

CAMA

Centre for Applied Macroeconomic Analysis

Efficient estimation of Bayesian VARMA with time-varying coefficients

CAMA Working Paper 19/2015
June 2015

Joshua C.C. Chan

Research School of Economics, ANU and
Centre for Applied Macroeconomic Analysis (CAMA), ANU

Eric Eisenstat

School of Economics, The University of Queensland

Abstract

Empirical work in macroeconometrics has been mostly restricted to using VARs, even though there are strong theoretical reasons to consider general VARMA. A number of articles in the last two decades have conjectured that this is because estimation of VARMA is perceived to be challenging and proposed various ways to simplify it. Nevertheless, VARMA continue to be largely dominated by VARs, particularly in terms of developing useful extensions. We address these computational challenges with a Bayesian approach. Specifically, we develop a Gibbs sampler for the basic VARMA, and demonstrate how it can be extended to models with time-varying VMA coefficients and stochastic volatility. We illustrate the methodology through a macroeconomic forecasting exercise. We show that in a class of models with stochastic volatility, VARMA produce better density forecasts than VARs, particularly for short forecast horizons.

Keywords

state space, stochastic volatility, factor model, macroeconomic forecasting, density forecast

JEL Classification

C11, C32, C53

Address for correspondence:

(E) cama.admin@anu.edu.au

[The Centre for Applied Macroeconomic Analysis](#) in the Crawford School of Public Policy has been established to build strong links between professional macroeconomists. It provides a forum for quality macroeconomic research and discussion of policy issues between academia, government and the private sector.

The Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

Efficient Estimation of Bayesian VARMA with Time-Varying Coefficients

Joshua C.C. Chan

Research School of Economics,
and Centre for Applied Macroeconomic Analysis,
Australian National University

Eric Eisenstat*

School of Economics,
The University of Queensland

June 2015

Abstract

Empirical work in macroeconometrics has been mostly restricted to using VARs, even though there are strong theoretical reasons to consider general VARMA. A number of articles in the last two decades have conjectured that this is because estimation of VARMA is perceived to be challenging and proposed various ways to simplify it. Nevertheless, VARMA continue to be largely dominated by VARs, particularly in terms of developing useful extensions. We address these computational challenges with a Bayesian approach. Specifically, we develop a Gibbs sampler for the basic VARMA, and demonstrate how it can be extended to models with time-varying VMA coefficients and stochastic volatility. We illustrate the methodology through a macroeconomic forecasting exercise. We show that in a class of models with stochastic volatility, VARMA produce better density forecasts than VARs, particularly for short forecast horizons.

Keywords: state space, stochastic volatility, factor model, macroeconomic forecasting, density forecast

JEL codes: C11, C32, C53

*Corresponding author: School of Economics, The University of Queensland, Level 6, Colin Clark Building (39), St. Lucia, Brisbane Qld 4072 (Australia), eric.eisenstat@gmail.com. Joshua Chan would like to acknowledge financial support by the Australian Research Council via a Discovery Early Career Researcher Award (DE150100795).

1 Introduction

Vector autoregressive moving average (VARMA) models have long been considered an appropriate framework for modeling covariance stationary time series. As is well known, by the Wold decomposition theorem, any covariance stationary time series has an infinite moving average representation. Whenever this is characterized by a rational transfer function, the multivariate series can be exactly represented by a finite-order VARMA model. When the transfer function is irrational, the VARMA specification can be used to provide an arbitrarily close approximation (Lütkepohl and Poskitt, 1996).

However, in most empirical work, only purely autoregressive models are considered. In fact, since the seminal work of Sims (1980), VARs have become the most prominent approach in empirical macroeconometrics. This is in spite of long-standing criticisms of VARs, especially with short lag orders, as being theoretically deficient for macroeconomic applications. There are two main theoretical drawbacks of VARs in this context: first, linearized DSGE models typically result in VARMA, not VARs (e.g., Cooley and Dwyer, 1998; Yang, 2005; Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson, 2007; Leeper, Walker, and Yang, 2008). Second, even if a particular set of variables can be adequately described by a VAR, any linear combination, temporal aggregation, or subsets of these variables will follow a VARMA process.

