Manuscript version: Author's Accepted Manuscript
The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:
http://wrap.warwick.ac.uk/109567

How to cite:
Please refer to published version for the most recent bibliographic citation information.
If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

warwick.ac.uk/lib-publications

Bayesian Vector Autoregressions:
Applications

Silvia Miranda-Agrippino[*]

Bank of England
Centre for Macroeconomics (LSE)

Giovanni Ricco[†]

University of Warwick
OFCE - SciencesPo

Revised 16 August 2018

Summary

Bayesian Vector Autoregressions (BVARs) are standard multivariate autoregressive models routinely used in empirical macroeconomics and finance for structural analysis, forecasting and scenario analysis in an ever growing number of applications.

A preeminent field of application of BVARs is forecasting. BVARs with informative priors have often proved to be superior tools compared to standard frequentist/flat-prior VARs. In fact, VARs are highly parametrised autoregressive models, whose number of parameters grows with the square of the number of variables times the number of lags included. Prior information, in the form of prior distributions on the model parameters, helps in forming sharper posterior distributions of parameters, conditional on an observed sample. Hence, BVARs can be effective in reducing parameters uncertainty, and improving forecast accuracy compared to standard frequentist/flat-prior VARs.

This feature in particular has favoured the use of Bayesian techniques to address 'big data' problems, in what is arguably one of the most active frontiers in

[*]Monetary Analysis, Bank of England, Threadneedle Street, London EC2R 8AH, UK. Email: silvia.miranda-agrippino@bankofengland.co.uk   Web: www.silviamirandaagrippino.com

[†]Department of Economics, The University of Warwick, The Social Sciences Building, Coventry, West Midlands CV4 7AL, UK. Email: G.Ricco@warwick.ac.uk   Web: www.giovanni-ricco.com

the BVAR literature. Large-information BVARs have in fact proven to be valuable tools to handle empirical analysis in data-rich environments.

BVARs are also routinely employed to produce conditional forecasts and scenario analysis. Of particular interest for policy institutions, these applications permit evaluating 'counterfactual' time evolution of the variables of interests conditional on a pre-determined path for some other variables, such as the path of interest rates over a certain horizon.

The 'structural interpretation' of estimated VARs as the data generating process of the observed data requires the adoption of strict 'identifying restrictions'. From a Bayesian perspective, such restrictions can be seen as dogmatic prior beliefs about some regions of the parameter space that determine the contemporaneous interactions among variables, and for which the data are uninformative. More generally, Bayesian techniques offer a framework for structural analysis through priors that incorporate uncertainty about the identifying assumptions themselves.

Keywords: Bayesian inference, Vector Autoregression Models, BVAR, SVAR, forecasting
JEL Classification: C30, C32

Contents

1 Introduction

Bayesian Vector Autoregressions (BVARs) have been applied to an increasingly large number of empirical problems. In this article we review some of the most common applications of BVARs in macroeconomics and finance, such as structural identification, forecasting and scenario analysis. A companion article surveys Bayesian inference methods for Vector Autoregression models, commonly used priors for economic and financial variables, and it introduces the notation used throughout this paper (Miranda-Agrippino and Ricco, 2018).

Forecasting has featured predominantly in the development of BVARs. In this context, BVARs with informative priors have often proved to be superior tools compared to standard frequentist/flat-prior VARs. VARs are highly parametrised autoregressive models, whose number of parameters grows with the square of the number of variables times the number of lags included. Given the limited length of standard macroeconomic datasets – that usually involve monthly, quarterly, or even annual observations –, such overparametrisation makes the estimation of VARs impossible with standard (frequentist) techniques, already for relatively small sets of variables. This is known in the literature as the 'curse of dimensionality'. BVARs efficiently deal with the problem of over-parametrisation through the use of prior information about the model coefficients. The general idea is to use informative priors that shrink the unrestricted model towards a parsimonious naïve benchmark, thereby reducing parameter uncertainty, and improving forecast accuracy. Section 2 discusses forecasting with BVARs, while Section 3 focusses on conditional forecast and scenario analysis.

Another important area of application is the study of causal relationships among economic variables with Structural (B)VARs (Sims and Zha, 1998). It is common practice to present results from SVARs in the form of impulse response functions – i.e. causal responses over time of a given variable of interest to an 'identified' economic shock – together with bands that characterise the shape of the posterior distribution of the model (see Sims and Zha, 1999).[1] Section 4 reviews Bayesian techniques in SVARs.

---

[1] An extreme version of lack of sample information arises in this context. In fact Structural VARs can be parametrised in terms of reduced form VARs that capture the joint dynamics of economic variables,

The application of Bayesian techniques to 'big data' problems is one of the most active frontiers in the BVAR literature. Indeed, because they can efficiently deal with parameters proliferation, large BVARs are valuable tools to handle empirical analysis in data-rich environments (Bańbura et al., 2010). Important applications in this case also concern forecasting and structural analysis, where large-information BVARs can efficiently address issues related to misspecification and non-fundamentalness. De Mol et al. (2008) have discussed the connection between BVARs and factor models, another popular way to handle large datatsets. We review large BVARs in Section 5.

2 Forecasting with BVARs

Reduced form Bayesian Vector Autoregressions, that are written as

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + c + u_t \, , \tag{1}$$

when estimated with informative priors over the autoregressive parameters $\{A_i, \ldots A_p\}$ and the variance-covariance of the errors Σ, usually outperform VARs estimated with frequentist techniques (or flat priors). Using the frequentist terminology, reasonably specified priors reduce estimated parameters variance and hence improve forecast accuracy, at the cost of the introduction of relatively small biases. From a more Bayesian perspective, the prior information that may not be apparent in short samples – as for example the long-run properties of economic variables captured by the Minnesota priors – helps in forming sharper posterior distributions for the VAR parameters, conditional on an observed sample (see e.g. Todd, 1984, for an early treatment of forecasting with BVARs).

---

and an 'impact matrix' describing the casual connection between stochastic disturbances and economic variables. This matrix is not uniquely identified by sample information and hence the investigator has to elicit prior beliefs on it (see Sims and Zha, 1998; Baumeister and Hamilton, 2015).

