilde{z} = \begin{pmatrix} z'_1 - d'_1\Psi' \\ \vdots \\ z'_T - d'_T\Psi' \end{pmatrix} = \begin{pmatrix} z'_1 \\ \vdots \\ z'_T \end{pmatrix} - \begin{pmatrix} d'_1\Psi' \\ \vdots \\ d'_T\Psi' \end{pmatrix} = z - \begin{pmatrix} \psi'(d_1 \otimes I_p) \\ \vdots \\ \psi'(d_T \otimes I_p) \end{pmatrix}$$

In companion form, we thus have

$$\tilde{Z}_t = \begin{pmatrix} \tilde{z}_t \\ \tilde{z}_{t-1} \\ \vdots \\ \tilde{z}_{t-p+1} \end{pmatrix}, \quad \tilde{Z}_{1:T} = \begin{pmatrix} \tilde{Z}_1' \\ \tilde{Z}_2' \\ \vdots \\ \tilde{Z}_T' \end{pmatrix} = \begin{pmatrix} \tilde{z}_1' & \tilde{z}_0' & \cdots & \tilde{z}_{2-p}' \\ \tilde{z}_2' & \tilde{z}_1' & \cdots & \tilde{z}_{3-p}' \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{z}_T' & \tilde{z}_{T-1}' & \cdots & \tilde{z}_{T-p+1}' \end{pmatrix}.$$

The posterior for the error covariance is

$$\begin{aligned} \Sigma|Z, \psi &\sim IW(\overline{S}, \overline{\nu}),\ \overline{\nu} = T + \underline{\nu} \\ \overline{S} &= \underline{S} + S + (\underline{\Pi} - \widehat{\Pi})' \left(\underline{\Omega}_{\Pi} + (\tilde{Z}_{1:T-1}' \tilde{Z}_{1:T-1})^{-1} \right)^{-1} (\underline{\Pi} - \widehat{\Pi}) \\ \widehat{\Pi} &= (\tilde{Z}_{1:T-1}' \tilde{Z}_{1:T-1})^{-1} \tilde{Z}_{1:T-1}' \tilde{z}_{2:T}, \\ S &= (\tilde{z}_{2:T} - \tilde{Z}_{1:T-1} \widehat{\Pi})' (\tilde{z}_{2:T} - \tilde{Z}_{1:T-1} \widehat{\Pi}). \end{aligned}$$

To find the conditional posterior of Ψ, the results of Villani (2009) still apply as we are

now conditioning on Z. Let

$$\tilde{\mathbf{Y}} = \begin{pmatrix} z_1'\Pi(L)' \\ \vdots \\ z_T'\Pi(L)' \end{pmatrix} = \begin{pmatrix} z_1' - z_0'\Pi_1' - z_{-1}'\Pi_2' - \cdots - z_{-p+1}'\Pi_p' \\ \vdots \\ z_T' - z_{T-1}'\Pi_1' - z_{T-2}'\Pi_2' - \cdots - z_{T-p}'\Pi_p' \end{pmatrix} = \begin{pmatrix} z_1' - \begin{pmatrix} z_0' & \cdots & z_{-p+1}' \end{pmatrix}\Pi \\ \vdots \\ z_T' - \begin{pmatrix} z_{T-1}' & \cdots & z_{T-p}' \end{pmatrix}\Pi \end{pmatrix}$$

$$= \begin{pmatrix} z_1' - Z_0'\Pi \\ \vdots \\ z_T' - Z_{T-1}'\Pi \end{pmatrix} = z_{1:T} - Z_{0:T-1}\Pi$$

$$\mathbf{D} = \begin{pmatrix} D_1^- \\ \vdots \\ D_T^- \end{pmatrix} = \begin{pmatrix} d_1' & -d_0' & \cdots & -d_{1-p}' \\ d_2' & -d_1' & \cdots & -d_{2-p}' \\ \vdots & \vdots & \ddots & \vdots \\ d_T' & -d_{T-1}' & \cdots & -d_{T-p}' \end{pmatrix}, \quad D_t^- = \begin{pmatrix} d_t' & -d_{t-1}' & \ldots & -d_{t-p}' \end{pmatrix}$$

and

$$\mathbf{\Theta}' = \begin{pmatrix} \psi & \Pi_1\psi & \ldots & \Pi_p\psi \end{pmatrix}.$$