Over the past two decades, a number of authors (e.g., Lütkepohl and Poskitt, 1996; Lütkepohl and Claessen, 1997; Athanasopoulos and Vahid, 2008; Athanasopoulos, Poskitt, and Vahid, 2012; Poskitt, 2011; Dufour and Stevanović, 2013; Dufour and Pelletier, 2014; Kascha and Trenkler, 2014) have pointed out this unfortunate phenomenon and various approaches have been proposed aimed at making VARMA accessible to applied macroeconomists. Nevertheless, VARs continue to dominate in this field. One possible reason for this is that many flexible extensions of the basic VAR have been developed. For example, VARs are now routinely augmented with time-varying coefficients (Canova, 1993; Koop and Korobilis, 2013), Markov switching or regime switching processes (Paap and van Dijk, 2003; Koop and Potter, 2007), and stochastic volatility (Cogley and Sargent, 2005; Primiceri, 2005). Recently, specifications targeting a large number of variables such as factor augmented VARs (Bernanke, Boivin, and Elias, 2005; Korobilis, 2012), among many others (see, e.g., Koop and Korobilis, 2010) have been introduced.

While these extensions have made VARs extremely flexible, papers such as Banbura, Giannone, and Reichlin (2010); Koop (2011); Korobilis (2013); Eisenstat, Chan, and Strachan (2014) have focused on achieving parsimony and controlling over-parameterization. This balance between flexibility and parsimony has contributed to the success of VARs in forecasting macroeconomic time series. Seemingly, these developments surrounding VARs have largely overshadowed the advantages inherent to VARMA specifications. Indeed, VARMA remain largely underdeveloped in this sense, and this is a central concern of the present paper.

The lack of development of VARMA is unfortunate because well-specified VARMA are

naturally parsimonious, and recent evidence suggests that VARMA do forecast macroeconomic variables better than basic VARs (Athanasopoulos and Vahid, 2008; Athanasopoulos, Poskitt, and Vahid, 2012; Kascha and Trenkler, 2014).¹ Therefore, similar extensions such as time-varying coefficients and stochastic volatility in a VARMA specification may offer even further forecasting gains. Moreover, for structural analysis such as estimation of impulse response functions, VARs remain fundamentally deficient. Cooley and Dwyer (1998) argue that typical applications of structural VARs in macroeconomics use lag lengths that are much too short to adequately approximate the underlying, theoretically founded VARMA processes. In certain cases—such as in DSGEs with fiscal foresight—the VARMA process arising in equilibrium is not invertible, and therefore, a VAR approximation does not exist at all (Yang, 2005; Leeper, Walker, and Yang, 2008).

The identification and estimation of VARMA is, however, far more involved than with VARs. Even for a pure VMA process, the likelihood is highly nonlinear in the parameters, and identification further requires that a system of nonlinear invertibility conditions be imposed. In consequence, estimation using maximum likelihood or Bayesian methods is difficult, and most practical applications rely on approximate methods rather than exact inference from the likelihood (see Kascha, 2012, for a review). Combining VMA with VAR terms gives rise to further problems in terms of specification and identification (see Lütkepohl, 2005, for a textbook treatment), thereby complicating matters even more.

We propose a Bayesian approach that draws on a few recent developments in the state space literature. First, we make use of a convenient state space representation of the VARMA introduced in Metaxoglou and Smith (2007). Specifically, by using the fact that a VMA plus white noise remains a VMA (Peiris, 1988), the authors write a VARMA as a latent factor model, with the unusual feature that lagged factors also enter the current measurement equation. This linear state space form is an *equivalent*, but over-parameterized representation of the original VARMA. To estimate the model in this form, Metaxoglou and Smith (2007) set *ex ante* certain parameters to pre-determined values and estimate the remaining parameters using the EM algorithm based on the Kalman filter.