2.1 Bayesian Forecasting

The fundamental object in Bayesian forecasting is the posterior predictive density.[2] That is, the distribution of future data points $y_{T+1:T+H} = [y'_{T+1}, \ldots, y'_{T+H}]'$, conditional on past data $y_{1-p:T}$. Choosing a particular forecast $\mathcal{F}$ – e.g. the mode or median of the predictive distribution, alongside appropriate probability intervals –, is essentially a decision problem, given a specified loss function $\mathcal{L}(\cdot)$. The Bayesian decision corresponds to choosing the forecast that minimises the expected loss, conditional on past data

$$\mathbb{E}[\mathcal{L}(\mathcal{F}, y_{T+1:T+H}|y_{1-p:T})] = \int \mathcal{L}(\mathcal{F}, y_{T+1:T+H})p(y_{T+1:T+H}|y_{1-p:T})d_{y_{T+1:T+H}}. \tag{2}$$

For a given loss function, the solution to the minimisation problem is a function of the data, i.e. $\mathcal{F}(y_{1-p:T})$. For example, with quadratic loss function $\mathcal{L}(\mathcal{F}, y_{T+1:T+H}|y_{1-p:T}) = (\mathcal{F} - y_{T+1:T+H})'(\mathcal{F} - y_{T+1:T+H})$, the solution is the conditional expectation $\mathcal{F}(y_{1-p:T}) = \mathbb{E}[y_{T+1:T+H}|y_{1-p:T}]$. The predictive density is given by

$$p(y_{T+1:T+H}|y_{1-p:T}) = \int p(y_{T+1:T+H}|y_{1-p:T}, \theta)p(\theta|y_{1-p:T})d\theta, \tag{3}$$

where θ is the vector collecting all the VAR parameters, i.e. A and Σ, $p(\theta|y_{1-p:T})$ is the posterior distribution of the parameters, and $p(y_{T+1:T+H}|y_{1-p:T}, \theta)$ is the likelihood of future data. Eq. (3) highlights how Bayesian forecasts account for both the uncertainty related to future events via $p(y_{T+1:T+H}|y_{1-p:T}, \theta)$, and that related to parameters values via $p(\theta|y_{1-p:T})$.

The posterior predictive density for $h > 1$ is not given by any standard density function. However, if it is possible to sample directly from the posterior probability for the parameters, Eq. (3) provides an easy way to generate draws from this predictive density.

Algorithm 1: Sampling from the Posterior Predictive Density.

For $s = 1, \ldots, n_{sim}$:

1. Draw $\theta^{(s)}$ from the posterior $p(\theta|y_{1-p:T})$.

---

[2]The exposition in this section follows Karlsson (2013). See also Geweke and Whiteman (2006).

2. Generate $u_{T+1}^{(s)}, \ldots, u_{T+H}^{(s)}$ from the distribution of the errors and calculate recursively $\tilde{y}_{T+1}^{(s)}, \ldots, \tilde{y}_{T+H}^{(s)}$ from the VAR equations with parameters $A^{(s)}$.

The set $\left\{ \tilde{y}_{T+1}^{(s)}, \ldots, \tilde{y}_{T+H}^{(s)} \right\}_{s=1}^{n_{sim}}$ is a sample of independent draws from the joint predictive distribution.

Kadiyala and Karlsson (1993) analyse the forecasting performance of different priors and find that those that induce correlation among the VAR coefficients, e.g. the sums-of-coefficient priors (Doan et al., 1984) and the co-persistence prior (Sims, 1993), tend to do better.

Carriero et al. (2015a) conduct an extensive assessment of Bayesian VARs under different specifications and evaluate the relative merits of different model specifications and treatments of the data. In particular, starting from a benchmark VAR in levels and with NIW, sums-of-coefficients, and co-persistence priors, they evaluate (1) the effects of the optimal choice of the tightness hyperparameters, (2) of the lag length, (3) of the relative merits of modelling in levels or growth rates, (4) of direct, iterated and pseudo-iterated h-step-ahead forecasts, (5) the treatment of the error variance Σ and (6) of cross-variable shrinkage $f(\ell)$. They find that in general simpler specifications tend to be very effective.[3,4]

2.2   Bayesian Model Averaging and Prediction Pools

Bayesian analysis offers a straightforward way to deal with model uncertainty. Consider for instance the two competing models $\mathcal{M}_1$ and $\mathcal{M}_2$ with likelihood $p(\mathbf{y}|\theta_1, \mathcal{M}_1, y_{1-p:0})$ and $p(\mathbf{y}|\theta_2, \mathcal{M}_2, y_{1-p:0})$ and prior probabilities $p(\theta_1|\mathcal{M}_1)$ and $p(\theta_2|\mathcal{M}_2)$ respectively. Bayesian Model Averaging (BMA) obtains the marginalised (with respect to the models)

---

[3]Since the work of Sims et al. (1990), it is common practice to keep variables in (log-)levels. However, whether to employ in VARs variables in growth rates, log-levels, or levels remains an empirically important question. In the context of forecasting, for example, Carriero et al. (2015a) recommend the use of differenced data.

[4]Carriero et al. (2015a) also find that overall the differences between the iterated and direct forecasts are small, but there are large gains from the direct forecast for some of the variables. This is presumably because the direct forecast is more robust to misspecification.

predictive distribution as

$$p(y_{T+1:T+H}|\mathbf{y}) = p(y_{T+1:T+H}|\mathbf{y}, \mathcal{M}_1)p(\mathcal{M}_1) + p(y_{T+1:T+H}|\mathbf{y}, \mathcal{M}_2)p(\mathcal{M}_2), \qquad (4)$$

where $p(\mathcal{M}_j)$ is the prior probability assigned to model $\mathcal{M}_j$, and $p(y_{T+1:T+H}|\mathbf{y}, \mathcal{M}_j)$ is the model's marginal likelihood. Eq. (4) can be extended to allow for M different models. This can be seen as a generalisation of the predictive distribution in Eq. (3) where instead of conditioning on a single model, M different models are considered. BMA was introduced in economic forecasting by the seminal work of Geweke (1999) and its applications in the context of forecast combinations and pooling have been numerous. Notable extensions to BMA include the Linear Optimal Prediction Pools of Geweke and Amisano (2011, 2012), and the Dynamic Prediction Pools of Del Negro et al. (2016).[5] Earlier reviews of BMA and forecast combinations are in Geweke and Whiteman (2006) and Timmermann (2006). The evolution of forecast density combinations is discussed in detail in Aastveit et al. (2018)'s chapter in this collection. We refer the reader to their paper for further details.

3 Conditional Forecasts and Scenario Analysis

Forecasts that condition on a specific path for one of the variables, such as e.g. a preferred path for the policy interest rate, are of particular interest to central banks. Early treatment of such forecasts, also referred to as scenario analysis, is in Doan et al. (1984), who note that a conditional forecast is equivalent to imposing restrictions on the disturbances $u_{T+1}, \ldots, u_{t+H}$. Waggoner and Zha (2012) suggest a way to compute conditional forecasts which does not condition on specific parameters values (for example the posterior means) and produces minimum squared forecast errors conditional on the restrictions. Moreover, it yields posterior distributions for the parameters which are

---

[5]Other relevant contributions on density forecast combination are Waggoner and Zha (2012); Geweke and Amisano (2011); Hall and Mitchell (2007); Billio et al. (2013); Amisano and Geweke (2017); Raftery et al. (2010). Applications are in e.g. Hwang (2017); Koop and Korobilis (2012) and Aastveit et al. (2017) in the context of real-time forecasting.

consistent with the constrained paths. Let

$$Ry_{T+1:T+H} = r \qquad (5)$$

denote the desired restrictions on the future path of some of the variables in y_t. These can be rewritten as

$$R\left[\mathbb{E}(y_{T+1:T+H}|\mathbf{y},\theta) + C'u_{T+1:T+H})\right] = r \ , \qquad (6)$$

where

$$C = \begin{bmatrix} C_0 & C_1 & \cdots & C_{H-1} \\ 0 & C_0 & \cdots & C_{H-2} \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & C_0 \end{bmatrix} , \qquad (7)$$

and C_j are the coefficients of the MA representation with

$$C_0 = \mathbb{I}_n$$
$$C_j = \sum_{i=0}^{p} A_i C_{j-i} \quad \forall j > 0 \ . \qquad (8)$$

