Bayesian Vector Autoregressions

Silvia Miranda-Agrippino*
Bank of England
Centre for Macroeconomics (LSE)

Giovanni Ricco†
University of Warwick
OFCE - SciencesPo

Revised 2 September 2018

Summary

Vector Autoregressions (VARs) are linear multivariate time-series models able to capture the joint dynamics of multiple time series. Bayesian inference treats the VAR parameters as random variables, and it provides a framework to estimate 'posterior' probability distribution of the location of the model parameters by combining information provided by a sample of observed data and prior information derived from a variety of sources, such as other macro o micro datasets, theoretical models, other macroeconomic phenomena, or introspection.

In empirical work in Economics and Finance, informative prior probability distributions are often adopted. These are intended to summarise stylised representations of the data generating process. For example, 'Minnesota' priors, one of the most commonly adopted macroeconomic priors for the VAR coefficients, express the belief that an independent random-walk model for each variable in the

*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`

This is a draft of an article that has been accepted for publication by Oxford University Press in the forthcoming Economics and Finance Oxford Research Encyclopedia edited by Jonathan Hamilton, and due for publication in 2018.

Acknowledgments. We thank the Editor of the Oxford Research Encyclopedia and an anonymous referee for useful comments and suggestions. We are also grateful to Fabio Canova, Andrea Carriero, Matteo Ciccarelli, Domenico Giannone, Marek Jarociński, Marco del Negro, Massimiliano Marcellino, Giorgio Primiceri, Lucrezia Reichlin, and Frank Shorfheide for helpful comments and discussions. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of England or any of its Committees.

system is a reasonable 'centre' for the beliefs about their time series behaviour. Other commonly adopted priors, the 'single-unit-root' and the 'sum-of-coefficients' priors are used to enforce beliefs about relations among the VAR coefficients, such as for example the existence of co-integrating relationships among variables, or of independent unit-roots.

Priors for macroeconomic variables are often adopted as 'conjugate prior distributions' – i.e. distributions that yields a posterior distribution in the same family as the prior p.d.f. –, in the form of Normal-Inverse-Wishart distributions that are conjugate prior for the likelihood of a VAR with normally distributed disturbances. Conjugate priors allow direct sampling from the posterior distribution and fast estimation. When this is not possible, numerical techniques such as Gibbs and Metropolis-Hastings sampling algorithms are adopted.

Bayesian techniques allow for the estimation of an ever-expanding class of sophisticated autoregressive models that includes conventional fixed-parameters VAR models; Large VARs incorporating hundreds of variables; Panel VARs, that permit analysing the joint dynamics of multiple time series of heterogeneous and interacting units; and VAR models that relax the assumption of fixed coefficients, such as Time-Varying Parameters, Threshold, and Markov Switching VARs.

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 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

Vector Autoregressions (VARs) are linear multivariate time-series models able to capture the joint dynamics of multiple time series. The pioneering work of Sims (1980) proposed to replace the large-scale macroeconomic models popular in the 1960s with VARs, and suggested that Bayesian methods could have improved upon frequentist ones in estimating the model coefficients. Bayesian VARs (BVARs) with macroeconomic variables were first employed in forecasting by Litterman (1979) and Doan et al. (1984). Since then, VARs and BVARs have been a standard macroeconometric tool routinely used by scholars and policy makers for structural analysis, forecasting and scenario analysis in an ever growing number of applications.

The aim of this article is to review key ideas and contributions in the BVAR literature. A companion paper provides a brief survey of applications of BVARs in Economics and Finance, such as forecasting, scenario analysis and structural identification (Miranda-Agrippino and Ricco, 2018). An exhaustive survey of the literature is beyond the scope of this article due to space limitations. Readers are referred to a number of monographs and more detailed surveys available on different topics in the BVARs literature.[1,2]

Differently from frequentist statistics, Bayesian inference treats the VAR parameters as random variables, and provides a framework to update probability distributions about

[1]Several books provide excellent in-depth treatments of Bayesian inference. Among others, Zellner (1971), Gelman et al. (2003), Koop (2003) and Geweke (2005). Canova (2007) provides a book treatment of VARs and BVARs in the context of the methods for applied macroeconomic research. Several recent articles survey the literature on BVARs. Del Negro and Schorfheide (2011) have a deep and insightful discussion of BVAR with a broader focus on Bayesian macroeconometrics and DSGE models. Koop and Korobilis (2010) propose a discussion of Bayesian multivariate time series models with an in-depth discussion of time-varying parameters and stochastic volatility models. Geweke and Whiteman (2006a) and Karlsson (2013b) provide a detailed survey with a focus on forecasting with Bayesian Vector Autoregression. Ciccarelli and Rebucci (2003) survey BVARs in forecasting analysis with Euro Area data. Canova and Ciccarelli (2009, 2013) discuss panel Bayesian VARs. Finally, the reader is referred to Timmermann (2006) for an in-depth discussion on model averaging and forecast combination, a natural extension of the Bayesian framework.

[2]Dieppe et al. (2016) have developed the ready-to-use BEAR toolbox that implements many of the methods described in this article (`https://www.ecb.europa.eu/pub/research/working-papers/html/bear-toolbox.en.html`). Other useful code sources are those related to the books of Kroese and Chan (2014) (`http://joshuachan.org/code.html`) and Koop and Korobilis (2010) (`https://sites.google.com/site/dimitriskorobilis/matlab`).

the unobserved parameters conditional on the observed data. By providing such a framework, the Bayesian approach allows to incorporate prior information about the model parameters into post-sample probability statements. The 'prior' distributions about the location of the model parameters summarise pre-sample information available from a variety of sources, such as other macro o micro datasets, theoretical models, other macroeconomic phenomena, or introspection.

In the absence of pre-sample information, Bayesian VAR inference can be thought of as adopting 'non-informative' (or 'diffuse' or 'flat') priors, that express complete ignorance about the model parameters, in the light of the sample evidence summarised by the likelihood function (i.e. the probability density function of the data as a function of the parameters). Often, in such a case, Bayesian probability statements about the unknown parameters (conditional on the data) are very similar to classical confidence statements about the probability of random intervals around the true parameters value. For example, for a VAR with Gaussian errors and a flat prior on the model coefficients, the posterior distribution is centred at the maximum likelihood estimator (MLE), with variance given by the variance-covariance matrix of the residuals. Section 2 discusses inference in BVARs and 'non-informative' priors.

While non-informative priors can provide a useful benchmark, in empirical work with macroeconomic and financial variables informative priors are often adopted. In scientific data analysis, priors on the model coefficients do not incorporate the investigator's 'subjective' beliefs, instead, they summarise stylised representations of the data generating process. Conditional on a model, these widely held standardised priors aim at making the likelihood-based description of the data useful to investigators with potentially diverse prior beliefs (Sims, 2010b).[3]

The most commonly adopted macroeconomic priors for VARs are the the so-called 'Minnesota' priors (Litterman, 1980). They express the belief that an independent random-walk model for each variable in the system is a reasonable 'centre' for the be-

[3]Bayesian priors can often be interpreted as frequentist penalised regressions (see, for example, De Mol et al., 2008). A Gaussian prior for the regression coefficients, for example, can be thought of as a Ridge penalised regression. Having a double exponential (Laplace) prior on the coefficients is instead equivalent to a Lasso regularisation problem.

liefs about their time series behaviour. While not motivated by economic theory, they are computationally convenient priors, meant to capture commonly held beliefs about how economic time series behave. Minnesota priors can be cast in the form of a Normal-Inverse-Wishart (NIW) prior, which is the conjugate prior for the likelihood of a VAR with normally distributed disturbances (see Kadiyala and Karlsson, 1997). Conjugate priors are such that the posterior distribution belongs to the same family as the prior probability distribution. Hence, they allow for analytical tractability of the posterior, and computational speed. Because the data is incorporated into the posterior distribution only through the sufficient statistics, formulas for updating the prior into the posterior are in this case conveniently simple. It is often useful to think of the parameters of a prior distribution – known as 'hyperparameters' – as corresponding to having observed a certain number of 'dummy' or 'pseudo-' observations with properties specified by the prior beliefs on the VAR parameters. Minnesota priors can be formulated in terms of artificial data featuring pseudo observations for each of the regression coefficients, and that directly assert the prior on them.

Dummy observations can also implement prior beliefs about relations among the VAR coefficients, such as e.g. co-integration among variables. In this case, commonly used priors are formulated directly as linear joint stochastic restrictions among the coefficients.[4] This is, for example, the case of the 'single-unit root' prior, that is centred on a region of the VAR parameter space where either there is no intercept and the system contains at least one unit root, or the system is stationary and close to its steady state at the beginning of the sample (Sims, 1993).[5] Another instance in which dummy observations are used to establish relations among several coefficients is the 'sum-of-coefficients' prior, that incorporates the widely shared prior beliefs that economic variables can be represented by a process with unit roots and weak cross-sectional linkages (Litterman,

[4]In principle, dummy observations can also implement prior beliefs about nonlinear functions of the parameters (a short discussion on this is in Sims, 2005b).

[5]Such a prior is adopted to capture the belief that it is not plausible to assume that initial transients can explain a large part of observed long-run variation in economic time series. Since in a sample of given size there is no information on the behaviour of time series at frequencies longer than the sample size, the prior assumptions implicitly or explicitly elicited in the analysis will inform results. This is a clear example, in the inference in VARs, of an issue for which Bayesian inference provides a framework to make prior information explicit and available to scientific discussion on the inference in VAR models.

1979).[6] Section 3 discusses some of the priors commonly adopted in the economic literature.

The hyperparameters can be either fixed using prior information (and sometimes 'unorthodoxly' using sample information), or associated to hyperprior distributions that express beliefs about their values. A Bayesian model with more than one level of priors is called a hierarchical Bayes model. In empirical macroeconomic modelling, the hyperparameters associated with the informativeness of the prior beliefs (i.e. the tightness of the prior distribution) are usually left to the investigator's judgement. In order to select a value for these hyperparameters, the VAR literature has adopted mostly heuristic methodologies that minimise pre-specified loss functions over a pre-sample (e.g. the out-of-sample mean squared forecast error in Litterman, 1979, or the in-sample fit in Bańbura et al., 2010). Conversely, Giannone et al. (2015) specify hyperprior distributions and choose the hyperparameters that maximise their posterior probability distribution conditional on the data. Section 4 discusses hierarchical modelling and common approaches to choose hyperparameters not specified by prior information.

In Section 5 we discuss Bayesian inference in VAR models that relax the assumption of fixed coefficients in order to capture changes in the time series dynamics of macroeconomic and financial variables, such as VARs with autoregressive coefficients, Threshold and Markov Switching VARs. In Section 6 we review Panel Bayesian VARs that generalise VAR models by describing the joint dynamics of multiple time series of potentially heterogenous and interacting units – as for examples, the economies of several countries, regions, or sectors.

Bayesian Vector Autoregressions (BVARs) have been applied to an increasingly large number of empirical problems. 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

[6]Several sets of pseudo-observations can be adopted at the same time. In fact, successive dummy observations modify the prior distribution as if they reflected successive observations of functions of the VAR parameters, affected by stochastic disturbances.

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 7 discusses forecasting with BVARs, while Section 8 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).[7] Section 9 reviews Bayesian techniques in SVARs.

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 10.

[7]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, 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 Inference in BVARs

Vector Autoregressions (VARs) are linear stochastic models that describe the joint dynamics of multiple time series. Let y_t be an $n \times 1$ random vector that takes values in $\mathbb{R}_n$. The evolution of y_t – the endogenous variables – is described by a system of p-th order difference equations – the VAR(p):

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

In Eq. (67), A_j, $j = 1, \ldots, p$ are $n \times n$ matrices of autoregressive coefficients, c is a vector of n intercepts, and u_t is an n-dimensional vector of one-step-ahead forecast errors, or reduced-form innovations. The vector of stochastic innovations, u_t, is an independent and identically distributed random variable for each t. The distribution from which u_t is drawn determines the distribution of y_t, conditional on its past $y_{1-p:t-1} \equiv \{y_{1-p}, \ldots, y_0, \ldots, y_{t-2}, y_{t-1}\}$. The standard assumption in the macroeconometric literature is that errors are Gaussian

$$u_t \sim i.i.d. \ \mathcal{N}(0, \Sigma) \ . \tag{2}$$

This implies that also the conditional distribution of y_t is Normal.[8,9]

Bayesian inference on the model in Eq. (67) amounts to updating prior beliefs about the VAR parameters, that are seen as stochastic variables, after having observed a sample $y_{1-p:t} \equiv \{y_{1-p}, \ldots, y_0, \ldots, y_{t-2}, y_t\}$. Prior beliefs about the VAR coefficients are

[8]While the assumption of normally distributed errors makes the posterior p.d.f. tractable, modern computational methods permit straightforward characterisation of posterior distributions obtained under different assumptions. Among others, Chiu et al. (2017) and Panagiotelis and Smith (2008) depart from the normality assumption and allow for t-distributed errors.