The model in (3) can be written

$$\Pi(L)z_t = \Pi(L)\Psi d_t + u_t = \Psi d_t - \Pi_1\Psi d_{t-1} - \cdots - \Pi_p\Psi d_{t-p} + u_t$$

such that

$$\tilde{\mathbf{Y}} = \mathbf{D}\mathbf{\Theta} + \mathbf{U}$$

where $\mathbf{U} = (u_1, \ldots, u_T)'$. The conditional posterior $\psi|Z, \Sigma, \Pi$ follows from multivariate

regression and hence

$$\text{vec}(\mathbf{\Theta}') = \mathbf{E}\psi = \begin{pmatrix} \mathbf{I}_{pm} \\ \mathbf{I}_m \otimes \Pi_1 \\ \vdots \\ \mathbf{I}_m \otimes \Pi_p \end{pmatrix} \psi.$$

$$\psi | Z, \Pi, \Sigma \sim N(\overline{\psi}, \overline{\Omega}_\psi)$$

$$\overline{\Omega}_\psi^{-1} = \mathbf{E}'(\mathbf{D}'\mathbf{D} \otimes \Sigma^{-1})\mathbf{E} + \underline{\Omega}_\psi^{-1}$$

$$\overline{\psi} = \overline{\Omega}_\psi(\mathbf{E}' \,\text{vec}(\Sigma^{-1}\mathbf{Y}'\mathbf{D}) + \underline{\Omega}_\psi^{-1}\underline{\psi}).$$

In summary, the three conditional posteriors for the parameters are

$$\Sigma | Z, \psi \sim IW(\overline{S}, \overline{\nu})$$
$$\text{vec}(\Pi') | Z, \Sigma, \psi \sim N(\text{vec}(\overline{\Pi}'), \Sigma \otimes \overline{\Omega}_\Pi)$$
$$\psi | Z, \Pi, \Sigma \sim N(\overline{\psi}, \overline{\Omega}_\psi).$$

D. Simulation smoother

Handling the deterministic terms

Let W_t be as in (5) so that $W_t\psi = E(Z_t)$. The original state-space model is

$$\begin{aligned} y_t &= M_t \Lambda Z_t \\ Z_{t+1} &= W_{t+1}\psi + F(\Pi)(Z_t - W_t\psi) + \varepsilon_t. \end{aligned} \tag{8}$$

We can note that

$$y_t^* \equiv y_t - M_t\Lambda W_t\psi = M_t\Lambda(Z_t - W_t\psi) \equiv M_t\Lambda Z_t^*$$

is the same as $y_t = M_t \Lambda Z_t$. Using $Z_t^* = Z_t - W_t \psi$, the formulation in (8) is equivalent to

$$\begin{aligned} y_t^* &= M_t \Lambda Z_t^* \\ Z_{t+1}^* &= F(\Pi) Z_t^* + \varepsilon_{t+1}, \\ \varepsilon_t &\sim N(0, \Omega(\Sigma)). \end{aligned} \tag{9}$$

Thus, it is sufficient to provide a treatment of the Kalman filter without deterministic components in what follows, as we can simply do any Kalman filter steps based on y_t^* and Z_t^*, and then add the deterministic term appropriately.

The Durbin and Koopman (2002) simulation smoother

Assume the state-space system in (9) for the time period $t = 1, \ldots, T$ and treat in the following the parameters as fixed and known. Applying the Kalman filter means to apply the following equations recursively:

$$\begin{aligned} a_t = F(\Pi) a_{t-1} + K_{t-1} v_{t-1}, \quad P_t = F(\Pi) P_{t-1} L_{t-1}' + \Omega(\Sigma), \quad v_t = y_t - M_t \Lambda a_t \\ F_t = M_t \Lambda P_t \Lambda' M_t', \quad K_t = F(\Pi) P_t \Lambda' M_t' F_t^{-1}, \quad L_t = F(\Pi) - K_t M_t \Lambda \end{aligned} \tag{10}$$

for $t = 1, \ldots, T$. Note that the dimensions of v_t, F_t and K_t vary deterministically as a function of what observations are available at time t. The smoothed state vector is computed by first evaluating

$$r_{t-1} = M_t \Lambda F_t^{-1} v_t + L_t' r_t \tag{11}$$

backwards, i.e. for $t = T, T-1, \ldots, 1$ with $r_T = 0$. Then, the smoothed mean of the latent state, $\hat{Z}_t = E(Z_t | Y_{1:T})$, is computed by applying the forwards recursion