Rearranging Eq. (6) as

$$RC'u_{T+1:T+H} = r - R\mathbb{E}(y_{T+1:T+H}|y_{1-p:T},\theta) \ , \qquad (9)$$

defining $G \equiv RC'$ and $g \equiv r - R\mathbb{E}(y_{T+1:T+H}|y_{1-p:T},\theta)$, and noting that $u_{T+1:T+H} \sim \mathcal{N}(0,\mathbb{I}_H \otimes \Sigma)$, one obtains the conditional distribution of $u_{T+1:T+H}$ as

$$u_{T+1:T+H}|(Gu_{T+1:T+H} = g) \sim \mathcal{N}\left(\Sigma_H G'(G\Sigma_H G')^{-1}g, \ \Sigma_H - \Sigma_H G'(G\Sigma_H G')^{-1}G\Sigma_H\right) \qquad (10)$$

which can be used to draw from the predictive distribution. In order to ensure consistency of the posterior distribution with the restriction in Eq. (9), Waggoner and Zha

(2012) suggest treating $y_{T+1:T+H}$ as latent variables and simulating the joint posterior of the parameters and the future observations using the following MCMC sampler.

Algorithm 2:   MCMC Sampler for VAR with restrictions on $y_{T+1:T+H}$.
Given restrictions as in Eq. (9), select starting values for $A^{(0)}$ and $\Sigma^{(0)}$ using e.g. simulation on historical data. For $s = 1, \ldots, n_{sim}$:

1. Draw $u_{T+1:T+H}$ from the distribution in Eq. (10) and recursively calculate

$$y_{T+h}^{(s)} = \sum_{j=1}^{h-1} y_{T+h-j}^{(s)\prime} A_j^{(s-1)} + \sum_{j=h}^{p} y_{T+h-j}^\prime A_j^{(s-1)} + u_{T+h}^{(s)\prime} \ .$$

2. Augment $y_{1-p:T}$ with $y_{T+1:T+h}^{(s)}$ and draw $A^{(s)}$ and $\Sigma^{(s)}$ from the full conditional posteriors

$$\Sigma^{(s)} | y_{1-p:T}, y_{T+1:T+h}^{(s)}, A^{(s-1)},$$
$$A^{(s)} | y_{1-p:T}, y_{T+1:T+h}^{(s)}, \Sigma^{(s)},$$

    using an appropriate sampling given the chosen VAR specification and priors.

3. Discard the parameters to obtain a draw $\left\{ y_{T+1}^{(s)}, \ldots, y_{T+h}^{(s)} \right\}$ from the joint predictive density consistent with the restrictions in Eq. (9).

Jarociński (2010) suggests an efficient way to sample $u_{T+1:T+H}$ that reduces the computational burden of the algorithm discussed above. An extension to this method is in Andersson et al. (2010), who restrict the forecasts $y_{T+1:T+H}$ to be in a specified region $\mathbb{S} \in \mathbb{R}_{nH}$. This is a case of 'soft' restrictions, as opposed to those in Eq. (9). Robertson et al. (2005) follow a different approach and propose exponential tilting as a way to enforce moment conditions on the path of future y_t. This is the approach also implemented in Cogley et al. (2005). These methods are typically used in conjunction with small VARs, and become quickly computationally cumbersome as the system's dimension increases.

Bańbura et al. (2015) propose instead a Kalman Filter-based algorithm to produce conditional forecasts in large systems which admit a state-space representation such as large Bayesian VARs and Factor Models. Intuitively, this method improves on computational efficiency due to the recursive nature of filtering techniques which allow to tackle the problem period by period.

Antolin-Diaz et al. (2018) propose a method to conduct 'structural scenario analysis' that can be supported by economic interpretation by choosing which structural shock is responsible for the conditioning path.

4 Structural VARs

Reduced form VARs can capture the autocovariance properties of multiple time-series. However, their 'structural interpretation' as the data generating process of the observed data, and of their one-step-ahead forecast errors in terms of economically meaningful shocks, requires additional identifying restrictions.

A VAR in structural form (SVAR) can be written as

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + B_c + e_t, \qquad e_t \sim i.i.d. \mathcal{N}(0, \mathbb{I}_n) , \qquad (11)$$

where B_0 is a matrix of contemporaneous (causal) relationships among the variables, and e_t is a vector of structural shocks that are mutually uncorrelated and have an economic interpretation. All structural shocks are generally assumed to be of unitary variance. This does not imply a loss of generality, however, since the diagonal elements of B_0 are unrestricted. In the structural representation, the coefficients have a direct behavioural interpretation, and it is possible to provide a causal assessment of the effects of economic shocks on variables – e.g. the effect of a monetary policy shock onto prices and output. Premultiplying the SVAR in Eq. (11) by B_0^{-1} yields its reduced-form representation, i.e. the VAR in Eq. (1). Comparing the two representations one obtains that $A_i = B_0^{-1} B_i$,

$i = 1, \ldots, p$, and $u_t = B_0^{-1} e_t$. The variance of the reduced form forecast errors, u_t is

$$\Sigma = B_0^{-1} B_0^{-1'} . \tag{12}$$

Since Σ is symmetric, it has only $n(n+1)/2$ independent parameters. This implies that the data can provide information to uniquely identify only $n(n+1)/2$ out of the n^2 parameters in B_0. In fact, given a positive definite matrix Σ, it is possible to write B_0 as the product of the unique lower triangular Cholesky factor of Σ ($\Sigma = \Sigma_{Chol} \Sigma_{Chol}'$) times an orthogonal matrix Q

$$B_0 = Q \Sigma_{Chol} . \tag{13}$$

From this decomposition it is clear that while Σ_{Chol} is uniquely determined for a given Σ, the $n(n-1)/2$ unrestricted parameters span the space of the $O(n)$ group of $n \times n$ orthogonal matrices. The central question in structural identification is how to recover the elements of B_0 given the variance-covariance matrix of the one-step-ahead forecast errors, Σ. That is, how to choose Q out of the many possible n-dimensional orthogonal matrices.[6]

From a Bayesian perspective, the issue is that since y_t depends only on Σ and not on its specific factorisation, the conditional distribution of the parameter Q does not get updated by the information provided in the data, i.e.

$$p(Q|Y, A, \Sigma) = p(Q|A, \Sigma) . \tag{14}$$

For some regions of the parameter space, posterior inference will be determined purely by prior beliefs even if the sample size is infinite, since the data are uninformative. This is a standard property of Bayesian inference in partially identified models, as discussed for example in Kadane (1975), Poirier (1998), and Moon and Schorfheide (2012).