[9]It is interesting to observe that in large samples, and under certain regularity conditions, the likelihood function converges to a Gaussian distribution, with mean at the maximum likelihood estimator (MLE) and covariance matrix given by the usual MLE estimator for the covariance matrix. This implies that conditioning on the MLE and using its asymptotic Gaussian distribution is, approximately in large samples, as good as conditioning on all the data (see discussion in Sims, 2010b).

summarised by a probability density function (p.d.f.), and updated using Bayes' Law

$$p(A, \Sigma | y_{1-p:t}) = \frac{p(A, \Sigma) p(y_{1-p:t} | A, \Sigma)}{p(y_{1-p:t})} \propto p(A, \Sigma) p(y_{1-p:t} | A, \Sigma) \ , \tag{3}$$

where we define $A \equiv [A_1, \ldots, A_p, c]'$ as a $k \times n$ matrix, with $k = np + 1$. The joint posterior distribution of the VAR(p) coefficients $p(A, \Sigma | y_{1-p:t})$ incorporates the information contained in the prior distribution $p(A, \Sigma)$ – summarising the initial information about the model parameters –, and the sample information summarised by $p(y_{1-p:t} | A, \Sigma)$. Viewed as a function of the parameters, the sample information is the likelihood function.[10] The posterior distribution summarises the entire information available, and is used to conduct inference on the VAR parameters.

Given the autoregressive structure of the model, and the i.i.d. innovations, the (conditional) likelihood function of the sample observations $y_{1:T}$ – conditional on A, Σ and on the first p observations $y_{1-p:0}$ –, can be written as the product of the conditional distribution of each observation

$$p\left(y_{1:T} | A, \Sigma, y_{1-p:0}\right) = \prod_{t=1}^{T} p\left(y_t | A, \Sigma, y_{t-p:t-1}\right). \tag{4}$$

Under the assumption of Gaussian errors, the conditional likelihood of the VAR in Eq. (67) is

$$p\left(y_{1:T} | A, \Sigma, y_{1-p:0}\right) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{n/2}} |\Sigma|^{-1} \exp\left\{ -\frac{1}{2} \left(y_t - A' x_t\right)' \Sigma^{-1} \left(y_t - A' x_t\right) \right\} \ , \tag{5}$$

where $x_t' \equiv \left[\begin{array}{cccc} y_{t-1}' & \ldots & y_{t-p}' & 1 \end{array} \right]$.

The likelihood in Eq. (5) can be written in compact form, by using the seemingly unrelated regression (SUR) representation of the VAR

$$y = xA + u, \tag{6}$$

[10] The marginal p.d.f. for the observations, denoted as $p(y_{1-p:t})$, is a normalising constant and as such can be dropped when making inference about the model parameters.

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

$$y = \begin{bmatrix} y_1' \\ \vdots \\ y_T' \end{bmatrix}, \quad x = \begin{bmatrix} x_1' \\ \vdots \\ x_T' \end{bmatrix}, \quad u = \begin{bmatrix} u_1' \\ \vdots \\ u_T' \end{bmatrix}. \tag{7}$$

Using this notation and standard properties of the trace operator, the conditional likelihood function can be equivalently expressed as

$$\begin{aligned} p\left(y_{1:T}|A, \Sigma, y_{1-p:0}\right) = \frac{1}{(2\pi)^{Tn/2}}|\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\widehat{S}\right]\right\} \\ \times \exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}(A-\widehat{A})'x'x(A-\widehat{A})\right]\right\}, \end{aligned} \tag{8}$$

where $\widehat{A}$ is the maximum-likelihood estimator (MLE) of A, and $\widehat{S}$ the matrix of sums of squared residuals, i.e.

$$\widehat{A} = (x'x)^{-1}x'y, \qquad \widehat{S} = (y - x\widehat{A})'(y - x\widehat{A}). \tag{9}$$

The likelihood can also be written in terms of the vectorised representation of the VAR

$$\mathbf{y} = (\mathbb{I}_n \otimes x)\alpha + \mathbf{u}, \qquad \mathbf{u} \sim (0, \Sigma \otimes \mathbb{I}_T), \tag{10}$$

where $\mathbf{y} \equiv vec(y)$ and $\mathbf{u} \equiv vec(u)$ are $Tn \times 1$ vectors, and $\alpha \equiv vec(A)$ is $nk \times 1$. In this vectorised notation the likelihood function is written as

$$\begin{aligned} p\left(y_{1:T}|A, \Sigma, y_{1-p:0}\right) = \frac{1}{(2\pi)^{Tn/2}}|\Sigma|^{-T/2} \exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\widehat{S}\right]\right\} \\ \times \exp\left\{-\frac{1}{2}(\alpha-\hat{\alpha})'[\Sigma^{-1} \otimes (x'x)](\alpha-\hat{\alpha})\right\}, \end{aligned} \tag{11}$$

where, consistently, $\hat{\alpha} \equiv vec(\widehat{A})$ is $nk \times 1$. Detailed derivations for the multivariate Gaussian linear regression model can be found in Zellner (1971).

Given the likelihood function, Eq. (3) is used to update the prior information regarding the VAR parameters. An interesting case arises when we assume the absence of any information on the location of the model parameters. This setting can be formalised by assuming that α and Σ are independently distributed, i.e.,

$$p(\alpha, \Sigma) = p(\alpha)p(\Sigma), \tag{12}$$

with prior p.d.f.

$$\begin{aligned} p(\alpha) &\propto const., \\ p(\Sigma) &\propto |\Sigma|^{-(n+1)/2} . \end{aligned} \tag{13}$$

These priors are known as diffuse or Jeffreys' prior (Geisser, 1965; Tiao and Zellner, 1964). Jeffreys priors are proportional to the square root of the determinant of the Fisher information matrix, and are derived from the Jeffreys' 'invariance principle', meaning that the prior is invariant to re-parameterization (see Zellner, 1971).[11]

Given this set of priors, it is straightforward to derive the posterior distribution of the VAR parameters as

$$\begin{aligned} & p\left(A, \Sigma | y_{1:T}\right) \propto \\ & \quad |\Sigma|^{-(T+n+1)/2} \exp\left\{-\frac{1}{2} tr\left(\Sigma^{-1} \otimes \mathbb{I}_T\right)\left[\mathbf{y} - \left(\mathbb{I}_n \otimes x\right)\hat{\alpha}\right]'\left[\mathbf{y} - \left(\mathbb{I}_n \otimes x\right)\hat{\alpha}\right]\right\} \\ & \qquad \times \exp\left\{-\frac{1}{2}\left(\alpha - \hat{\alpha}\right)'\left(\Sigma^{-1} \otimes x'x\right)\left(\alpha - \hat{\alpha}\right)\right\} , \end{aligned} \tag{14}$$

where the proportionality factor has been dropped for convenience.

From the joint posterior in Eq. (14) one can readily deduce the form of the posterior for α, conditional on Σ and the observed sample. Also, the posterior can be integrated

[11] 'Non-informative' or 'flat' priors are designed to extract the maximum amount of expected information from the data. They maximise the difference (measured by Kullback-Leibler distance) between the posterior and the prior when the number of samples drawn goes to infinity. Jeffreys priors for VARs are 'improper', in the sense that they do not integrate to one over the parameter space. Hence, they cannot be thought of as well specified p.d.f. distributions. However, they can be obtained as degenerate limit of the Normal-Inverse-Wishart conjugate distribution, and their posterior is proper. For an in-depth discussion on non-informative priors in multi-parameter settings see Zellner (1971) and Bernardo and Smith (2009).

over α to obtain the marginal posterior for Σ. Therefore, it is possible to conveniently write the posterior distribution of the parameters as

$$p(\alpha, \Sigma | y_{1:T}) = p(\alpha | \Sigma, y_{1:T}) p(\Sigma | y_{1:T}) \tag{15}$$

where

$$\Sigma | y \sim \mathcal{IW}\left((y - x\widehat{A})'(y - x\widehat{A}), T - k\right) \tag{16}$$

$$\alpha | \Sigma, y \sim \mathcal{N}\left(\hat{\alpha}, \Sigma \otimes (x'x)^{-1}\right) \ . \tag{17}$$

Hence, given the diffuse priors on α and Σ, the posterior for the autoregressive coefficients is centred at the MLE, with posterior variance $\Sigma \otimes (x'x)^{-1}$.[12] Interestingly, in this standard normal multivariate linear regression model, Bayesian probability statements about the parameters (given the data) have the same form as the frequentist pre-sample probability statements about the parameters' estimator (see also Sims, 2010b). This is a more general property, in fact, Kwan (1998) has shown that, under widely applicable regularity conditions, an estimator $\hat{\alpha}_T$ for which

$$\sqrt{T}(\hat{\alpha}_T - \alpha) | \alpha \xrightarrow[T \to \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma)$$

allows, with high accuracy, to approximate the distribution of $\sqrt{T}(\alpha - \hat{\alpha}_T) | \hat{\alpha}$ as $\mathcal{N}(0, \Sigma)$ in large samples. Hence, it is often possible to interpret $(1 - \rho)$ approximate confidence sets generated from the frequentist asymptotic approximate distribution as if they were sets in the parameter space with posterior probability $(1 - \rho)$.

In potentially misspecified models for which linear regression coefficients are the object of interest, Müller (2013) proposes to adopt an artificial Gaussian posterior centred at the MLE but with a sandwich estimator for the covariance matrix. In fact, in the case of a misspecified model, the shape of the likelihood (the posterior) is asymptotically

[12] The marginal posterior distribution of the $k \times n$ matrix A is matricvariate t (see Kadiyala and Karlsson, 1997)

$$A | \mathbf{y} \propto |(y - x\widehat{A})'(y - x\widehat{A}) + (A - \widehat{A})'x'x(A - \widehat{A})|^{-T/2} \ . \tag{18}$$

Gaussian and centred at the MLE, but of a different variance than the asymptotically normal sampling distribution of the MLE. This argument can be seen as a 'flipping' of the frequentist asymptotic statement that supports the use of a sandwich estimator for the covariance matrix in misspecified models, in line with the results in Kwan (1998).[13]

An important case in which frequentist pre-sample probability statements and Bayesian post-sample probability statements about parameters diverge, is the case of time-series regression models with unit roots. In such cases, while the frequentist distribution of the estimator is skewed asymptotically, the likelihood, and hence the posterior p.d.f., remain unaffected (see Sims and Uhlig, 1991; Kim, 1994).

3 Informative Priors for Reduced-Form VARs

Informative prior probability distributions incorporate information about the VAR parameters that is available before some sample is observed. Such prior information can be contained in samples of past data – from the same or a related system –, or can be elicited from introspection, casual observation, and theoretical models. The first case is sometimes referred to as a 'data-based' prior, while the second as a 'nondata-based' prior.

An important case arises when the prior probability distribution yields a posterior distribution for the parameters in the same family as the prior p.d.f. In this case the prior is called a natural conjugate prior for the likelihood function (Raiffa and Schlaifer, 1961). In general, it has been shown that exponential distributions are the only class of distributions that admit a natural conjugate prior, due to these having a fixed number of sufficient statistics that does not increase as the sample size T increases (see e.g. Gelman et al., 2013). Because the data is incorporated into the posterior distribution

[13]Müller (2013) shows that a Bayesian decision-maker can justify using OLS with a sandwich covariance matrix when the probability limit of the OLS estimator is the object of interest, despite the fact that the linear regression model is known not to be the true model (see discussion in Sims, 2010b). Miranda-Agrippino and Ricco (2017) use this intuition to construct coverage bands for impulse responses estimated with Bayesian Local Projections (BLP). This method can be thought of as a generalisation of BVARs that estimates a different model for different forecast horizons – as in direct forecasts – and hence induces autocorrelation in the reduced-form residuals that violate the the i.i.d. assumption in Eq. (61).

only through the sufficient statistics, formulas for updating the prior into the posterior are in these cases conveniently simple.

Prior distributions can be expressed in terms of coefficients, known as hyperparameters, whose functions are sufficient statistics for the model parameters. It is often useful to think of the hyperparameters of a conjugate prior distribution as corresponding to having observed a certain number of pseudo-observations with properties specified by the priors on the parameters. In general, for nearly all conjugate prior distributions, the hyperparameters can be interpreted in terms of 'dummy' or pseudo-observations. The basic idea is to add to the observed sample extra 'data' that express prior beliefs about the hyperparameters. The prior then takes the form of the likelihood function of these dummy observations. Hyperparameters can be either fixed using prior information, or associated to hyperprior distributions that express beliefs about their values. A Bayesian model with more than one level of priors is called a hierarchical Bayes model. In this section we review some of the most commonly used priors for VARs with macroeconomic and financial variables, while we discuss the choice of the hyperpriors and hierarchical modelling in Section 4.