$$\hat{Z}_{t+1} = F(\Pi) \hat{Z}_t + \Omega(\Sigma) r_t, \tag{12}$$

which is initialized by $\hat{Z}_1 = a_1 + P_1 r_0$. To draw from the density $p(Z_{1:T}|Y_{1:T})$, the main idea is to use (9) to generate pseudo-variables y^+ and Z^+ by drawing ε_t from $N(0, \Omega(\Sigma))$ and using the recursions in (9). Given these pseudo-variables, the latent state is smoothed to yield $\hat{Z}^+ = E(Z^+_{1:T}|Y^+_{1:T})$. The final draw is then obtained by computing

$$\tilde{Z}_{1:T} = \hat{Z}_{1:T} + \left(Z^+_{1:T} - \hat{Z}^+_{1:T}\right).$$

Note that $\hat{Z}_{1:T}$ is obtained by first applying the Kalman smoother based on y^*_t and Z^*_t, which yields $\hat{Z}^*_{1:T}$. Then, $\hat{Z}_{1:T} = \hat{Z}^*_{1:T} + (I_N \otimes \psi')W_{1:T}$, where $W_{1:T} = (W_1, \ldots, W_T)'$.

Algorithm 3 Simulation smoother

1: Draw $\varepsilon^+_1, \ldots, \varepsilon^+_T$ from $N(0, \Omega(\Sigma))$ and use (9) to construct $Y^+_{1:T}$ and $Z^+_{1:T}$
2: Compute $\hat{Z}^*_{1:T} = E(Z^*_{1:T}|Y^*_{1:T})$ and $\hat{Z}^+_{1:T} = E(Z^+_{1:T}|Y^+_{1:T})$ by applying (10) forwards, (11) backwards and (12) forwards
3: Compute the final draw as $\tilde{Z}_{1:T} = \hat{Z}^*_{1:T} + (I_T \otimes \psi')W_{1:T} + \left(Z^+_{1:T} - \hat{Z}^+_{1:T}\right)$

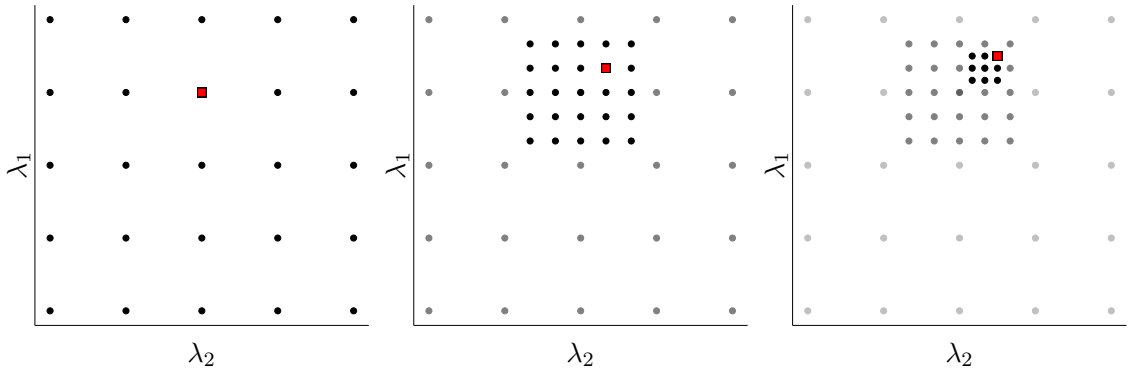
The simulation smoother can be time consuming, but the computational burden can be alleviated by using the computational refinements presented in the Online Appendix to Schorfheide and Song (2015).

Initialization

The simulation smoother needs to be initialized by a_0 and P_0. To do this, we fix $Z_0 = (z_0, \ldots, z_{-p+1})'$ at its observed values where applicable and fill the remaining missing entries with the previous period's observation. If initial observations are missing we set them to the next available observation.