Much of ingenuity and creativity in the SVAR literature has been devoted to provide

---

[6]It is assumed that the information in the history of y_t is sufficient to recover the structural shocks e_t, i.e., that it is possible to write the structural shocks as a linear combination of the reduced form innovations u_t. In this case, it is said that the shocks are fundamental for y_t. Departures from this case are discussed in Section 5. Relevant references are provided therein.

arguments – i.e. 'identification schemes' – about the appropriate choice of $p(Q|A, \Sigma)$.[7] These arguments translate into what can be viewed as Bayesian inference with dogmatic prior beliefs – i.e. distributions with singularities – about the conditional distribution of Q, given the reduced form parameters. For example, the commonly applied recursive identification amounts, from a Bayesian perspective, to assuming with dogmatic certainty that all of the upper diagonal elements of B_0 are zero, while we do not have any information on the other values of B_0. Equivalently, it assumes with certainty that $Q = \mathbb{I}_n$. Similarly, other commonly used identifications – e.g. long-run, medium-run, sign restrictions, etc. – can be expressed in terms of probabilistic a priori statements about the parameters in B_0.

Once a B_0 matrix is selected, dynamic causal effects of the identified structural shocks on the variables in y_t are usually summarised by the structural impulse response functions (IRFs). In a VAR(p), they can be recursively calculated as

$$IRF_h = \Theta_h B_0^{-1} \qquad h = 0, \ldots, H \ , \tag{15}$$

where

$$\Theta_h = \sum_{\tau=1}^{h} \Theta_{h-\tau} A_\tau \qquad h = 1, \ldots, H \ , \tag{16}$$

$\Theta_0 = \mathbb{I}_n$, and A_τ are the reduced form autoregressive coefficients of Eq. (1) with $A_\tau = 0$ for $\tau > p$. The (i, j) element of IRF_h denotes the response of variable i to shock j at horizon h. Uncertainty about dynamic responses to identified structural shocks is typically reported in the Bayesian literature as point-wise coverage sets around the posterior mean or median IRFs, at each horizon – i.e. as the appropriate quantiles of the IRFs posterior distribution. For example, 68% coverage intervals can be reported as two lines representing the posterior 16th and 84th percentiles of the distribution of the IRFs. Such credible sets usually need to be interpreted as point-wise, i.e. as credible sets for the response of a specific variable, to a specific shock, at a given horizon. However, point-wise bands effectively ignore the existing correlation between responses at different

---

[7]A survey of the identification schemes proposed in the literature goes beyond the scope of this article. A recent textbook treatment on the subject is in Kilian and Lütkepohl (2017).

horizons. To account for the time (horizon) dependence, Sims and Zha (1999) suggest to use the first principal components of the covariance matrix of the IRFs.

Sims and Zha (1998) discuss a very general framework for Bayesian inference on the structural representation in Eq. (11). Rewrite the SVAR as

$$yB_0 = xB + e \ , \tag{17}$$

where the $T \times n$ matrices y and e and the $T \times k$ matrix x are defined as

$$y = \begin{bmatrix} y_1' \\ \vdots \\ y_T' \end{bmatrix} , \quad x = \begin{bmatrix} x_1' \\ \vdots \\ x_T' \end{bmatrix} , \quad e = \begin{bmatrix} e_1' \\ \vdots \\ e_T' \end{bmatrix} , \tag{18}$$

and $B = [B_1, \ldots, B_p, B_c]$. The likelihood can be written as

$$p(y|B_0, B) \propto |B_0|^T exp \left\{ -\frac{1}{2} tr \left[(yB_0 - xB)'(yB_0 - xB) \right] \right\} \ , \tag{19}$$

where $|B_0|$ is the determinant of B_0 (and the Jacobian of the transformation of e in y). Conditional on B_0, the likelihood function is a normal distribution in B. Define $\beta \equiv vec(B)$ and $\beta_0 \equiv vec(B_0)$. A prior for the SVAR coefficients can be conveniently factorised as

$$p(\beta_0, \beta) = p(\beta|\beta_0)p(\beta_0), \tag{20}$$

where $p(\beta_0)$ is the marginal distribution for β_0, and can include singularities generated by e.g. zero restrictions. The (conditional) prior for β can be chosen to be a normal p.d.f.[8]

$$\beta|\beta_0 \sim \mathcal{N} \left(\underline{\beta}_0, \lambda^{-1} \mathbb{I}_n \otimes \underline{\Gamma}_{\beta_0} \right) \ . \tag{21}$$

The posterior distribution of β is hence of the standard form

$$\beta|\beta_0, \mathbf{y} \sim \mathcal{N} \left(\overline{\beta}_0, \mathbb{I}_n \otimes \overline{\Gamma}_{\beta_0} \right) \ , \tag{22}$$

---

[8]As it is usually done in the literature, Sims and Zha (1998) suggest to preserve the Kronecker structure of the likelihood to avoid the inversion of $nk \times nk$ matrices and gain computational speed.

where the posterior moments are updated as in the standard VAR with Normal-Inverse Wishart priors (see e.g. Kadiyala and Karlsson, 1997). The posterior for β_0 will depend on the assumed prior.[9]

Baumeister and Hamilton (2015) apply a streamlined version of this framework to provide analytical characterisation of the informative prior distributions for impulse-response functions that are implicit in a commonly used algorithm for sign restrictions. Sign restrictions are a popular identification scheme, pioneered in a Bayesian framework by Canova and De Nicolo (2002) and Uhlig (2005). The scheme selects sets of models whose B_0 comply with restrictions on the sign of the responses of variables of interests over a given horizon. Bayesian SVARs with sign restrictions are typically estimated using algorithms such as in Rubio-Ramírez et al. (2010), where a uniform (or Haar) prior is assumed for the orthogonal matrix. Operationally, a $n \times n$ matrix X of independent $\mathcal{N}(0, 1)$ values is generated, and decomposed using a QR decomposition where Q is the orthogonal factor and R is upper triangular. The orthogonal matrix is used as candidate rotation Q and the signs of the responses of variables at the horizons of interest are assessed against the desired sign restrictions. Baumeister and Hamilton (2015) show that this procedure implies informative distributions on the structural objects of interest. In fact, it implies that the impact of a one standard-deviation structural shock is regarded (before seeing the data) as coming from a distribution with more mass around zero when the number of variables n in the VAR is greater than 3 (and with more mass at large values when $n = 2$). It also implies Cauchy priors for structural parameters such as elasticities. The influence of these priors does not vanish even asymptotically, since the data do not contain information about Q. In fact, as the sample size goes to infinity, the height of the posterior distribution for the impact parameters is proportional to that of the prior distribution for all the points in the parameter space for which the structural coefficients satisfy the set restrictions that orthogonalise the true variance-covariance matrix.

Giacomini and Kitagawa (2015) suggest the use of 'ambiguous' prior for the struc-

---

[9]Canova and Pérez Forero (2015) provide a general procedure to estimate structural VARs also in the case of overidentified systems where identification restrictions are of linear or of nonlinear form.

tural rotation matrix in order to account for the uncertainty about the structural parameters in all under-identified SVARs. The methodology consists in formally incorporating in the inference all classes of priors for the structural rotation matrix which are consistent with the a priori 'dogmatic' restrictions. In a similar vein, Baumeister and Hamilton (2017) discuss how to generalise priors on B_0 to a less restrictive formulation that incorporates uncertainty about the identifying assumptions themselves, and use this approach to study the importance of shocks to oil supply and demand.