3.1 Natural Conjugate Normal-Inverse Wishart Priors

The Normal-Inverse Wishart (NIW) conjugate priors, part of the exponential family, are commonly used prior distributions for (A, Σ) in VARs with Gaussian errors. These assume a multivariate normal distribution for the regression coefficients, and an Inverse Wishart specification for the covariance matrix of the error term, and can be written as

$$\Sigma \sim \mathcal{IW}(\underline{S},\ \underline{d}) \tag{19}$$

$$\alpha|\Sigma \sim \mathcal{N}(\underline{\alpha},\ \Sigma \otimes \underline{\Omega}) \ , \tag{20}$$

where $(\underline{S}, \underline{d}, \underline{\alpha}, \underline{\Omega})$ are the priors' hyperparameters. $\underline{d}$ and $\underline{S}$ denote, respectively, the degrees of freedom and the scale of the prior Inverse-Wishart distribution for the variance-covariance matrix of the residuals. $\underline{\alpha}$ is the prior mean of the VAR coefficients, and

$\underline{\Omega}$ acts as a prior on the variance-covariance matrix of the dummy regressors.[14] The posterior distribution can be analytically derived and is given by

$$\Sigma|\mathbf{y} \sim \mathcal{IW}(\overline{S},\ \overline{d}) \tag{21}$$

$$\alpha|\Sigma, \mathbf{y} \sim \mathcal{N}(\overline{\alpha},\ \Sigma \otimes \overline{\Omega}), \tag{22}$$

where

$$\overline{\Omega} = (\underline{\Omega} + x'x)^{-1}, \tag{23}$$

$$\overline{\alpha} \equiv vec(\overline{A}) = vec\left(\overline{\Omega}\ \left(\underline{\Omega}^{-1}\underline{A} + x'x\widehat{A}\right)\right), \tag{24}$$

$$\overline{S} = \widehat{A}'x'x\widehat{A} + \underline{A}'\underline{\Omega}^{-1}\underline{A} + \underline{S} + (y - x\widehat{A})'(y - x\widehat{A}) - \overline{A}'(\underline{\Omega}^{-1} + x'x)\overline{A}\ . \tag{25}$$

Comparing Eqs. (16) - (17) to Eqs. (19) - (20), it is evident that informative priors can be thought of as equivalent to having observed dummy observations (y_d, x_d) of size T_d, such that

$$\underline{S} = (y_d - x_d\underline{A})'(y_d - x_d\underline{A}), \tag{26}$$

$$\underline{d} = T_d - k, \tag{27}$$

$$\underline{\alpha} = vec(\underline{A}) = vec\left((x_d'x_d)^{-1}x_d'y_d\right), \tag{28}$$

$$\underline{\Omega} = (x_d'x_d)^{-1}\ . \tag{29}$$

This idea was first proposed for a classical estimator for stochastically restricted coefficients by Theil (1963). Once a set of pseudo-observations able to match the wished hyperparameters is found, the posterior can be equivalently estimated using the extended samples $y_* = [y', y_d']'$, $x_* = [x', x_d']'$ of size $T_* = T + T_d$ obtaining

$$\Sigma|\mathbf{y} \sim \mathcal{IW}\left(S_*,\ T_* + \underline{d}\right) \tag{30}$$

$$\alpha|\Sigma, \mathbf{y} \sim \mathcal{N}\left(\alpha_*,\ \Sigma \otimes (x_*'x_*)^{-1}\right). \tag{31}$$

[14]The prior mean of the VAR coefficients is $\mathbb{E}[\alpha] = \underline{\alpha}$, for $\underline{d} > n$, while the variance is $\mathbb{V}ar[\alpha] = (\underline{d} - n - 1)^{-1}\underline{S} \otimes \underline{\Omega}$, for $\underline{d} > n + 1$. Setting $\underline{d} = \max\{n + 2, n + 2h - T\}$ ensures that both the prior variances of A and the posterior variances of the forecasts at $T + h$ are defined.

Indeed, it is easy to verify that the posterior moments obtained with the starred variables coincide with those in Eqs. (21) - (22). The posterior estimator efficiently combines sample and prior information using their precisions as weights in the spirit of the mixed estimation of Theil and Goldberger (1961). Posterior inference can be conducted via direct sampling.

Algorithm 1: Direct Monte Carlo Sampling from Posterior of VAR Parameters.

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

1. Draw $\Sigma^{(s)}$ from the Inverse-Wishart distribution $\Sigma|\mathbf{y} \sim \mathcal{IW}(S_*,\ T_* + \underline{d})$.

2. Draw $A^{(s)}$ from the Normal distribution of $A^{(s)}|\Sigma^{(s)}, \mathbf{y} \sim \mathcal{N}\left(\alpha_*,\ \Sigma^{(s)} \otimes (x_*' x_*)^{-1}\right)$.

When it is not possible to sample directly from the posterior distribution, as in this case, Markov chain Monte Carlo (MCMC) algorithms are usually adopted (see e.g. Chib, 2001).[15]

An important feature of the NIW priors in Eqs. (19) - (20) is the Kronecker factorisation that appears in the Gaussian prior for α. As discussed in the previous section, because the same set of regressors appears in each equation, homoskedastic VARs can be written as SUR models. This symmetry across equations means that homoskedastic VAR models have a Kronecker factorisation in the likelihood, which in turn implies that estimation can be broken into n separate least-squares calculations, each only of dimension $np + 1$. The symmetry in the likelihood can be inherited by the posterior, if the prior adopted also features a Kronecker structure as in Eq. (20). This is a desirable

[15]The key idea of MCMC algorithms is to construct a Markov chain for $\theta \equiv (A, \Sigma)$ which has the posterior as its (unique) limiting stationary distribution, and such that random draws can be sampled from the transition kernel $p(\theta^{(s+1)}|\theta^{(s)})$. Tierney (1994) and Geweke (2005) discuss the conditions for the convergence of the chain to the posterior distribution when starting from an arbitrary point in the parameter space. Typically, a large number of initial draws (known as burn-in sample) is discarded to avoid including portions of the chain which have not yet converged to the posterior. Also, even if convergent, the chain may move very slowly in the parameter space due to e.g. autocorrelation between the draws, and a very large number of draws may be needed. See also Karlsson (2013a) for a discussion on this point and on empirical diagnostic tests to assess the chain convergence. References include Geweke (1999); Chib and Greenberg (1995); Geweke and Whiteman (2006b).

property that guarantees tractability of the posterior p.d.f. and computational speed. However, such a specification can result in unappealing restrictions and may not fit the actual prior beliefs one has – see discussions in Kadiyala and Karlsson (1997), and Sims and Zha (1998). In fact, it forces symmetry across equations, because the coefficients of each equation have the same prior variance matrix (up to a scale factor given by the elements of Σ). There may be situations in which theory suggests 'asymmetric restrictions' may be desirable instead, e.g. money neutrality implies that the money supply does not Granger-cause real output.[16] Also, the Kronecker structure implies that prior beliefs must be correlated across the equations of the reduced form representation of the VAR, with a correlation structure that is proportional to that of the disturbances.

3.2 Minnesota Prior

In macroeconomic and financial applications, the parameters of the NIW prior in Eqs. (19) - (20) are often chosen so that prior expectations and variances of A coincide with the so-called 'Minnesota' prior, that was originally proposed in Litterman (1980, 1986).[17] The basic intuition behind this prior is that the behaviour of most macroeconomic variables is well approximated by a random walk with drift. Hence, it 'centres' the distribution of the coefficients in A at a value that implies a random-walk behaviour for all the elements in y_t

$$y_t = c + y_{t-1} + u_t. \tag{32}$$

While not motivated by economic theory, these are computationally convenient priors, meant to capture commonly held beliefs about how economic time series behave.

[16]Such restrictions can be accommodated by replacing Eq. (19) with a truncated Normal distribution. In this case, however, posterior moments are not available analytically and must be evaluated numerically, with consequential complications and loss of efficiency with respect to the MCMC algorithm discussed above (see Hajivassiliou and Ruud, 1994; Kadiyala and Karlsson, 1997, for further details).

[17]The original formulation of Litterman (1980)'s prior was of the form

$$\alpha \sim \mathcal{N}(\underline{\alpha}, \underline{\Gamma}),$$

where $\underline{\Gamma} \equiv diag([\underline{\gamma}_1^2, \ldots, \underline{\gamma}_n^2])$ is assumed to be fixed, known, and diagonal. Highfield (1992) and Kadiyala and Karlsson (1997) observed that by modifying Litterman's prior to make it symmetric across equations in the form of a NIW prior, the posterior p.d.f. was tractable.

The Minnesota prior assumes the coefficients $A_1, \ldots, A_p$ to be a priori independent and normally distributed, with the following moments

$$\mathbb{E}\left[(A_\ell)_{ij}|\Sigma\right] = \begin{cases} \delta_i & i = j,\ \ell = 1 \\ 0 & \text{otherwise} \end{cases} \qquad \mathbb{V}ar\left[(A_\ell)_{ij}|\Sigma\right] = \begin{cases} \frac{\lambda_1^2}{f(\ell)} & \text{for } i = j, \forall \ell \\ \frac{\lambda_1^2}{f(\ell)} \frac{\Sigma_{ij}}{\omega_j^2} & \text{for } i \neq j, \forall \ell. \end{cases} \tag{33}$$

In Eq. (33), $(A_\ell)_{ij}$ denotes the coefficient of variable j in equation i at lag ℓ. In the original formulation of the prior $\delta_i = 1$, in accordance with Eq. (32). The random-walk assumption, however, may not be appropriate if the variables in y_t were characterised by substantial mean-reversion. For stationary series, or series that have been transformed to achieve stationarity, Bańbura et al. (2010) centre the distribution around zero (i.e. $\delta_i = 0$). The prior also assumes that lags of other variables are less informative than own lags, and that most recent lags of a variable tend to be more informative than more distant lags. This intuition is formalised with $f(\ell)$. A common choice for this function is a harmonic lag decay – i.e. $f(\ell) = \ell^{\lambda_2}$, a special case of which is $f(\ell) = \ell$ –, where the severity of the lag decay is regulated by the hyperparameter λ_2. The factor Σ_{ij}/ω_j^2 accounts for the different scales of variables i and j. The hyperparameters ω_j^2 are often fixed using sample information, for example from univariate regressions of each variable onto its own lags.

Importantly, λ_1 is a hyperparameter that controls the overall tightness of the random walk prior. If $\lambda_1 = 0$ the prior information dominates, and the VAR reduces to a vector of univariate models. Conversely, as $\lambda_1 \to \infty$ the prior becomes less informative, and the posterior mostly mirrors sample information. We discuss the choice of the free hyperparameters in Section 4.

The Minnesota prior can be implemented using dummy observations. Priors on the A coefficients are implemented via the following pseudo-observations

$$\begin{aligned} y_d^{(1)} &= \begin{bmatrix} diag([\delta_1\omega_1, \ldots, \delta_n\omega_n])/\lambda_1 \\ 0_{n(p-1)\times n} \end{bmatrix}, \\ x_d^{(1)} &= \begin{bmatrix} J_p \otimes diag([\omega_1, \ldots, \omega_n])/\lambda_1 & 0_{np\times 1} \end{bmatrix}, \end{aligned} \tag{34}$$

where $J_p = diag([1^{\lambda_2}, 2^{\lambda_2}, \ldots, p^{\lambda_2}])$ with geometric lag decay.[18] To provide intuition on how the prior is implemented using artificial observations, we consider the simplified case of a $n = 2,\ p = 2$ VAR for the pseudo-observations . The first n rows of Eq. (34) impose priors on A_1; that is, on the coefficients of the first lag. In the $n = 2,\ p = 2$ case one obtains,

$$\begin{pmatrix} \frac{\delta_1\omega_1}{\lambda_1} & 0 \\ 0 & \frac{\delta_2\omega_2}{\lambda_1} \end{pmatrix} = \begin{pmatrix} \frac{\omega_1}{\lambda_1} & 0 & 0 & 0 & 0 \\ 0 & \frac{\omega_2}{\lambda_1} & 0 & 0 & 0 \end{pmatrix} A + \begin{pmatrix} (u_d^{(1)})_{1,1} & (u_d^{(1)})_{2,1} \\ (u_d^{(1)})_{1,2} & (u_d^{(1)})_{2,2} \end{pmatrix} \tag{35}$$

that implies, for example, the following equations for the elements $(1,1)$ and $(1,2)$ of A_1

$$\frac{\delta_1\omega_1}{\lambda_1} = \frac{\omega_1}{\lambda_1}(A_1)_{1,1} + (u_d^{(1)})_{1,1} \implies (A_1)_{1,1} \sim \mathcal{N}\left(\delta_1, \frac{\Sigma_{1,1}\lambda_1^2}{\omega_1^2}\right),$$

$$0 = \frac{\omega_1}{\lambda_1}(A_1)_{2,1} + (u_d^{(1)})_{2,1} \implies (A_1)_{2,1} \sim \mathcal{N}\left(0, \frac{\Sigma_{2,1}\lambda_1^2}{\omega_1^2}\right).$$

Similar restrictions are obtained for the elements the elements $(2,1)$ and $(2,2)$ of A_1. The following $(n-1)p$ rows in Eq. (34) implement priors on the coefficients of the other lags. In fact, we readily obtain

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \frac{2^{\lambda_2}\omega_1}{\lambda_1} & 0 & 0 \\ 0 & 0 & 0 & \frac{2^{\lambda_2}\omega_2}{\lambda_1} & 0 \end{pmatrix} A + \begin{pmatrix} (u_d^{(1)})_{1,1} & (u_d^{(1)})_{2,1} \\ (u_d^{(1)})_{1,2} & (u_d^{(1)})_{2,2} \end{pmatrix} \tag{36}$$

which for example implies the following restriction for the element $(1,1)$ of A_2

$$0 = \frac{2^{\lambda_2}\omega_1}{\lambda_1}(A_2)_{1,1} + (u_d^{(1)})_{1,1} \implies (A_2)_{1,1} \sim \mathcal{N}\left(0, \frac{\Sigma_{1,1}\lambda_1^2}{2^{2\lambda_2}\omega_1^2}\right).$$