E. Hyperparameter selection

Both the Minnesota and the steady-state prior are parameterized by the same two hyperparameters: λ_1 for overall tightness, and λ_2 for lag decay. To select appropriate values for



(a) Step 1 (b) Step 2 (c) Step 3

Figure 11. Adaptive grid search
Notes: Circles represent evaluated points in the grid and squares the maximizing pair in each step. The figure illustrates grids of size 5, 5 and 3 in each step; in the application we use 7, 5 and 3.

these, we employ an adaptive grid search. First, we compute the marginal data density for a 7×7 two-dimensional grid of hyperparameter values. Next, we calculate the marginal data density for a 5×5 grid centered on the maximizing point from the first grid. Let $\lambda_1^{(j)}$ denote the jth value in the first grid for λ_1 and suppose that this value maximizes the marginal data density. The endpoints in the second step's grid are set to $\lambda_1^{(j-1)} + (\lambda_1^{(j)} - \lambda_1^{(j-1)})/3$ and $\lambda_1^{(j+1)} - (\lambda_1^{(j+1)} - \lambda_1^{(j)})/3$. If $\lambda_1^{(j)}$ is a boundary point, we instead let the upper (or lower) endpoint be $\lambda_1^{(j)}$.

We take the same approach for the second grid of values for λ_2, and thus end up with a rectangular grid centered on the first step's maximizer with corners inside the neighboring points in the first grid. Finally, the third step is conducted in a similar fashion using a 3×3 grid. Figure 11 illustrates the method visually.[4]

The adaptive grid search is conducted for all models at all forecasting origins. The final forecast used is the forecast made by the model with the largest marginal data density at that specific origin. Following some preliminary runs, the grids in the first step are set to

[4]For space considerations, the first step in the Figure shows only a 5×5 grid. In the application, we instead use a 7×7 grid in the first step.

seven equally-spaced values between 0.01 and 1 for λ_1. The models with the steady-state prior use seven equally-spaced values between 0.01 and 4 for λ_2, whereas the Minnesota-based models use seven values between 0.01 and 8.

Research Papers 2018

2018-14: Cristina Amado, Annastiina Silvennoinen and Timo Teräsvirta: Models with Multiplicative Decomposition of Conditional Variances and Correlations

2018-15: Changli He, Jian Kang, Timo Teräsvirta and Shuhua Zhang: The Shifting Seasonal Mean Autoregressive Model and Seasonality in the Central England Monthly Temperature Series, 1772-2016

2018-16: Ulrich Hounyo and Rasmus T. Varneskov: Inference for Local Distributions at High Sampling Frequencies: A Bootstrap Approach

2018-17: Søren Johansen and Morten Ørregaard Nielsen: Nonstationary cointegration in the fractionally cointegrated VAR model

2018-18: Giorgio Mirone: Cross-sectional noise reduction and more efficient estimation of Integrated Variance

2018-19: Kim Christensen, Martin Thyrsgaard and Bezirgen Veliyev: The realized empirical distribution function of stochastic variance with application to goodness-of-fit testing

2018-20: Ruijun Bu, Kaddour Hadri and Dennis Kristensen: Diffusion Copulas: Identification and Estimation

2018-21: Kim Christensen, Roel Oomen and Roberto Renò: The drift burst hypothesis

2018-22: Russell Davidson and Niels S. Grønborg: Time-varying parameters: New test tailored to applications in finance and macroeconomics

2018-23: Emilio Zanetti Chini: Forecasters' utility and forecast coherence

2018-24: Tom Engsted and Thomas Q. Pedersen: Disappearing money illusion

2018-25: Erik Christian Montes Schütte: In Search of a Job: Forecasting Employment Growth in the US using Google Trends

2018-26: Maxime Morariu-Patrichi and Mikko Pakkanen: State-dependent Hawkes processes and their application to limit order book modelling

2018-27: Tue Gørgens and Allan H. Würtz: Threshold regression with endogeneity for short panels

2018-28: Mark Podolskij, Bezirgen Veliyev and Nakahiro Yoshida: Edgeworth expansion for Euler approximation of continuous diffusion processes

2018-29: Isabel Casas, Jiti Gao and Shangyu Xie: Modelling Time-Varying Income Elasticities of Health Care Expenditure for the OECD

2018-30: Yukai Yang and Luc Bauwens: State-Space Models on the Stiefel Manifold with A New Approach to Nonlinear Filtering

2018-31: Stan Hurn, Nicholas Johnson, Annastiina Silvennoinen and Timo Teräsvirta: Transition from the Taylor rule to the zero lower bound

2018-32: Sebastian Ankargren, Måns Unosson and Yukai Yang: A mixed-frequency Bayesian vector autoregression with a steady-state prior