5   Large Bayesian VARs

The size of the VARs typically used in empirical applications ranges from three to a dozen variables. VARs with larger sets of variables are impossible to estimate with standard techniques, due the 'curse of dimensionality' induced by the densely parametrised structure of the model.[10] However, in many applications there may be concerns about the omission of many potentially relevant economic indicators, that may affect both structural analysis and forecasting.[11] Additionally, big datasets are increasingly important in economics to study phenomena in a connected and globalised world, where economic developments in one region can propagate and affect others.[12]

VARs involving tens or even hundreds of variables have become increasingly popular following the work of Bańbura et al. (2010), that showed that standard macroeconomic priors – Minnesota and sums-of-coefficients – with a careful setting of the tightness

---

[10]The number of parameters to be estimated in an unrestricted VAR increases in the square of n, the number of variables in y_t. Even when mechanically feasible, that is, when the number of available data points allows to produce point estimates for the parameters of interest, the tiny number of available degrees of freedom implies that parameters are estimated with substantial degrees of uncertainty, and typically yield very imprecise out-of-sample forecasts.

[11]A standard example of this has been the debate about the so called 'price puzzle' – positive reaction of prices in response to a monetary tightening – that is often found in small scale VARs (see for example Christiano et al., 1999). The literature has often connected such a puzzling result as an artefact resulting from the omission of forward looking variables, like the commodity price index. In fact, one of the first instances of VARs incorporating more than a few variables was the 19-variable BVAR in Leeper et al. (1996) to study the effects of monetary policy shocks.

[12]Large datasets of macroeconomic and financial variables are increasingly common. For example, in the US, the Federal Reserve Bank of St. Louis maintains the FRED-MD monthly database for well over 100 macroeconomic variables from 1960 to the present (see McCracken and Ng, 2015), and several other countries and economic areas have similarly sized datasets.

parameters allowed to effectively incorporate very large sets of endogenous variables. Indeed, a stream of papers have found large VARs to forecast well (see, e.g. Bańbura et al. 2010, Carriero et al. 2015a, Carriero et al. 2009, Giannone et al. 2014 and Koop 2013).

Early examples of higher-dimensional VARs are Panel VARs, where small country-specific VARs are interacted to allow for international spillovers (see e.g. Canova and Ciccarelli, 2004, 2009). These models can be seen as large scale models that impose more structure on the system of equations. Koop and Korobilis (2015) study methods for high-dimensional panel VARs. In the study of international spillovers, an alternative to Panel VARs are Global VARs (Pesaran et al., 2004). A Bayesian treatment to G-VARs is in e.g. Cuaresma et al. (2016).

A recent development in this literature has been the inclusion of stochastic volatility in Large BVAR models. Carriero et al. (2016a) assume a factor structure in the stochastic volatility of macroeconomic and financial variables in Large BVARs. In Carriero et al. (2016b), stochastic volatility and asymmetric priors for large n are instead handled using a triangularisation method which allows to simulate the conditional mean coefficients of the VAR by drawing them equation by equation. Chan et al. (2017) propound composite likelihood methods for large BVARs with multivariate stochastic volatility which involve estimating large numbers of parsimonious sub-models and then taking a weighted average across them. Koop et al. (2016) discuss large Bayesian VARMA. Koop (2017) reviews the applications of big data in macroeconomics.

5.1 Bayesian VARs and Dynamic Factor Models

Research started with Bańbura et al. (2010) has shown that large BVARs are competitive models in leading with large-n problems in empirical macroeconomics, along with factor models (see e.g. Forni et al., 2000; Stock and Watson, 2002) and Factor-Augmented VARs (FAVARs, see e.g. Bernanke et al., 2005). Indeed, Bayesian VARs are strictly connected to factor models as shown by De Mol et al. (2008) and Bańbura et al. (2015).

The link can be better understood in terms of data that have been transformed to

achieve stationarity, Δy_t, and that have been standardised to have zero mean and unit variance. A VAR in first differences can be written as

$$\Delta y_t = \Phi_1 \Delta y_{t-1} + \cdots + \Phi_p \Delta y_{t-p} + v_t. \tag{23}$$

Imposing the requirement that the level of each variable y_t must follow an independent random walk process is equivalent to requiring its first difference Δy_t to follow an independent white noise process. Hence, the prior on the autoregressive coefficients in Eq. (23) can be characterised by the following first and second moments:

$$\mathbb{E}\left[(\Phi_\ell)_{ij}|\Psi\right] = 0, \ \forall \ell \qquad \mathbb{V}ar\left[(\Phi_\ell)_{ij}|\Psi\right] = \begin{cases} \frac{\lambda_1^2}{f(\ell)} & \text{for } i = j, \forall \ell \\ \frac{\lambda_1^2}{f(\ell)} \frac{\Sigma_{ij}}{\omega_j^2} & \text{for } i \neq j, \forall \ell. \end{cases} \tag{24}$$

The covariance between coefficients at different lags is set to zero. Since the variables have been rescaled to have the same variance, we can set $\Sigma = \sigma \mathbb{I}_n$, where $\Sigma = \mathbb{E}[v_t v_t']$.

Denote the eigenvalues of the variance-covariance matrix of the standardised data by ζ_j, and the associated eigenvectors by ν_j, for $j = 1, \ldots, n$, i.e.

$$\left[\frac{1}{T}\sum_{t=1}^{T} \Delta y_t \Delta y_t'\right] \nu_j = \nu_j \zeta_j, \tag{25}$$

where $\nu_i' \nu_j = 1$ if $i = j$ and zero otherwise. We assume an ordering such that $\zeta_1 \geq \zeta_2 \geq \cdots \geq \zeta_n$. The sample principal components of Δy_t are defined as

$$z_t = \left[\frac{\nu_1}{\sqrt{\zeta_1}} \ \cdots \ \frac{\nu_n}{\sqrt{\zeta_n}}\right]' y_t \equiv W \Delta y_t \ . \tag{26}$$

The principal components transform correlated data, Δy_t, into linear combinations which are cross-sectionally uncorrelated and have unit variance, i.e. $T^{-1}\sum_{t=1}^{T} z_t z_t' = \mathbb{I}_n$. The principal components can be ordered according to their ability to explain the variability in the data, as the total variance explained by each principal component is equal to ζ_j.

Rewrite the model in Eq. (23) in terms of the ordered principal components, as

$$\Delta y_t = \Phi_1 W^{-1} z_{t-1} + \cdots + \Phi_p W^{-1} z_{t-p} + v_t \ . \tag{27}$$

The priors that impose a uniform shrinkage on the parameters in Eq. (24) map into a non-uniform shrinkage on the parameters in Eq. (27):

$$\mathbb{E}\left[(\Phi_\ell W^{-1})_{ij}|\Psi\right] = 0, \ \forall \ell \quad \mathbb{V}ar\left[(\Phi_\ell W^{-1})_{ij}|\Psi\right] = \begin{cases} \frac{\lambda_1^2 \zeta_j}{f(\ell)} & \text{for } i = j, \forall \ell \\ \frac{\lambda_1^2 \zeta_j}{f(\ell)} \frac{\Psi_{ij}}{\omega_j^2} & \text{for } i \neq j, \forall \ell. \end{cases} \tag{28}$$

Importantly, the prior variance for the coefficients on the j-th principal component is proportional to its share of explained variance of the data ζ_j.