Similar restrictions obtain for the other elements of A_2. Priors beliefs on the residual

[18] Given the dummy observations in Eq. (34), the matrix Ω in Eq. (19) is diagonal and of the form

$$\underset{[k\times k]}{\Omega} = ((x_d^{(1)})' x_d^{(1)})^{-1} = diag\left(\left[\frac{\lambda_1^2}{\omega_1^2}, \ldots, \frac{\lambda_1^2}{\omega_n^2}, \frac{\lambda_1^2}{2^{2\lambda_2}\omega_1^2}, \ldots, \frac{\lambda_1^2}{2^{2\lambda_2}\omega_n^2}, \ldots, \frac{\lambda_1^2}{p^{2\lambda_2}\omega_1^2}, \ldots, \frac{\lambda_1^2}{p^{2\lambda_2}\omega_n^2}, 0\right]\right).$$

covariance matrix Σ can instead be implemented by the following block of dummies

$$y_d^{(2)} = \left[\begin{array}{c} 1_{\lambda_3 \times 1} \otimes diag([\omega_1, \ldots, \omega_n]) \end{array} \right] \quad (37)$$

$$x_d^{(2)} = \left[\begin{array}{cc} 0_{\lambda_3 n \times np} & 0_{\lambda_3 n \times 1} \end{array} \right] . \quad (38)$$

In the $n = 2$, $p = 2$ case, they correspond to appending to the VAR equations λ_3 replications of

$$\begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} A + \begin{pmatrix} (u_d^{(2)})_{1,1} & (u_d^{(2)})_{2,1} \\ (u_d^{(2)})_{1,2} & (u_d^{(2)})_{2,2} \end{pmatrix} . \quad (39)$$

λ_3 is the hyperparameter that determines the tightness of the prior on Σ. To understand how this works, it is sufficient to consider that given λ_3 artificial observations z_i, with $z_i \sim \mathcal{N}(0, \sigma_z^2)$, an estimator for the covariance is given by $\lambda_3^{-1} \sum_{i=1}^{\lambda_3} z_i^2$.

Finally, uninformative priors for the intercept are often implemented with the following set of pseudo-observations

$$y_d^{(3)} = \left[\begin{array}{c} 0_{1 \times n} \end{array} \right], \qquad x_d^{(3)} = \left[\begin{array}{cc} 0_{1 \times np} & \epsilon \end{array} \right],$$

where ϵ is a hyperparameter usually set to a very small number.[19]

3.3 Priors for VAR with Unit Roots and Trends

Sims (1996, 2000) observed that flat-prior VARs, or more generally estimation methods that condition on initial values, tend to attribute an implausibly large share of the variation in observed time series to deterministic – and hence entirely predictable – components. The issue stems from the fact that ML and OLS estimators that condition on the initial observations and treat them as non-stochastic do not apply any penalisation to parameters values that imply that these observations are very distant from the variables' steady state (or their trend if non-stationary). As a consequence,

[19] Canova (1992, 1993) propose a set of artificial observations to account for seasonal patterns and potentially other peaks in the spectral densities.

complex transient dynamics from the initial conditions to the steady state are treated as plausible, and can explain an 'implausibly' large share of the low-frequency variation of the data. This typically translates into poor out-of-sample forecasts. To understand the intuition, consider the univariate model

$$y_t = c + ay_{t-1} + u_t. \tag{40}$$

Iterating Eq. (40) backward yields

$$y_t = \left[a^t y_0 + \sum_{j=0}^{t-1} a^j c\right] + \left[\sum_{j=0}^{t-1} a^j u_{t-j}\right], \tag{41}$$

which, if $|a| < 1$, reduces to

$$y_t = \left[a^t \left(y_0 - \frac{c}{1-a}\right) + \frac{c}{1-a}\right] + \left[\sum_{j=0}^{t-1} a^j u_{t-j}\right]. \tag{42}$$

The first term in square brackets in Eq. (41) is the deterministic component: the evolution of y_t from the initial conditions y_0, absent any shocks. The second term instead captures the stochastic evolution of y_t due to the shocks realised between $[0, t-1]$. $c/(1-a)$ in Eq. (42) is the unconditional mean of y_t. If y_t is close to non-stationary – i.e. $a \simeq 1$ –, the MLE estimator of the unconditional mean of y_t may be very far from y_0, and the 'reversion to the mean' from y_0 is then used to fit the data (see Eq. 42).

One way to deal with this issue is to use the unconditional likelihood, by explicitly incorporating the density of the initial observations in the inference. However, because most macroeconomic time series are effectively nonstationary, it is not obvious how the density of the initial observations should be specified.[20] Another approach, following Sims and Zha (1998); Sims (2000), is to instead specify priors that downplay the im-

[20]This approach requires the use of iterative nonlinear optimisation methods. The main issue with this approach is that nonstationary models have no unconditional – viz. ergodic – distribution of the initial conditions. Also, while near-nonstationary models may have an ergodic distribution, the time required to arrive at the ergodic distribution from arbitrary initial conditions may be very long. For this reason, using such a method requires strong beliefs about the stationarity of the model, which is rarely the case in macroeconomics, and imposing the ergodic distribution on the first p observations may be unreasonable (see Sims, 2005a).

portance of the initial observations, and hence reduce the explanatory power of the deterministic component.

These types of priors, implemented through artificial observations, aim to reduce the importance that the deterministic component has in explaining a large share of the in-sample variation of the data, eventually improving forecasting performances out-of-sample (see Sims, 1996; Sims and Zha, 1998, for a richer discussion on this point).[21]

The 'co-persistence' (or 'single-unit-root' or 'dummy initial observation') prior (Sims, 1993) reflects the belief that when all lagged y_t's are at some level $\bar{y}_0$, y_t tends to persist at that level. It is implemented using the following artificial observation

$$\underset{[1\times n]}{y_d^{(4)}} = \left[\frac{\bar{y}_{0,1}}{\lambda_4}, \ldots, \frac{\bar{y}_{0,n}}{\lambda_4}\right] \qquad \underset{[1\times k]}{x_d^{(4)}} = [y_d^{(4)}, \ldots, y_d^{(4)}, 1/\lambda_4], \tag{43}$$

where $\bar{y}_{0,i},\ i = 1, \ldots, n$ are the average of the initial values of each variable, and usually set to be equal to the average of the first p observations in the sample, and $k = np + 1$. Writing down the implied system of equations $y_d^{(4)} = Ax_d^{(4)} + u_d^{(4)}$ one obtains the following stochastic restriction on the VAR coefficients

$$[\mathbb{I}_n - A(1)]\, \bar{y}_0 - c = \lambda_4 u_d^{(4)}, \tag{44}$$

where $\mathbb{I}_n - A(1) = (\mathbb{I}_n - A_1 - \ldots - A_p)$. The hyperparameter λ_4 controls the tightness of this stochastic constraint. The prior is uninformative for $\lambda_4 \to \infty$. Conversely, as $\lambda_4 \to 0$ the model tends to a form where either there is at least one explosive common unit root and the constant c is equal to zero ($\bar{y}_0$ is the eigenvector of the unit root), or the VAR is stationary, c is different from zero, and the initial conditions are close to the implied unconditional mean ($\bar{y}_0 = [\mathbb{I}_n - A(1)]^{-1}\, c$). In the stationary form, this prior does not rule out cointegrated models. This prior induces prior correlation among all the VAR coefficients in each equation, including the constant.[22]

[21]The treatment of unit root in Bayesian and frequentist inference has been hotly debated. Among others, important contributions are Sims (1988, 1991), Sims and Uhlig (1991), Koop and Steel (1991), Phillips (1991a,b), Uhlig (1994a,b), Müller and Elliott (2003); Jarociński and Marcet (2011, 2014). The Journal of Applied Econometrics October/December 1991 Volume 6, Issue 4 has been entirely dedicated to this debate.

[22]To put a heavier weight on the presence of a unit root, one could add to the observation in Eq.

The 'sums-of-coefficients' (or 'no-cointegration') prior (Doan et al., 1984), captures the belief that when the average lagged values of a variable $y_{j,t}$ is at some level $\bar{y}_{0,j}$, then $\bar{y}_{0,j}$ is likely to be a good forecast of $y_{j,t}$. It also implies that knowing the average of lagged values of variable j does not help in predicting a variable $i \neq j$. This prior is implemented using n artificial observations, one for each variable in y_t

$$\underset{[n\times n]}{y_d^{(5)}} = diag\left(\left[\frac{\bar{y}_{0,1}}{\lambda_5}, \ldots, \frac{\bar{y}_{0,n}}{\lambda_5}\right]\right) \qquad\qquad \underset{[n\times k]}{x_d^{(5)}} = [y_d^{(5)}, \ldots, y_d^{(5)}, 0]. \tag{45}$$

The prior implied by these dummy observations is centred at 1 for the sum of coefficients on own lags for each variable, and at 0 for the sum of coefficients on other variables' lags. It also introduces correlation among the coefficients of each variable in each equation. In fact, it is easy to show that equation by equation this priors implies the stochastic constraint

$$[1 - (A_1)_{jj} - \ldots - (A_p)_{jj}]\, \bar{y}_{0,j} = \lambda_5 (u_d^{(5)})_j \qquad \forall j \ , \tag{46}$$

where $(A_\ell)_{jj}$ denotes the coefficient of variable j in equation j at lag ℓ. The hyperparameter λ_5 controls the variance of these prior beliefs. As $\lambda_5 \to \infty$ the prior becomes uninformative, while $\lambda_5 \to 0$ implies that each variable is an independent unit-root process, and there are no co-integration relationships.[23]

The Bayesian analysis of cointegrated VARs is an active area of research, (a detailed survey is in Koop et al. 2006).[24] Giannone et al. (2016) elicit theory-based priors for the long run of persistent variables which shrink towards a random walk those linear

(43) an additional artificial observation that enforces the belief that $c = 0$. Alternatively, one could modify Eq. (43) to have a zero in place of λ_4^{-1} as the observation corresponding to the intercept. In this case, the prior gives no plausibility to stationary models and, if used in isolation, allows for at least a single unit root without any restriction on c. Hence, despite the presence of a unit root, it may not necessarily reduce the importance of the deterministic component (see Sims, 2005a).

[23] The sums-of-coefficients observations of Eq. (45) do not imply any restriction on the vector of intercepts c, since the artificial observations loading on the constant are set to zero. Therefore, this prior allows for a non-zero constant, and hence for a linearly trending drift. To assign smaller probability to versions of the model in which deterministic transient components are much more important than the error term in explaining the series variance, one has to add to Eq. (45) artificial observations that favour $c = 0$ (see Sims, 2005a).

[24] Among many others, contributions to the treatment of cointegration in Bayesian VARs are in Kleibergen and van Dijk (1994), Geweke (1996), Villani (2001), Kleibergen and Paap (2002), Strachan and Inder (2004), Koop et al. (2011), Jochmann and Koop (2015).

combination of variables that are likely to have a unit root. Conversely, combinations which are likely to be stationary (i.e. cointegrating relationships among variables) are shrunk towards stationary processes. Operationally, this is achieved by rewriting the VAR in Eq. (67) as

$$\begin{aligned}\Delta y_t &= \Pi y_{t-1} + P_1 \Delta y_{t-1} + \ldots + P_p \Delta y_{t-p+1} + c + \xi_t \\ &= \Pi F^{-1} F y_{t-1} + P_1 \Delta y_{t-1} + \ldots + P_p \Delta y_{t-p+1} + c + \xi_t,\end{aligned} \tag{47}$$

where $\Pi = A_1 + \ldots + A_p - \mathbb{I}_n$, $P_j = -(A_{j+1} + \ldots + A_p)$, and F is any invertible n-dimensional matrix. The problem is then specified as setting a prior for $\widetilde{\Pi} \equiv \Pi F^{-1}$, conditional on a specific choice of F. F defines the relevant linear combinations of the variables in y_t which macroeconomic theory suggest to be a priori stationary or otherwise.

Another alternative is in Villani (2009). Here the VAR is written as

$$y_t = \rho_0 + \rho_1 t + \tilde{y}_t, \qquad \tilde{y}_t = A_1 \tilde{y}_{t-1} + \ldots + A_p \tilde{y}_{t-p} + u_t, \qquad u_t \sim i.i.d.\mathcal{N}(0, \Sigma) \tag{48}$$

where ρ_0 and ρ_1 are $n \times 1$ vectors. The first term, $\rho_0 + \rho_1 t$, captures a linear deterministic trend of y_t, whereas the law of motion of $\tilde{y}_t$ captures stochastic fluctuations around the deterministic trend, which can be either stationary or non-stationary. This alternative specification allows to separate beliefs about the deterministic trend component from beliefs about the persistence of fluctuations around this trend. Let $A = [A_1, \ldots, A_p]'$ and $\rho = [\rho_0', \rho_1']'$. It can be shown that if the prior distribution of ρ conditional on A and Σ is Normal, the (conditional) posterior distribution of ρ is also Normal (see also Del Negro and Schorfheide, 2011, for details). Hence, posterior inference can be implemented via Gibbs sampling.