If the data are characterised by a factor structure, then, as n and T increase, ζ_j will go to infinity at a rate n for $j = 1, \ldots, r$ where r is the number of common factors. Conversely, $\zeta_{r+1}, \ldots, \zeta_n$ will grow at a slower rate, which cannot be faster than $n/\sqrt{T}$. If λ_1 is set such that it converges to zero at a rate that is faster than that for the smaller eigenvalues and slower than that for the largest eigenvalues, e.g. $\lambda_1 \propto \frac{\sqrt{T}}{n} \frac{1}{T^\varrho}$, with $0 < \varrho < 1/2$, then $\lambda_1 \zeta_j$ will go to infinity for $j = 1, \ldots, r$ and the prior on the coefficients associated with the first r principal components will become flat (see Bańbura et al., 2015). Conversely, the coefficients related to the principal components associated with the bounded eigenvalues will be shrunk to zero, since $\lambda_1 \zeta_j$ will go to zero for $j > r$.

De Mol et al. (2008) show that, if the data are generated by a factor model and λ_1 is set according to the rate described above, the point forecasts obtained by using shrinkage estimators converge to the unfeasible optimal forecasts that would be obtained if the common factors were observed.

5.2 Large SVARs, non-fundamentalness

One of the open problems in SVARs is the potential 'non-fundamentalness' of structural shocks for commonly employed VARs (a review on this issue is in Alessi et al. 2011).

Non-fundamentalness implies that the true structural shocks (i.e. e_t in Eq. 11) cannot be retrieved from current and past forecast errors of the VARs of choice (see Hansen and Sargent, 1980; Lippi and Reichlin, 1994). This situation arises when for example the econometrician does not have all the information available to economic agents, such as news about future policy actions. This is notoriously the case for fiscal shocks, as explained in Leeper et al. (2013). In this case, economic agents' expectations may not be based only on the current and past y_t, implying that the residuals of the reduced-form model (i.e. u_t in Eq. 1) are not the agents' expectation/forecast errors. As a consequence, the shocks of interest may not be retrieved from the forecast errors, and may be non-fundamental. A possible solution is to allow for noninvertible moving average (MA) components. A different strategy is to view non-fundamentalness as an omitted variables problem. In this respect BVARs (and factor models) can offer a solution to the incorporation of larger information sets. For example, Ellahie and Ricco (2017) discuss the use of large BVARs to study the propagation of government purchases shocks, while controlling for potential non-fundamentalness of shocks in small VARs.[13]

5.3   Forecasting in Data-Rich Environments

A research frontier is the application of Bayesian VARs to forecasting in data-rich environment, where the predictive content of large datasets (typically counting 100 or more variables) is exploited to forecast variables of interest. A recent survey is in Bok et al. (2017).

Bańbura et al. (2010) study the forecasting performance of large Bayesian VARs. They find that while it increases with model size – provided that the shrinkage is appropriately chosen as a function of n –, most of the gains are in fact achieved by a 20-variable VAR. Evaluation of the forecasting performance of medium and large Bayesian VARs is also provided in Koop (2013). Carriero et al. (2011) evaluate the forecasting accuracy of reduced-rank Bayesian VARs in large datasets. The reduced-rank model adopted has a

---

[13]Lütkepohl (2014) has observed that while large information techniques can be of help in dealing with the problem, they are bound to distort the parameter estimates and also the estimated impulse responses, hence results have to been taken with some caution.

factor model underlying structure, with factors that evolve following a VAR. Koop and Korobilis (2013) extend the framework to allow for time-varying parameters. Giannone et al. (2017) argue in favour of dense (as opposed to sparse) representations of predictive models for economic forecasting and use a 'spike-and-slab' prior that allows for both variable selection and shrinkage.

BVARs are also a valuable tool for real-time forecasting and nowcasting with mixed-frequency datasets. In fact, they can be cast in state-space form and filtering techniques can be easily used to handle missing observations, data in real time, and data sampled at different frequencies. Ghysels (2016) introduces a class of mixed-frequency VAR models that incorporates data sampled at different frequencies and discusses Bayesian approaches to the estimation of these models. Recent examples of these applications include Korobilis (2013), Schorfheide and Song (2015); Carriero et al. (2015b); Brave et al. (2016); Clark (2011); Giannone et al. (2014); McCracken et al. (2015).

Koop et al. (2016) propose the use of Bayesian compressed VARs for high dimensional forecasting problems, and find that these tend to outperform both factor models and large VAR with prior shrinkage. More recently, Kastner and Huber (2017) develop BVARs that can handle vast dimensional information set and also allow for changes in the volatility of the error variance. This is done by assuming that the reduced-form residuals have a factor stochastic volatility structure (which allows for conditional equation-by-equation estimation) and by applying a Dirichlet-Laplace prior (Bhattacharya et al., 2015) to the VAR coefficients that heavily shrinks the coefficients towards zero while still allowing for some non-zero parameters. Kastner and Huber (2017) provide MCMC-based algorithms to sample from the posterior distributions and show that their proposed model typically outperforms simpler nested alternatives in forecasting output, inflation and the interest rate.

References

Aastveit, Knut Are, James Mitchell, Francesco Ravazzolo, and Herman K. van Dijk (2018) "The Evolution of Forecast Density Combinations in Economics," in Jonathan Hamilton ed. *Oxford Research Encyclopedia in Economics and Finance*: Oxford University Press.

Aastveit, Knut Are, Francesco Ravazzolo, and Herman K. Van Dijk (2017) "Combined Density Nowcasting in an Uncertain Economic Environment," *Journal of Business & Economic Statistics*, Vol. 0, No. 0, pp. 1–15.

Alessi, Lucia, Matteo Barigozzi, and Marco Capasso (2011) "Non-Fundamentalness in Structural Econometric Models: A Review," *International Statistical Review*, Vol. 79, No. 1, pp. 16–47, 04.

Amisano, Gianni and John Geweke (2017) "Prediction Using Several Macroeconomic Models," *The Review of Economics and Statistics*, Vol. 99, No. 5, pp. 912–925.

Andersson, Michael K., Stefan Palmqvist, and Daniel Waggoner (2010) "Density-Conditional Forecasts in Dynamic Multivariate Models," Working Paper Series 247, Sveriges Riksbank (Central Bank of Sweden).

Antolin-Diaz, Juan, Ivan Petrella, and Juan Francisco Rubio-Ramírez (2018) "Structural Scenario Analysis with SVARs," CEPR Discussion Papers 12579, C.E.P.R. Discussion Papers.

Bańbura, Marta, Domenico Giannone, and Michele Lenza (2015) "Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections," *International Journal of Forecasting*, Vol. 31, No. 3, pp. 739 – 756.

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010) "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, Vol. 25, No. 1, pp. 71–92.

Baumeister, Christiane and James D. Hamilton (2015) "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica*, Vol. 83, No. 5, pp. 1963–1999, September.

———— (2017) "Inference in Structural Vector Autoregressions When the Identifying Assumptions are Not Fully Believed: Re-evaluating the Role of Monetary Policy in Economic Fluctuations," mimeo, UCSD.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz (2005) "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *The Quarterly Journal of Economics*, Vol. 120, pp. 387–422.

Bhattacharya, Anirban, Debdeep Pati, Natesh S. Pillai, and David B. Dunson (2015) "Dirichlet-Laplace priors for optimal shrinkage," *Journal of the American Statistical Association*, Vol. 110, No. 512, pp. 1479–1490.

Billio, Monica, Roberto Casarin, Francesco Ravazzolo, and Herman K. van Dijk (2013) "Time-varying combinations of predictive densities using nonlinear filtering," *Journal of Econometrics*, Vol. 177, No. 2, pp. 213 – 232. Dynamic Econometric Modeling and Forecasting.

Bok, Brandyn, Daniele Caratelli, Domenico Giannone, Argia M. Sbordone, and Andrea Tambalotti (2017) "Macroeconomic nowcasting and forecasting with big data," Staff Reports 830, Federal Reserve Bank of New York.

Brave, Scott, R. Andrew Butters, and Alejandro Justiniano (2016) "Forecasting Economic Activity with Mixed Frequency Bayesian VARs," Working Paper Series WP-2016-5, Federal Reserve Bank of Chicago.

Canova, Fabio and Matteo Ciccarelli (2004) "Forecasting and turning point predictions in a Bayesian panel VAR model," *Journal of Econometrics*, Vol. 120, No. 2, pp. 327–359, June.

———— (2009) "Estimating Multicountry VAR Models," *International Economic Review*, Vol. 50, No. 3, pp. 929–959.

Canova, Fabio and Gianni De Nicolo (2002) "Monetary disturbances matter for business fluctuations in the G-7," *Journal of Monetary Economics*, Vol. 49, No. 6, pp. 1131–1159, September.

Canova, Fabio and Fernando J. Pérez Forero (2015) "Estimating overidentified, nonrecursive, time-varying coefficients structural vector autoregressions," *Quantitative Economics*, Vol. 6, No. 2, pp. 359–384.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2015a) "Bayesian VARs: Specification Choices and Forecast Accuracy," *Journal of Applied Econometrics*, Vol. 30, No. 1, pp. 46–73.

———— (2015b) "Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 178, No. 4, pp. 837–862.

———— (2016a) "Common Drifting Volatility in Large Bayesian VARs," *Journal of Business & Economic Statistics*, Vol. 34, No. 3, pp. 375–390.

———— (2016b) "Large Vector Autoregressions with Stochastic Volatility and Flexible Priors," Working Paper 1617, Federal Reserve Bank of Cleveland.

Carriero, Andrea, George Kapetanios, and Massimiliano Marcellino (2009) "Forecasting exchange rates with a large Bayesian VAR," *International Journal of Forecasting*, Vol. 25, No. 2, pp. 400–417.

———— (2011) "Forecasting large datasets with Bayesian reduced rank multivariate models," *Journal of Applied Econometrics*, Vol. 26, No. 5, pp. 735–761, August.

Chan, Joshua C. C., Eric Eisenstat, Chenghan Hou, and Gary Koop (2017) "Composite Likelihood Methods for Large Bayesian VARs with Stochastic Volatility," mimeo, University of Strathclyde.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (1999) "Monetary policy shocks: What have we learned and to what end?" in *Handbook of Macroeconomics*, Vol. 1, Part A: Elsevier, pp. 65 – 148.

Clark, Todd E. (2011) "Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility," *Journal of Business & Economic Statistics*, Vol. 29, No. 3, pp. 327–341.

Cogley, Timothy, Sergei Morozov, and Thomas J. Sargent (2005) "Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system," *Journal of Economic Dynamics and Control*, Vol. 29, No. 11, pp. 1893 – 1925. Expectations, learning, and monetary policy.

Cuaresma, Jesús Crespo, Martin Feldkircher, and Florian Huber (2016) "Forecasting with Global Vector Autoregressive Models: a Bayesian Approach," *Journal of Applied Econometrics*, Vol. 31, No. 7, pp. 1371–1391.

De Mol, Christine, Domenico Giannone, and Lucrezia Reichlin (2008) "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?" *Journal of Econometrics*, Vol. 146, No. 2, pp. 318 – 328. Honoring the research contributions of Charles R. Nelson.

Del Negro, Marco, Raiden B. Hasegawa, and Frank Schorfheide (2016) "Dynamic prediction pools: An investigation of financial frictions and forecasting performance," *Journal of Econometrics*, Vol. 192, No. 2, pp. 391 – 405. Innovations in Multiple Time Series Analysis.

Doan, Thomas, Robert Litterman, and Christopher Sims (1984) "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, Vol. 3, No. 1, pp. 1–100.

Ellahie, Atif and Giovanni Ricco (2017) "Government purchases reloaded: Informational insufficiency and heterogeneity in fiscal VARs," *Journal of Monetary Economics*, Vol. 90, No. C, pp. 13–27.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin (2000) "The Generalized Dynamic-Factor Model: Identification and Estimation," *Review of Economics and Statistics*, Vol. 82, No. 4, pp. 540–554.

Geweke, John (1999) "Using simulation methods for bayesian econometric models: inference, development,and communication," *Econometric Reviews*, Vol. 18, No. 1, pp. 1–73.

Geweke, John and Gianni Amisano (2011) "Optimal prediction pools," *Journal of Econometrics*, Vol. 164, No. 1, pp. 130 – 141. Annals Issue on Forecasting.

———— (2012) "Prediction with Misspecified Models," *American Economic Review*, Vol. 102, No. 3, pp. 482–86, May.

Geweke, John and Charles Whiteman (2006) "Bayesian Forecasting," in G. Elliott, C.W.J. Granger, and A. Timmermann eds. *Handbook of Economic Forecasting*, Vol. 1: Elsevier, pp. 3 – 80.

Ghysels, Eric (2016) "Macroeconomics and the reality of mixed frequency data," *Journal of Econometrics*, Vol. 193, No. 2, pp. 294–314.

Giacomini, Raffaella and Toru Kitagawa (2015) "Robust inference about partially identified SVARs," mimeo, UCL.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri (2017) "Economic Predictions with Big Data: The Illusion Of Sparsity," CEPR Discussion Papers 12256, C.E.P.R. Discussion Papers.

Giannone, Domenico, Michele Lenza, Daphne Momferatou, and Luca Onorante (2014) "Short-term inflation projections: A Bayesian vector autoregressive approach," *International Journal of Forecasting*, Vol. 30, No. 3, pp. 635–644.

Hall, Stephen G. and James Mitchell (2007) "Combining density forecasts," *International Journal of Forecasting*, Vol. 23, No. 1, pp. 1 – 13.

Hansen, Lars Peter and Thomas J. Sargent (1980) "Formulating and estimating dynamic linear rational expectations models," *Journal of Economic Dynamics and Control*, Vol. 2, No. 1, pp. 7–46, May.