3.4 Priors from Structural Models

DeJong et al. (1993), Ingram and Whiteman (1994), Del Negro and Schorfheide (2004) have proposed the use of priors for VARs that are derived from Dynamic Stochastic

General Equilibrium (DSGE) models. This approach bridges VARs and DSGEs by constructing families of prior distributions informed by the restrictions that a DSGE-model implies on the VAR coefficients. This modelling approach is sometimes referred to as DSGE-VAR. Ingram and Whiteman (1994) derive prior information from the basic stochastic growth model of King et al. (1988) and report that a BVAR based on the Real Business Cycle model prior outperforms a BVAR with a Litterman prior in forecasting real economic activity. Del Negro and Schorfheide (2004) extend and generalise this approach, and show how to conduct policy simulations within this framework.

Schematically, the exercises can be thought of as follows. First, time-series are simulated from a DSGE model. Second, a VAR is estimated from these simulated data. Population moments of the simulated data computed from the DSGE model solution are considered in place of sample moments. Since the DSGE model depends on unknown structural parameters, hierarchical prior modelling is adopted by specifying a distribution on the DSGE model parameters. A tightness parameter controls the weight of the DSGE model prior relative to the weight of the actual sample. Finally, Markov Chain Monte Carlo methods are used to generate draws from the joint posterior distribution of the VAR and DSGE model parameters.

3.5 Priors for Model Selection

It is standard practice in VAR models to pre-select the relevant variables to be included in the system (and with how many lags). This procedure may be thought of as having dogmatic priors about which variables have non-zero coefficients in the system. The challenge is in selecting among an expansive set of potential models. Indeed, for a VAR with n endogenous variables, q additional potentially exogenous variables including a constant, and p lags, there are $2^{(q+pn)n+n(n-1)/2}$ possible models.

Jarociński and Maćkowiak (2017) propose to select the variables to be included in the system by systematically assessing the posterior probability of 'Granger causal priority' (Sims, 2010a) in a BVAR with conjugate priors. Granger causal priority answers questions of the form "Is variable z relevant for variable x, after controlling for other variables in the system?" The authors provide a closed form expression for the posterior probability of Granger causal priority, and suggest that variables associated with high Granger causal priority probabilities can be omitted from a VAR with the variables of interest.

Alternatively, one can adopt priors that support model selection and enforce sparsity. A variety of techniques, including double exponential (Laplace) prior, spike-and-slab prior, etc., have been adopted to handle this issue. Some recent theoretical and empirical contributions on this topic are in Mitchell and Beauchamp (1988), George et al. (2008), Koop (2013), Korobilis (2013), Bhattacharya et al. (2015a), Griffin and Brown (2010, 2017), Giannone et al. (2017), Huber and Feldkircher (2017).

4 Hyperpriors and Hierarchical Modelling

As seen in the previous section, the informativeness of prior beliefs on the VAR parameters often depends on a set of free hyperparameters. Let $\lambda \equiv [\lambda_1, \lambda_2, \ldots]$ denote the vector collecting all the hyperparameters not fixed using (pre)sample information, and θ denote all the VAR parameters, i.e. A and Σ. The prior distribution of θ is thus effectively $p_\lambda(\theta)$. Choosing a value for λ alters the tightness of the prior distribution, and hence determines how strictly the prior is enforced on the data.

In order to set the informativeness of the prior distribution of the VAR coefficients, the literature has initially used mostly heuristic methodologies. Litterman (1980) and Doan et al. (1984), for example, choose a value for the hyperparameters that maximises the out-of-sample forecasting performance over a pre-sample. Conversely, Bańbura et al. (2010) propose to choose the shrinkage parameters that yield a desired in-sample fit, in order to control for overfitting. Subsequent studies have then either used these as 'default' values, or adopted either one of these criteria. Robertson and Tallman (1999); Wright (2009); Giannone et al. (2014) opt for the first, while e.g. Giannone et al. (2008); Bloor and Matheson (2011); Carriero et al. (2009); Koop (2013) follow Bańbura et al. (2010).

In VARs, Giannone et al. (2015) observe that, from a purely Bayesian perspective, choosing λ is conceptually identical to conducting inference on any other unknown parameter of the model. Specifically, the model is interpreted as a hierarchical one (Berger, 1985; Koop, 2003) and λ can be chosen as the maximiser of

$$\begin{aligned} p(\lambda|\mathbf{y}) &\propto \int p(\mathbf{y}|\theta, \lambda, y_{1-p:0}) p(\theta|\lambda) d\theta \cdot p(\lambda) \\ &= p(\mathbf{y}|\lambda, y_{1-p:0}) \cdot p(\lambda) \ . \end{aligned} \tag{49}$$

This method is also known in the literature as the Maximum Likelihood Type II (ML-II) approach to prior selection (Berger, 1985; Canova, 2007). In Eq. (49), $p(\lambda|\mathbf{y})$ is the posterior distribution of λ conditional on the data, and $p(\lambda)$ denotes a prior probability density specified on the hyperparameters themselves, and also known as the hyperprior distribution. In such hierarchical model, the prior distribution for the VAR coefficients is treated as a conditional prior, that is $p_\lambda(\theta)$ is replaced by $p(\theta|\lambda)$. In the case of a NIW family of distributions, the prior structure becomes $p(\alpha|\Sigma, \lambda)p(\Sigma|\lambda)p(\lambda)$. $p(\mathbf{y}|\lambda, y_{1-p:0})$ is the marginal likelihood (ML), and is obtained as the density of the data as a function of λ, after integrating out all the VAR parameters. Conveniently, with conjugate priors the ML is available in closed form.

Conversely, the joint posterior of α, Σ and λ is not available in closed form. However, with NIW priors for θ, Giannone et al. (2015) set up the following Metropolis-Hasting sampler for the joint distribution

Algorithm 2: MCMC Sampler for a VAR with Hierarchical Prior.

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

1. Draw a candidate vector λ^* from the random walk distribution $\lambda^* \sim \mathcal{N}(\lambda^{s-1}, \kappa H^{-1})$, where H is the Hessian of the negative of the log-posterior at the peak for λ, and κ is a tuning constant. Choose

$$\lambda^{(s)} = \begin{cases} \lambda^* & \text{with probability } = \min\left\{1, \frac{p(\mathbf{y}|\lambda^*)}{p(\mathbf{y}|\lambda^{(s-1)})}\right\} \\ \lambda^{(s-1)} & \text{otherwise.} \end{cases}$$

2. Draw $\Sigma^{(s)}$ form the full conditional posterior $\Sigma|\mathbf{y}, \lambda^{(s)}$ in Eq. (21).

3. Draw $A^{(s)}$ from the full conditional posterior $A^{(s)}|\mathbf{y}, \Sigma^{(s)}, \lambda^{(s)}$ in Eq. (22).

In a similar fashion, Belmonte et al. (2014) apply a hierarchical structure to time-varying parameters (TVP) models and specify priors for Bayesian Lasso shrinkage parameters to determine whether coefficients in a forecasting model for inflation are zero, constant, or time-varying in a data driven way.

Carriero et al. (2015a) evaluate the forecasting performance of BVARs where tightness hyperparameters are chosen as the maximisers of Eq. (49) or rather set to default values and find that the former route yields modest but statistically significant gains in forecasting accuracy particularly at short horizons.

5 Time-Varying Parameter, State-Dependent, Stochastic Volatility VARs

Models that allow parameters to change over time are increasingly popular in empirical research, in recognition of the fact that they can capture structural changes in the economy. In fact, it seems to be a common belief that the properties of many (if not most) macroeconomic time series have changed over time, and can change across regimes or phases of the business cycle. Model parameters either change frequently and gradually over time according to a multivariate autoregressive process – as in e.g. in Time-Varying Parameters VARs (TVP-VARs) –, or they change abruptly and infrequently as in e.g. Markov-switching or structural-break models.

5.1 Time-varying parameters VAR (TVP-VAR)

Time-varying parameters VARs differ from fixed-coefficient VARs in that they allow the parameters of the model to vary over time, according to a specified law of motion.[25] TVP-VARs often include also stochastic volatility (SV), which allows for time variation

[25]Review articles are in Del Negro and Schorfheide (2011); Koop and Korobilis (2010); Lubik and Matthes (2015).

in the variance of the stochastic disturbances.[26] Doan et al. (1984) were first to show how estimation of a TVP-VAR with Litterman priors could be conducted by casting the VAR in state space form and using Kalman filtering techniques. This same specification is in Sims (1993). Bayesian time varying parameter VARs have become popular in empirical macroeconomics following the work of Cogley and Sargent (2002, 2005) and Primiceri (2005) who provided the foundations for Bayesian inference in these models, and used then innovations in MCMC algorithms to improve on their computational feasibility.

The basic TVP-VAR is of the form

$$y_t = A_{1,t} y_{t-1} + \ldots + A_{p,t} y_{t-p} + c_t + u_t \ , \tag{50}$$

where the constant coefficients of Eq. (67) are replaced by the time-varying $A_{j,t}$. Eq. (50) can be rewritten in compact form as

$$y_t = x_t A_t + u_t \ , \tag{51}$$

where x_t is defined as in Eq. (5), and $A_t = [A_{1,t}, \ldots, A_{p,t}, c_t]'$. It is common to assume that the coefficients follow a random-walk process

$$\alpha_t = \alpha_{t-1} + \varsigma_t \qquad \varsigma_t \sim i.i.d.\ \mathcal{N}(0, \Upsilon) \ , \tag{52}$$

where $\alpha_t \equiv vec(A_t)$. The covariance matrix Υ is usually restricted to be diagonal, and the innovations ς_t to be uncorrelated with u_t, with u_t distributed as in Eq. (61). The law of motion for α_t in Eq. (52) – i.e. the state equation –, implies that $\alpha_{t+1}|\alpha_t, \Upsilon \sim \mathcal{N}(\alpha_t, \Upsilon)$, which can be used as a prior distribution for α_{t+1}. Hence, the prior for all the states (i.e. α_t $\forall t$) is a product of normal distributions. For the initial vector of the VAR coefficients Cogley and Sargent (2002, 2005) use a prior of the form $\alpha_1 \sim \mathcal{N}(\underline{\alpha}_{1|0}, \underline{\Upsilon}_{1|0})$, where $\underline{\alpha}_{1|0}$ and $\underline{\Upsilon}_{1|0}$ are set by estimating a fixed-coefficient VAR with a flat prior on a pre-sample.[27] If the Gaussian prior for the states is complemented with IW priors for

[26] Stochastic volatility in Bayesian VARs was initially introduced in Uhlig (1997).

[27] See also the discussion in Karlsson (2013a) for additional details on the specification of the prior for α_t.

both Σ and Υ, then sampling from the joint posterior is possible with a Gibbs sampling algorithm

Algorithm 3: Gibbs Sampling from Posterior of TVP-VAR Parameters. Select starting values for $\Sigma^{(0)}$ and $\Upsilon^{(0)}$. For $s = 1, \dots, n_{sim}$:

1. Draw $\alpha_T^{(s)}$ from the full conditional posterior

$$\alpha_T^{(s)} | y_{1:T}, \Sigma^{(s-1)}, \Upsilon^{(s-1)} \sim \mathcal{N}(\alpha_{T|T}, \Upsilon_{T|T})$$

   obtained from the Kalman filter. For $t = T-1, \dots, 1$ draw $\alpha_t^{(s)}$ from the full conditional posterior

$$\alpha_t^{(s)} | y_{1:T}, \Sigma^{(s-1)}, \Upsilon^{(s-1)} \sim \mathcal{N}(\alpha_{t|T}, \Upsilon_{t|T})$$

   obtained from a simulation smoother.

2. Draw $\Upsilon^{(s)}$ from

$$\Upsilon^{(s)} | \alpha_{1:T}^{(s)} \sim \mathcal{IW}\left(\underline{S}_\Upsilon + \sum_{t=1}^{T} \left[\alpha_{t+1}^{(s)} - \alpha_t^{(s)} \right] \left[\alpha_{t+1}^{(s)} - \alpha_t^{(s)} \right]', \underline{d}_\Upsilon + T \right).$$

3. Draw $\Sigma^{(s)}$ from

$$\Sigma^{(s)} | \mathbf{y}, \alpha_{1:T}^{(s)} \sim \mathcal{IW}\left(\underline{S} + \sum_{t=1}^{T} \left[\mathbf{y} - (\mathbb{I}_n \otimes x)\alpha_t^{(s)} \right] \left[\mathbf{y} - (\mathbb{I}_n \otimes x)\alpha_t^{(s)} \right]', \underline{d} + T \right).$$

When stochastic volatility is added to the framework, the VAR innovations are assumed to be still normally distributed, but with variance that evolves over time (see Cogley and Sargent, 2002, 2005; Primiceri, 2005)

$$u_t \sim \mathcal{N}(0, \Sigma_t) , \qquad \Sigma_t = K^{-1} \Xi_t (K^{-1})' , \tag{53}$$

where K is a lower-triangular matrix with ones on the main diagonal, and Ξ_t a diagonal matrix with elements evolving following a geometric random-walk process

$$ln(\Xi_t)_j = ln(\Xi_{t-1})_j + \eta_{j,t} \qquad \eta_{j,t} \sim i.i.d.\ \mathcal{N}(0, \sigma^2_{\eta,j}) \ . \tag{54}$$

The prior distributions for Υ and $\sigma^2_{\eta,j}$ $j = 1, \ldots, n$ can be used to express beliefs about the magnitude of the period-to-period drift in the VAR coefficients, and the changes in the volatility of the VAR innovations respectively. In practice, these priors are chosen to ensure that innovations to the parameters are small enough that the short- and medium-run dynamics of y_t are not swamped by the random-walk behaviour of A_t and Ξ_t. Primiceri (2005) extends the above TVP-VAR by also allowing the nonzero off-diagonal elements of the contemporaneous covariance matrix K to evolve as random-walk processes (i.e. K is replaced by K_t to allow for an arbitrary time-varying correlation structure). A Gibbs sampler to draw from the posterior distribution of the parameters is in Primiceri (2005).

5.2 Markov Switching, Threshold, and Smooth Transition VARs

Contrary to the drifting coefficients models discussed in the previous section, Markov switching (MS) VARs are designed to capture abrupt changes in the dynamics of y_t.[28] These can be viewed as models that allow for at least one structural break to occur within the sample, with the timing of the break being unknown. They are of the form

$$y_t = A(s_t)x_t + u_t, \qquad u_t \sim \mathcal{N}(0, \Sigma(s_t)), \tag{55}$$

where x_t is defined as in Eq. (5). The matrix of autoregressive coefficients $A(s_t)$ and the variance of the error term $\Sigma(s_t)$ are a function of a discrete m-state Markov process s_t with fixed transition probabilities

$$\pi_{ij} \equiv p(s_t = \mathcal{S}_i | s_t = \mathcal{S}_j) \text{ for } i, j \in [1, \ldots, m]. \tag{56}$$

[28]The book by Kim and Nelson (1999) is the standard reference for frequentist and Bayesian estimation of Markov switching models.

If $\pi_{ii} = 1$ for some $i \in [1, \dots, m]$, then $\mathcal{S}_i$ is an absorbing state from which the system is not allowed to move away. Suppose $m = 2$, and that both $A(s_t)$ and $\Sigma(s_t)$ change simultaneously when switching from $\mathcal{S}_1$ to $\mathcal{S}_2$ and vice versa. If a NIW prior is specified for $A(s_t)$ and $\Sigma(s_t)$, and π_{11} and π_{22} have independent Beta prior distributions, a Gibbs sampler can be used to sample from the posterior (see e.g. Del Negro and Schorfheide, 2011).

A MS-VAR with non-recurrent states is called a 'change-point' model (see Chib, 1998; Bauwens and Rombouts, 2012). Generalising the specification to allow for more states, with the appropriate transition probabilities, allows to adapt the change-point model to the case of several structural breaks (see also Koop and Potter, 2007, 2009; Liu et al., 2017, for models where the number of change-points is unknown). Important extensions regard the transmission of structural shocks in the presence of structural breaks and in a time-varying coefficient environment discussed in e.g. Sims and Zha (2006) and Koop et al. (2011) who also allow for cointegration.

In threshold VARs (TVARs), the coefficients of the model change across regimes when an observable variable exceeds a given threshold value. Bayesian inference in TVAR models is discussed in detail in Geweke and Terui (1993) and Chen and Lee (1995). A TVAR with two regimes can be written as

$$y_t = Ax_t + \Theta(\tau_{t-d} - \tau)A^*x_t + u_t, \tag{57}$$

where A and A^* are $n \times k$ matrices that collect the autoregressive coefficients of the two regimes, $\Theta(\cdot)$ is a Heaviside step function, i.e. a discontinuous function whose value is zero for a negative argument, and one for a positive argument, τ_{t-d} is threshold variable at lag d, and τ is a potentially unobserved threshold value. The system in Eq. (57) can be easily generalised to allow for multiple regimes. TVARs have been applied to several problems in the economic literature (see, for example Koop and Potter, 1999; Ricco et al., 2016; Alessandri and Mumtaz, 2017).

If the coefficients gradually migrate to the new state(s), the model is called a smooth-transition VAR (STVAR). A STVAR model with two regimes can be written as

$$y_t = (1 - G(w_t; \vartheta, w))Ax_t + G(w_t; \vartheta, w)A^* x_t + u_t, \tag{58}$$

where A^*, A, and x_t are defined as in Eq. (57). The function $G(w_t; \vartheta, w)$ governs the transition across states, and is a function of the observable variable w_t, and of the parameters ϑ and w. In an exponential smooth-transition (EST) VAR, typically

$$G(w_t; \vartheta, w) = \frac{1}{1 + \exp\{-\vartheta(w_t - w)/\sigma_w\}} \tag{59}$$

where $\vartheta > 0$ determines the speed of transition across regimes, w can be thought of as a threshold value, and σ_w is the sample standard deviation of w_t. The higher ϑ the more abrupt the transition, the more the model collapses into a fixed threshold VAR. Among others, Gefang and Strachan (2009) and Gefang (2012) apply Bayesian techniques to estimate Smooth-transition VAR models.

6 Bayesian Panel VARs

Panel VARs generalise VAR models by describing the joint dynamics of multiple time series of heterogenous and interacting units – as for examples, the economies of several countries, regions, or sectors. Thorough reviews are in Canova and Ciccarelli (2013) and in Dieppe et al. (2016).

A panel VAR describes the the evolution of $y_{t,i}$ – the vector of $n \times 1$ endogenous variables of each unit $i \in [1, \ldots N]$ – by a system of p-th order VARs

$$y_{t,i} = \sum_{j=1}^{N} [A_{1,ij} y_{t-1,j} + \ldots + A_{p,ij} y_{t-p,j} + c_j] + H_i w_t + u_{t,i} \ , \tag{60}$$

where w_t is a vector of m exogenous controls. The innovations are generally assumed to be *i.i.d.* and Gaussian

$$u_{t,i} \sim i.i.d.\ \mathcal{N}(0, \sigma_i^2) \ , \tag{61}$$

while being possibly correlated across units.

Stacking over the N units, the model assumes the form of a VAR(p) with exogenous controls

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + c + H w_t + u_t \ , \qquad u_t \sim i.i.d.\ \mathcal{N}(0, \Sigma) \ . \tag{62}$$

In Eq. (63)

$$\underset{[Nn \times 1]}{y_t} = \begin{pmatrix} y_{t,1} \\ \vdots \\ y_{t,N} \end{pmatrix}, \quad \underset{[Nn \times 1]}{u_t} = \begin{pmatrix} u_{t,1} \\ \vdots \\ u_{t,N} \end{pmatrix} , \tag{63}$$

moreover

$$\underset{[Nn \times Nn]}{A_\ell} = \begin{pmatrix} A_{\ell,11} & \ldots & A_{\ell,1N} \\ \vdots & \ddots & \vdots \\ A_{\ell,1N} & \ldots & A_{\ell,NN} \end{pmatrix} \ \ell = 1, \ldots, p \ ,$$

and

$$\underset{[Nn \times Nn]}{\Sigma} = \begin{pmatrix} \Sigma_{11} & \ldots & \Sigma_{1N} \\ \vdots & \ddots & \vdots \\ \Sigma_{1N} & \ldots & \Sigma_{NN} \end{pmatrix} \qquad \underset{[Nn \times m]}{H} = \begin{pmatrix} H_1 \\ \vdots \\ H_N \end{pmatrix} .$$

While in Eq. (63) the system appears as a standard VAR, its panel structure is captured by three properties: (i) Dynamic interdependencies – the dynamics of the variables in each unit depend on the lagged values of the other endogenous variables in the unit and possibly all other units, i.e. $A_{\ell,jk} \neq 0$ for $j \neq k$; (ii) Static interdependencies – the innovations $u_{t,i}$ can be correlated across units, i.e. $\Sigma_{ij} \neq 0$ for $i \neq j$; (iii) Cross-unit (sub-sectional) heterogeneity – the VAR coefficients and residual variances can be unit-specific, i.e. $A_{\ell,ik} \neq A_{\ell,jk}$, $H_i \neq H_j$ and $\Sigma_{ii} \neq \Sigma_{jj}$ for $i \neq j$.[29]

If all of these properties are present in the data, and relationships do not exist among the coefficients, the system is a VAR with a large cross-section (i.e. a Large VAR) and can be estimated with standard macroeconomic priors such as e.g. the Minnesota priors

[29] An additional potential property of the Panel VAR is the time-variation in the VAR coefficients. For ease of notation we abstract from this in the following exposition.

of Section 3.[30] However, it is often possible to assume that some of these properties are relevant to system of interest.

If the units do not have dynamic or static interdependencies (i.e. $A_{\ell,jk} = 0$ and $\Sigma_{jk} = 0$ for $j \neq k$) and the dynamic coefficients are homogenous across units (i.e. $A_{\ell,jj} = \bar{A}_\ell$, $H_j = \bar{H}$, and $\Sigma_{jj} = \bar{\Sigma}\ \forall j$), then

$$A_\ell = \mathbb{I}_N \otimes \bar{A}_\ell, \qquad H = 1_{N\times 1} \otimes \bar{H}, \qquad \Sigma = \mathbb{I}_N \otimes \bar{\Sigma} \tag{64}$$

and the system simplifies into a single pooled VAR, with only $n \times (np + m)$ coefficients to be estimated.[31] By stacking the observations first over different units and then over different times, the system can be cast in the standard SUR representation (Eq. 6) and estimated with standard priors (e.g. Normal-Inverse Wishart prirors) and techniques.

If the dynamic coefficients are heterogenous across units but no dynamic or static interdependencies are present, then the system breaks up into N independent VARs, with the following SUR representation

$$y_i = x_i A_i + u_i \ . \tag{65}$$

A random coefficients model for Eq. (65) assumes that the coefficients from each unit can be thought of as random draws from a common distribution. For example,

$$\alpha_i \sim \mathcal{N}(a, \Sigma_a) \ , \tag{66}$$

where $\alpha_i \equiv vec(A_i)$. Eq. (66) can be thought of as an exchangeable Bayesian prior on the

[30] A Panel VAR can always be estimated as a Large VAR using standard macroeconomic priors (Bańbura et al., 2010). However, this implies (comes to the cost of) treating all the variables symmetrically thus disregarding the unit structure, and the fact that different variables may measure the same quantities in different units. Also, for very large systems the need to adopt too tight priors to overcome the issue of dimensionality may distort the posterior distribution.

[31] Alternatively, one could use standard priors to estimate a VAR for each of the N units separately, and then average the results across units. Such a mean group estimator is inefficient relative to the pooled estimator under dynamic homogeneity, but gives consistent estimates of the average system dynamic effects if dynamic heterogeneity is present. Conversely, the pooled estimator is inconsistent under dynamic heterogeneity due to the presence of correlation between the regressors and the error term.

units' coefficients – viz. the unit indices i are uninformative, in the sense that they can be exchanged without any loss of information. This approach was proposed by Zellner and Hong (1989) who used a 'Minnesota' type prior with fixed and known residual covariance matrix, a diagonal $\Sigma_{\boldsymbol{a}}$ with overall tightness hyperparameter $\lambda_{\boldsymbol{a}}$, and a plug-in pooled estimator for $\boldsymbol{a}$. Jarociñski (2010a) proposes instead a fully Bayesian model in which all the parameters are treated as random variables, and a sophisticated hierarchical prior approach is adopted. Estimation can be archived using a Gibbs sampler.

If dynamic interdependencies are allowed, the estimation problem becomes more complex. Canova and Ciccarelli (2004, 2009) have suggested solutions based on different Bayesian and cross-sectional shrinkage techniques that can deal with the issue of parameters proliferation that arises in these cases. The approach works by assuming a factor structure for the the matrix of coefficients and can be estimated with standard Bayesian priors and a Gibbs sampler. This structural factor approach is very flexible, and can also be used to estimate Panel VARs with dynamic coefficients that evolve over time, as done in e.g. Ciccarelli et al. (2012) and Canova and Ciccarelli, 2013.

7 Forecasting with BVARs

Reduced form Bayesian Vector Autoregressions, that are written as

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + c + u_t \ , \tag{67}$$

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).

7.1 Bayesian Forecasting

The fundamental object in Bayesian forecasting is the posterior predictive density.[32] 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

$$\mathbb{E}[\mathcal{L}(\mathcal{F}, y_{T+1:T+H}|y_{1-p:T})] = \int \mathcal{L}(\mathcal{F}, y_{T+1:T+H})p(y_{T+1:T+H}|y_{1-p:T})d_{y_{T+1:T+H}}. \tag{68}$$

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

$$p(y_{T+1:T+H}|y_{1-p:T}) = \int p(y_{T+1:T+H}|y_{1-p:T}, \theta)p(\theta|y_{1-p:T})d\theta, \tag{69}$$

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. (69) 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. (69) provides an easy way to generate draws from this predictive density.

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

Algorithm 4: 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. 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.[33,34]

7.2 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 $p(\mathbf{y}|\theta_1, \mathcal{M}_1, y_{1-p:0})$

[33] 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.

[34] 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.