Hwang, Youngjin (2017) "Forecasting with Specification-Switching VARs," *Journal of Forecasting*, Vol. 36, No. 5, pp. 581–596. for.2455.

Jarociński, Marek (2010) "Conditional forecasts and uncertainty about forecast revisions in vector autoregressions," *Economics Letters*, Vol. 108, No. 3, pp. 257–259, September.

Kadane, Joseph B. (1975) "The Role of Identification in Bayesian Theory," in Stephen E. Fienberg and Arnold Zellner eds. *Studies in Bayesian Econometrics and Statistics*: Amsterdam: North-Holland, Chap. 5.2, pp. 175–191.

Kadiyala, Rao K. and Sune Karlsson (1993) "Forecasting with generalized bayesian vector auto regressions," *Journal of Forecasting*, Vol. 12, No. 3-4, pp. 365–378.

———— (1997) "Numerical Methods for Estimation and Inference in Bayesian VAR-Models," *Journal of Applied Econometrics*, Vol. 12, No. 2, pp. 99–132, March-Apr.

Karlsson, Sune (2013) "Forecasting with Bayesian Vector Autoregression," in Graham Elliott and Allan Timmermann eds. *Handbook of Economic Forecasting*, Vol. 2 of Handbook of Economic Forecasting: Elsevier, pp. 791 – 897.

Kastner, G. and F. Huber (2017) "Sparse Bayesian vector autoregressions in huge dimensions,"Technical report, WU Vienna University of Economics and Business.

Kilian, L. and H. Lütkepohl (2017) *Structural Vector Autoregressive Analysis*, Themes in Modern Econometrics: Cambridge University Press.

Koop, Gary (2013) "Forecasting with Medium and Large Bayesian VARS," *Journal of Applied Econometrics*, Vol. 28, No. 2, pp. 177–203, March.

———— (2017) "Bayesian methods for empirical macroeconomics with big data," *Review of Economic Analysis*, Vol. 9, pp. 33–56, March.

Koop, Gary and Dimitris Korobilis (2012) "Forecasting Inflation using Dynamic Model Averaging," *International Economic Review*, Vol. 53, No. 3, pp. 867–886.

———— (2013) "Large time-varying parameter VARs," *Journal of Econometrics*, Vol. 177, No. 2, pp. 185–198.

———— (2015) "Forecasting With High Dimensional Panel VARs," Working Papers 25, Business School - Economics, University of Glasgow.

Koop, Gary, Dimitris Korobilis, and Davide Pettenuzzo (2016) "Bayesian Compressed Vector Autoregressions," Working Papers 103, Brandeis University, Department of Economics and International Businesss School.

Korobilis, Dimitris (2013) "VAR Forecasting Using Bayesian Variable Selection," *Journal of Applied Econometrics*, Vol. 28, No. 2, pp. 204–230.

Leeper, Eric M., Christopher A. Sims, and Tao Zha (1996) "What Does Monetary Policy Do?," *Brookings Papers on Economic Activity*, Vol. 27, No. 2, pp. 1–78.

Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang (2013) "Fiscal Foresight and Information Flows," *Econometrica*, Vol. 81, No. 3, pp. 1115–1145, May.

Lippi, Marco and Lucrezia Reichlin (1994) "VAR analysis, nonfundamental representations, blaschke matrices," *Journal of Econometrics*, Vol. 63, No. 1, pp. 307–325, July.

Lütkepohl, Helmut (2014) "Fundamental Problems with Nonfundamental Shocks," in Niels Haldrup, Mika Meitz, and Pentti Saikkonen eds. *Essays in Nonlinear Time Series Econometrics*: Oxford University Press, Chap. 8, pp. 198–214.

McCracken, Michael W. and Serena Ng (2015) "FRED-MD: A Monthly Database for Macroeconomic Research," Working Papers 2015-12, Federal Reserve Bank of St. Louis.

McCracken, Michael W., Michael T. Owyang, and Tatevik Sekhposyan (2015) "Real-Time Forecasting with a Large, Mixed Frequency, Bayesian VAR," Working Papers 2015-30, Federal Reserve Bank of St. Louis.

Miranda-Agrippino, Silvia and Giovanni Ricco (2018) "Bayesian Vector Autoregressions: Estimation," in Jonathan Hamilton ed. *Oxford Research Encyclopedia in Economics and Finance*: Oxford University Press.

Moon, Hyungsik Roger and Frank Schorfheide (2012) "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, Vol. 80, No. 2, pp. 755–782, March.

Pesaran, M. Hashem, Til Schuermann, and Scott M Weiner (2004) "Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model," *Journal of Business & Economic Statistics*, Vol. 22, No. 2, pp. 129–162.

Poirier, Dale J. (1998) "Revising Beliefs in Nonidentified Models," *Econometric Theory*, Vol. 14, No. 4, pp. 483–509.

Raftery, Adrian E., Miroslav Kárný, and Pavel Ettler (2010) "Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill," *Technometrics*, Vol. 52, No. 1, pp. 52–66.

Robertson, John C., Ellis W. Tallman, and Charles H. Whiteman (2005) "Forecasting Using Relative Entropy," *Journal of Money, Credit and Banking*, Vol. 37, No. 3, pp. 383–401.

Rubio-Ramírez, Juan F., Daniel F. Waggoner, and Tao Zha (2010) "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *The Review of Economic Studies*, Vol. 77, No. 2, pp. 665–696.

Schorfheide, Frank and Dongho Song (2015) "Real-Time Forecasting With a Mixed-Frequency VAR," *Journal of Business & Economic Statistics*, Vol. 33, No. 3, pp. 366–380, July.

Sims, Christopher A. (1993) "A Nine-Variable Probabilistic Macroeconomic Forecasting Model," in *Business Cycles, Indicators and Forecasting*: National Bureau of Economic Research, Inc, pp. 179–212.

Sims, Christopher A. and Tao Zha (1998) "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, Vol. 39, No. 4, pp. 949–68, November.

——— (1999) "Error Bands for Impulse Responses," *Econometrica*, Vol. 67, No. 5, pp. 1113–1156, September.

Sims, Christopher, James Stock, and Mark Watson (1990) "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, Vol. 58, No. 1, pp. 113–44.

Stock, James H. and Mark W. Watson (2002) "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, Vol. 20, No. 2, pp. 147–162.

Timmermann, Allan (2006) "Forecast Combinations," in G. Elliott, C. Granger, and A. Timmermann eds. *Handbook of Economic Forecasting*, Vol. 1: Elsevier, Chap. 4, pp. 135–196.

Todd, Richard M. (1984) "Improving economic forecasting with Bayesian vector autoregression," *Quarterly Review*, Vol. 8, No. 4.

Uhlig, Harald (2005) "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, Vol. 52, No. 2, pp. 381–419, March.

Waggoner, Daniel F. and Tao Zha (2012) "Confronting model misspecification in macroeconomics," *Journal of Econometrics*, Vol. 171, No. 2, pp. 167 – 184. Bayesian Models, Methods and Applications.