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) predictive distribution as

$$p(y_{T+1:T+H}|\mathbf{y}) = p(y_{T+1:T+H}|\mathbf{y}, \mathcal{M}_1)p(\mathcal{M}_1) + p(y_{T+1:T+H}|\mathbf{y}, \mathcal{M}_2)p(\mathcal{M}_2), \tag{70}$$

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. (70) can be extended to allow for M different models. This can be seen as a generalisation of the predictive distribution in Eq. (69) 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).[35] Earlier reviews of BMA and forecast combinations are in Geweke and Whiteman (2006b) 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.

8 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

[35] 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.

restrictions. Moreover, it yields posterior distributions for the parameters which are consistent with the constrained paths. Let

$$R y_{T+1:T+H} = r \tag{71}$$

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

$$R\left[\mathbb{E}(y_{T+1:T+H}|\mathbf{y},\theta) + C' u_{T+1:T+H})\right] = r \ , \tag{72}$$

where

$$C = \begin{bmatrix} C_0 & C_1 & \cdots & C_{H-1} \\ 0 & C_0 & \cdots & C_{H-2} \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & C_0 \end{bmatrix} , \tag{73}$$

and C_j are the coefficients of the MA representation with

$$\begin{aligned} C_0 &= \mathbb{I}_n \\ C_j &= \sum_{i=0}^{p} A_i C_{j-i} \quad \forall j > 0 \ . \end{aligned} \tag{74}$$

Rearranging Eq. (72) as

$$RC' u_{T+1:T+H} = r - R\mathbb{E}(y_{T+1:T+H}|y_{1-p:T},\theta) \ , \tag{75}$$

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

$$u_{T+1:T+H}|(G u_{T+1:T+H} = g) \sim \mathcal{N}\left(\Sigma_H G'(G\Sigma_H G')^{-1} g,\ \Sigma_H - \Sigma_H G'(G\Sigma_H G')^{-1} G\Sigma_H\right) \tag{76}$$

which can be used to draw from the predictive distribution. In order to ensure consistency of the posterior distribution with the restriction in Eq. (75), 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 5: MCMC Sampler for VAR with restrictions on $y_{T+1:T+H}$. Given restrictions as in Eq. (75), select starting values for $A^{(0)}$ and $\Sigma^{(0)}$ using e.g. simulation on historical data. For $s = 1, \dots, n_{sim}$:

1. Draw $u_{T+1:T+H}$ from the distribution in Eq. (76) 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 $\left\{ y_{T+1}^{(s)}, \dots, y_{T+h}^{(s)} \right\}$ from the joint predictive density consistent with the restrictions in Eq. (75).

Jarociński (2010b) 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. (75). 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.

9 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

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + B_c + e_t, \qquad e_t \sim i.i.d.\mathcal{N}(0, \mathbb{I}_n) \ , \tag{77}$$

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. (77) by B_0^{-1} yields its reduced-form representation, i.e. the VAR in Eq. (67). 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

$$\Sigma = B_0^{-1} B_0^{-1'} \ . \tag{78}$$

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

$$B_0 = Q \Sigma_{Chol} \ . \tag{79}$$

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.[36]

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.

$$p(Q|Y, A, \Sigma) = p(Q|A, \Sigma) \ . \tag{80}$$

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

[36] 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 10. Relevant references are provided therein.

arguments – i.e. 'identification schemes' – about the appropriate choice of $p(Q|A, \Sigma)$.[37] 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

$$IRF_h = \Theta_h B_0^{-1} \qquad h = 0, \ldots, H \ , \tag{81}$$

where

$$\Theta_h = \sum_{\tau=1}^{h} \Theta_{h-\tau} A_\tau \qquad h = 1, \ldots, H \ , \tag{82}$$

$\Theta_0 = \mathbb{I}_n$, and A_τ are the reduced form autoregressive coefficients of Eq. (67) 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

[37] 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. (77). Rewrite the SVAR as

$$yB_0 = xB + e \ , \tag{83}$$

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

$$y = \begin{bmatrix} y_1' \\ \vdots \\ y_T' \end{bmatrix}, \quad x = \begin{bmatrix} x_1' \\ \vdots \\ x_T' \end{bmatrix}, \quad e = \begin{bmatrix} e_1' \\ \vdots \\ e_T' \end{bmatrix}, \tag{84}$$

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

$$p(y|B_0, B) \propto |B_0|^T exp\left\{-\frac{1}{2}tr\left[(yB_0 - xB)'(yB_0 - xB)\right]\right\} \ , \tag{85}$$

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

$$p(\beta_0, \beta) = p(\beta|\beta_0)p(\beta_0), \tag{86}$$

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.[38]

$$\beta|\beta_0 \sim \mathcal{N}\left(\underline{\beta}_0, \lambda^{-1}\mathbb{I}_n \otimes \underline{\Gamma}_{\beta_0}\right) \ . \tag{87}$$

The posterior distribution of β is hence of the standard form

$$\beta|\beta_0, \mathbf{y} \sim \mathcal{N}\left(\overline{\beta}_0, \mathbb{I}_n \otimes \overline{\Gamma}_{\beta_0}\right) \ , \tag{88}$$

[38] 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.[39]

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-

[39] 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.

10 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.[40] 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.[41] 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.[42]

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

[40]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.

[41]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.

[42]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.

10.1 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

$$\Delta y_t = \Phi_1 \Delta y_{t-1} + \cdots + \Phi_p \Delta y_{t-p} + v_t. \tag{89}$$

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. (89) can be characterised by the following first and second moments:

$$\mathbb{E}\left[(\Phi_\ell)_{ij}|\Psi\right] = 0, \ \forall \ell \qquad \mathbb{V}ar\left[(\Phi_\ell)_{ij}|\Psi\right] = \begin{cases} \frac{\lambda_1^2}{f(\ell)} & \text{for } i = j, \forall \ell \\ \frac{\lambda_1^2}{f(\ell)} \frac{\Sigma_{ij}}{\omega_j^2} & \text{for } i \neq j, \forall \ell. \end{cases} \tag{90}$$

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.

$$\left[\frac{1}{T}\sum_{t=1}^{T} \Delta y_t \Delta y_t'\right] \nu_j = \nu_j \zeta_j, \tag{91}$$

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

$$z_t = \left[\frac{\nu_1}{\sqrt{\zeta_1}} \ \ldots \ \frac{\nu_n}{\sqrt{\zeta_n}}\right]' y_t \equiv W \Delta y_t \ . \tag{92}$$

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. (89) in terms of the ordered principal components, as

$$\Delta y_t = \Phi_1 W^{-1} z_{t-1} + \cdots + \Phi_p W^{-1} z_{t-p} + v_t \ . \tag{93}$$

The priors that impose a uniform shrinkage on the parameters in Eq. (90) map into a non-uniform shrinkage on the parameters in Eq. (93):

$$\mathbb{E}\left[(\Phi_\ell W^{-1})_{ij}|\Psi\right] = 0, \ \forall \ell \qquad \mathbb{V}ar\left[(\Phi_\ell W^{-1})_{ij}|\Psi\right] = \begin{cases} \frac{\lambda_1^2 \zeta_j}{f(\ell)} & \text{for } i = j, \forall \ell \\ \frac{\lambda_1^2 \zeta_j}{f(\ell)} \frac{\Psi_{ij}}{\omega_j^2} & \text{for } i \neq j, \forall \ell. \end{cases} \tag{94}$$

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.

10.2 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. 77) 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. 67) 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.[43]

10.3 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

[43]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., 2015b) 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.

Alessandri, Piergiorgio and Haroon Mumtaz (2017) "Financial conditions and density forecasts for US output and inflation," *Review of Economic Dynamics*, Vol. 24, pp. 66–78, March.

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.

Bauwens, Luc and Jeroen V.K. Rombouts (2012) "On marginal likelihood computation in change-point models," *Computational Statistics & Data Analysis*, Vol. 56, No. 11, pp. 3415–3429.

Belmonte, Miguel A.G., Gary Koop, and Dimitris Korobilis (2014) "Hierarchical Shrinkage in Time-Varying Parameter Models," *Journal of Forecasting*, Vol. 33, No. 1, pp. 80–94, January.

Berger, James O. (1985) *Statistical Decision Theory and Bayesian Analysis*: Springer.

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.

Bernardo, J.M. and A.F.M. Smith (2009) *Bayesian Theory*, Wiley Series in Probability and Statistics: Wiley.

Bhattacharya, Anirban, Debdeep Pati, Natesh S. Pillai, and David B. Dunson (2015a) "Dirichlet-Laplace Priors for Optimal Shrinkage," *Journal of the American Statistical Association*, Vol. 110, No. 512, pp. 1479–1490. PMID: 27019543.

——— (2015b) "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.

Bloor, Chris and Troy Matheson (2011) "Real-time conditional forecasts with Bayesian VARs: An application to New Zealand," *The North American Journal of Economics and Finance*, Vol. 22, No. 1, pp. 26 – 42. Symposium on Nowcasting and Model Combination.

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 (1992) "An Alternative Approach to Modeling and Forecasting Seasonal Time Series," *Journal of Business & Economic Statistics*, Vol. 10, No. 1, pp. 97–108.

——— (1993) "Forecasting time series with common seasonal patterns," *Journal of Econometrics*, Vol. 55, No. 1-2, pp. 173–200.

——— (2007) *Methods for Applied Macroeconomic Research*: Princeton University Press.

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.

——— (2013) "Panel Vector Autoregressive Models: A Survey," in Thomas B. Fomby, Lutz Kilian, and Anthony Murphy eds. *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims (Advances in Econometrics)*, Vol. 32: Emerald Group Publishing Limited, pp. 205 – 246.

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.

Chen, Cathy W. S. and Jack C. Lee (1995) "Bayesian Inference of Threshold Autoregressive Models," *Journal of Time Series Analysis*, Vol. 16, No. 5, pp. 483–492.

Chib, Siddhartha (1998) "Estimation and comparison of multiple change-point models," *Journal of Econometrics*, Vol. 86, No. 2, pp. 221 – 241.

——— (2001) "Markov Chain Monte Carlo Methods: Computation and Inference," in James J. Heckman and Edward Leamer eds. *Handbook of Econometrics*, Vol. 5: Elsevier, pp. 3569 – 3649.

Chib, Siddhartha and Edward Greenberg (1995) "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, Vol. 49, No. 4, pp. 327–335.

Chiu, Ching-Wai (Jeremy), Haroon Mumtaz, and Gábor Pintér (2017) "Forecasting with VAR models: Fat tails and stochastic volatility," *International Journal of Forecasting*, Vol. 33, No. 4, pp. 1124 – 1143.

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.

Ciccarelli, Matteo and Alessandro Rebucci (2003) "Bayesian Vars; A Survey of the Recent Literature with An Application to the European Monetary System," IMF Working Papers 03/102, International Monetary Fund.

Ciccarelli, Matteo, Eva Ortega, and Maria Teresa Valderrama (2012) "Heterogeneity and cross-country spillovers in macroeconomic-financial linkages," Working Paper Series 1498, European Central Bank.

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 and Thomas J. Sargent (2002) "Evolving Post-World War II U.S. Inflation Dynamics," in *NBER Macroeconomics Annual 2001, Volume 16*: National Bureau of Economic Research, Inc, pp. 331–388.

——— (2005) "Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S," *Review of Economic Dynamics*, Vol. 8, No. 2, pp. 262–302, April.

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.

DeJong, David N., Beth Ingram, and Charles H. Whiteman (1993) "Analyzing VARs with monetary business cycle model priors," *Proceedings of the American Statistical Association, Bayesian Statistics Section*, pp. 160–169.

Del Negro, Marco and Frank Schorfheide (2004) "Priors from General Equilibrium Models for VARS," *International Economic Review*, Vol. 45, No. 2, pp. 643–673, May.

——— (2011) "Bayesian Macroeconometrics," in John Geweke, Gary Koop, and Herman Van Dijk eds. *The Oxford Handbook of Bayesian Econometrics*: Oxford University Press, pp. 293–389.

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.

Dieppe, Alistair, Björn van Roye, and Romain Legrand (2016) "The BEAR toolbox," Working Paper Series 1934, European Central Bank.

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.

Gefang, Deborah (2012) "Money-output Causality Revisited – A Bayesian Logistic Smooth Transition VECM Perspective," *Oxford Bulletin of Economics and Statistics*, Vol. 74, No. 1, pp. 131–151, February.

Gefang, Deborah and Rodney Strachan (2009) "Nonlinear Impacts of International Business Cycles on the U.K. – A Bayesian Smooth Transition VAR Approach," *Studies in Nonlinear Dynamics & Econometrics*, Vol. 14, No. 1, pp. 1–33, December.

Geisser, Seymour (1965) "Bayesian Estimation in Multivariate Analysis," *The Annals of Mathematical Statistics*, Vol. 36, No. 1, pp. 150–159, 02.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin (2003) *Bayesian Data Analysis, Second Edition*, Chapman & Hall/CRC Texts in Statistical Science: Taylor & Francis.

Gelman, A., J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin (2013) *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science: Taylor & Francis.

George, Edward I., Dongchu Sun, and Shawn Ni (2008) "Bayesian stochastic search for VAR model restrictions," *Journal of Econometrics*, Vol. 142, No. 1, pp. 553 – 580.

Geweke, John (1996) "Bayesian reduced rank regression in econometrics," *Journal of Econometrics*, Vol. 75, No. 1, pp. 121–146, November.

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

——— (2005) *Contemporary Bayesian Econometrics and Statistics*, Wiley Series in Probability and Statistics: Wiley.

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 Nobuhiko Terui (1993) "Bayesian Threshold Autoregressive Models for Nonlinear Time Series," *Journal of Time Series Analysis*, Vol. 14, No. 5, pp. 441–454.

Geweke, John and Charles Whiteman (2006a) "Bayesian Forecasting," in G. Elliott, C. Granger, and A. Timmermann eds. *Handbook of Economic Forecasting*, Vol. 1: Elsevier, 1st edition, Chap. 01, pp. 3–80.

——— (2006b) "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 (2015) "Prior Selection for Vector Autoregressions," *The Review of Economics and Statistics*, Vol. 2, No. 97, pp. 436–451, May.

——— (2016) "Priors for the Long Run," CEPR Discussion Papers 11261, C.E.P.R. Discussion Papers.

——— (2017) "Economic Predictions with Big Data: The Illusion Of Sparsity," CEPR Discussion Papers 12256, C.E.P.R. Discussion Papers.

Giannone, Domenico, Michele Lenza, and Lucrezia Reichlin (2008) "Explaining The Great Moderation: It Is Not The Shocks," *Journal of the European Economic Association*, Vol. 6, No. 2-3, pp. 621–633, 04-05.

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.

Griffin, Jim E. and Philip J. Brown (2010) "Inference with normal-gamma prior distributions in regression problems," *Bayesian Anal.*, Vol. 5, No. 1, pp. 171–188, 03.

Griffin, Jim and Phil Brown (2017) "Hierarchical Shrinkage Priors for Regression Models," *Bayesian Analysis*, Vol. 12, No. 1, pp. 135–159, 03.

Hajivassiliou, Vassilis A. and Paul A. Ruud (1994) "Classical estimation methods for LDV models using simulation," in *Handbook of Econometrics*, Vol. 4: Elsevier, pp. 2383 – 2441.

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.

Highfield, Richard A. (1992) “Forecasting Similar Time Series with Bayesian Pooling Methods: Application to Forecasting European Output Growth,” in Prem K. Goel and N. Sreenivas Iyengar eds. *Bayesian Analysis in Statistics and Econometrics*, pp. 303–326, New York, NY: Springer New York.

Huber, Florian and Martin Feldkircher (2017) “Adaptive Shrinkage in Bayesian Vector Autoregressive Models,” *Journal of Business & Economic Statistics*, Vol. 0, No. 0, pp. 1–13.

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

Ingram, Beth F. and Charles H. Whiteman (1994) “Supplanting the ’Minnesota’ prior: Forecasting macroeconomic time series using real business cycle model priors,” *Journal of Monetary Economics*, Vol. 34, No. 3, pp. 497–510, December.

Jarociñski, Marek (2010a) “Responses to monetary policy shocks in the east and the west of Europe: a comparison,” *Journal of Applied Econometrics*, Vol. 25, No. 5, pp. 833–868.

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

Jarociński, Marek and Bartosz Maćkowiak (2017) “Granger Causal Priority and Choice of Variables in Vector Autoregressions,” *The Review of Economics and Statistics*, Vol. 99, No. 2, pp. 319–329, May.

Jarociński, Marek and Albert Marcet (2011) “Autoregressions in Small Samples, Priors about Observables and Initial Conditions,” CEP Discussion Papers dp1061, Centre for Economic Performance, LSE.

——— (2014) “Contrasting Bayesian and Frequentist Approaches to Autoregressions: the Role of the Initial Condition,” Working Papers 776, Barcelona Graduate School of Economics.

Jochmann, Markus and Gary Koop (2015) “Regime-switching cointegration,” *Studies in Nonlinear Dynamics & Econometrics*, Vol. 19, No. 1, pp. 35–48, February.

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 (2013a) “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.

——— (2013b) *Forecasting with Bayesian Vector Autoregression*, Vol. 2 of Handbook of Economic Forecasting, Chap. 0, pp. 791–897: Elsevier.

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.

Kim, Jae Young (1994) "Bayesian Asymptotic Theory in a Time Series Model with a Possible Nonstationary Process," *Econometric Theory*, Vol. 10, No. 3/4, pp. 764–773.

Kim, Chang Jin and Charles Nelson (1999) *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, Vol. 1: The MIT Press, 1st edition.

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo (1988) "Production, growth and business cycles: I. The basic neoclassical model," *Journal of Monetary Economics*, Vol. 21, No. 2, pp. 195 – 232.

Kleibergen, Frank and Richard Paap (2002) "Priors, posteriors and bayes factors for a Bayesian analysis of cointegration," *Journal of Econometrics*, Vol. 111, No. 2, pp. 223–249, December.

Kleibergen, Frank and Herman K. van Dijk (1994) "On the Shape of the Likelihood/Posterior in Cointegration Models," *Econometric Theory*, Vol. 10, No. 3-4, pp. 514–551, August.

Koop, Gary (2003) *Bayesian Econometrics*: Wiley.

——— (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 (2010) "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics," *Foundations and Trends(R) in Econometrics*, Vol. 3, No. 4, pp. 267–358, July.

——— (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 and Simon M. Potter (1999) "Dynamic Asymmetries in U.S. Unemployment," *Journal of Business & Economic Statistics*, Vol. 17, No. 3, pp. 298–312.

——— (2007) "Estimation and Forecasting in Models with Multiple Breaks," *The Review of Economic Studies*, Vol. 74, No. 3, pp. 763–789.

——— (2009) "Prior Elicitation In Multiple Change-Point Models," *International Economic Review*, Vol. 50, No. 3, pp. 751–772, August.

Koop, Gary and Mark F. J. Steel (1991) "A comment on: "To criticize the critics: An objective bayesian analysis of stochastic trends", By Peter C. B. Phillips," *Journal of Applied Econometrics*, Vol. 6, No. 4, pp. 365–370.

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

Koop, Gary, Roberto Leon-Gonzalez, and Rodney W. Strachan (2011) "Bayesian inference in a time varying cointegration model," *Journal of Econometrics*, Vol. 165, No. 2, pp. 210–220.

Koop, G., R.W. Strachan, H.K. van Dijk, and M. Villani (2006) "Monetary policy shocks: What have we learned and to what end?" in T. C.Mills and K. P. Patterson eds. *Palgrave Handbook of Econometrics*, Vol. 1: Basingstoke: Palgrave Macmillan, pp. 871–98.

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

Kroese, Dirk P. and Joshua Chan (2014) *Statistical Modeling and Computation*: Springer.

Kwan, Yum K. (1998) "Asymptotic Bayesian analysis based on a limited information estimator," *Journal of Econometrics*, Vol. 88, No. 1, pp. 99–121, November.

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.

Litterman, Robert B. (1979) "Techniques of forecasting using vector autoregressions," Working Papers 115, Federal Reserve Bank of Minneapolis.

——— (1980) "A Bayesian Procedure for Forecasting with Vector Autoregression," working papers, MIT Department of Economics.

——— (1986) "Forecasting with Bayesian Vector Autoregressions-Five Years of Experience," *Journal of Business & Economic Statistics*, Vol. 4, No. 1, pp. 25–38, January.

Liu, Philip, Haroon Mumtaz, Konstantinos Theodoridis, and Francesco Zanetti (2017) "Changing Macroeconomic Dynamics at the Zero Lower Bound," *Journal of Business & Economic Statistics*, Vol. 0, No. ja, pp. 0–0.

Lubik, Thomas A. and Christian Matthes (2015) "Time-Varying Parameter Vector Autoregressions: Specification, Estimation, and an Application," *Economic Quarterly*, Vol. 4Q, pp. 323–352.

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 (2017) "The transmission of monetary policy shocks," Bank of England working papers 657, Bank of England.

——— (2018) "Bayesian Vector Autoregressions: Applications," in Jonathan Hamilton ed. *Oxford Research Encyclopedia in Economics and Finance*: Oxford University Press.

Mitchell, T. J. and J. J. Beauchamp (1988) "Bayesian Variable Selection in Linear Regression," *Journal of the American Statistical Association*, Vol. 83, No. 404, pp. 1023–1032.

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

Müller, Ulrich K. (2013) "Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix," *Econometrica*, Vol. 81, No. 5, pp. 1805–1849.

Müller, Ulrich K. and Graham Elliott (2003) "Tests for Unit Roots and the Initial Condition," *Econometrica*, Vol. 71, No. 4, pp. 1269–1286.

Panagiotelis, Anastasios and Michael Smith (2008) "Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions," *International Journal of Forecasting*, Vol. 24, No. 4, pp. 710–727.

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.

Phillips, P. C. B. (1991a) "Bayesian routes and unit roots: De rebus prioribus semper est disputandum," *Journal of Applied Econometrics*, Vol. 6, No. 4, pp. 435–473.

——— (1991b) "To criticize the critics: An objective bayesian analysis of stochastic trends," *Journal of Applied Econometrics*, Vol. 6, No. 4, pp. 333–364.

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

Primiceri, Giorgio E. (2005) "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, Vol. 72, No. 3, pp. 821–852.

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.

Raiffa, H. and R. Schlaifer (1961) *Applied statistical decision theory*, Studies in managerial economics: Division of Research, Graduate School of Business Adminitration, Harvard University.

Ricco, Giovanni, Giovanni Callegari, and Jacopo Cimadomo (2016) "Signals from the government: Policy disagreement and the transmission of fiscal shocks," *Journal of Monetary Economics*, Vol. 82, No. C, pp. 107–118.

Robertson, John C. and Ellis W. Tallman (1999) "Vector autoregressions: forecasting and reality," *Economic Review*, Vol. Issue Q1, pp. 4–18.

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. (1980) "Macroeconomics and Reality," *Econometrica*, Vol. 48, No. 1, pp. 1–48, January.

——— (1988) "Bayesian skepticism on unit root econometrics," *Journal of Economic Dynamics and Control*, Vol. 12, No. 2, pp. 463 – 474.

——— (1991) "Comment by Christopher A. Sims on 'To criticize the critics', by Peter C. B. Phillips," *Journal of Applied Econometrics*, Vol. 6, No. 4, pp. 423–434.

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

——— (1996) "Inference For Multivariate Time Series Models With Trend,"Technical report, Princeton University, mimeo.

——— (2000) "Using a likelihood perspective to sharpen econometric discourse: Three examples," *Journal of Econometrics*, Vol. 95, No. 2, pp. 443–462.

——— (2005a) "Conjugate Dummy Observation Priors for VARs,"Technical report, Princeton University, mimeo.

——— (2005b) "Dummy Observation Priors Revisited,"Technical report, Princeton University, mimeo.

——— (2010a) "Causal Ordering and Exogeneity,"Technical report, Princeton University, mimeo.

——— (2010b) "Understanding Non-Bayesians,"Technical report, Princeton University, mimeo.

Sims, Christopher A and Harald Uhlig (1991) "Understanding Unit Rooters: A Helicopter Tour," *Econometrica*, Vol. 59, No. 6, pp. 1591–1599, November.

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.

——— (2006) "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review*, Vol. 96, No. 1, pp. 54–81, March.

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.

Strachan, Rodney W. and Brett Inder (2004) "Bayesian analysis of the error correction model," *Journal of Econometrics*, Vol. 123, No. 2, pp. 307–325, December.

Theil, H. (1963) "On the Use of Incomplete Prior Information in Regression Analysis," *Journal of the American Statistical Association*, Vol. 58, No. 302, pp. 401–414.

Theil, H. and A. S. Goldberger (1961) "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review*, Vol. 2, No. 1, pp. 65–78.

Tiao, George C. and Arnold Zellner (1964) "On the Bayesian Estimation of Multivariate Regression," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 26, No. 2, pp. 277–285.

Tierney, Luke (1994) "Markov Chains for Exploring Posterior Distributions," *Ann. Statist.*, Vol. 22, No. 4, pp. 1701–1728, 12.

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 (1994a) "On Jeffreys Prior when Using the Exact Likelihood Function," *Econometric Theory*, Vol. 10, No. 3-4, pp. 633–644.

——— (1994b) "What Macroeconomists Should Know about Unit Roots: A Bayesian Perspective," *Econometric Theory*, Vol. 10, No. 3-4, pp. 645–671.

——— (1997) "Bayesian Vector Autoregressions with Stochastic Volatility," *Econometrica*, Vol. 65, No. 1, pp. 59–74, January.

——— (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.

Villani, Mattias (2001) "Bayesian prediction with cointegrated vector autoregressions," *International Journal of Forecasting*, Vol. 17, No. 4, pp. 585 – 605.

——— (2009) "Steady-state priors for vector autoregressions," *Journal of Applied Econometrics*, Vol. 24, No. 4, pp. 630–650.

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.

Wright, Jonathan H. (2009) "Forecasting US inflation by Bayesian model averaging," *Journal of Forecasting*, Vol. 28, No. 2, pp. 131–144.

Zellner, Arnold (1971) *An introduction to Bayesian inference in econometrics*, Wiley Classics Library: Wiley-Interscience.

Zellner, Arnold and Chansik Hong (1989) "Forecasting international growth rates using Bayesian shrinkage and other procedures," *Journal of Econometrics*, Vol. 40, No. 1, pp. 183–202, January.