Our point of departure is to develop an efficient Gibbs sampler for this state space representation of the VARMA. First, we show that the pre-determined parameter restrictions in this case are neither desirable nor necessary—in a Bayesian setting, we work directly with the “unidentified” model and recover the identified VARMA parameters *ex post*. We emphasize from the start and demonstrate below that doing so *does not* require restrictive priors on the coefficients. Another advantage of this approach is that invertibility restrictions can be imposed in the post-processing of draws, rather than directly in the sampling scheme. To further accelerate computation, instead of the conventional forward-filtering and backward-smoothing algorithms based on the Kalman filter, we make use of the more efficient precision sampler of Chan and Jeliaskov (2009) to simulate the latent factors.

The significance of our contribution lies in the realization that once the basic VARMA

¹Chan (2013) arrives at a similar conclusion in forecasting inflation with univariate MA models with stochastic volatility.

can be efficiently estimated via the Gibbs sampler, a wide variety of generalizations, analogous to those mentioned earlier extending the basic VAR, can also be fitted easily using the machinery of Markov chain Monte Carlo (MCMC) techniques. We will focus on a particular generalization of the VARMA: allowing for time-varying VMA coefficients and stochastic volatility. Within this scope, we do not address the important issue of specifying a VARMA in canonical form, except to point out that the methods developed below can be readily used to estimate a VARMA in echelon form, conditional on knowledge of the Kronecker indices. An in-depth investigation of Bayesian approaches to specifying and estimating an echelon form VARMA is undertaken in our related work in Chan, Eisenstat, and Koop (2015). We show in the latter that building on the foundation laid out in the present paper, the same extensions may be incorporated in straightforward fashion in the fully canonical echelon form specification (where Kronecker indices are estimated jointly with the model parameters) as well.

In the present work, our main aim is to assess the forecasting potential of VARMA with time-varying coefficients and stochastic volatility. Specifically, we investigate whether adding moving average components to VARs with stochastic volatility (i.e., a modern, widespread forecasting tool) improves forecasting performance. The sampling algorithm we develop is expressly suitable for this purpose, and we do find that VARMA with time-varying coefficients and stochastic volatility generate better density forecasts than their VAR counterparts, particularly for inflation over short horizons.

To our knowledge, few attempts have been made to apply Bayesian methods in specifying and estimating VARMA models. Two noteworthy exceptions are Ravishanker and Ray (1997), who consider a hybrid Metropolis-Hastings algorithm for a basic VARMA, and Li and Tsay (1998), who use stochastic search variable selection (SSVS) priors (e.g., George and McCulloch, 1993) to jointly sample the Kronecker indices and coefficients of a VARMA in echelon form. The latter sampler is based on the observation that each equation within a VARMA is a univariate ARMAX, conditional on the other variables in the system. Both of these approaches, however, are computationally intensive and do not provide a convenient framework for incorporating the type of extensions we develop here.

The rest of this article is organized as follows. Section 2 first introduces a state space representation of a VARMA(p, q)—which we term the *expanded VARMA form*—that facilitates efficient estimation, followed by a detailed discussion of the correspondence between this representation and the original VARMA. We develop a Gibbs sampler for the basic VARMA in Section 3, which is extended to models with time-varying coefficients and stochastic volatility in Section 4. In Section 5, the methodology is illustrated using a recursive forecasting exercise involving inflation and GDP growth. Lastly, we discuss some future research directions in Section 6.

2 The Expanded VMA Form

2.1 The Basic Setup

To build up the general framework, we start with the pure VMA(q) specification:

$$\mathbf{u}_t = \Theta_0 \boldsymbol{\varepsilon}_t + \Theta_1 \boldsymbol{\varepsilon}_{t-1} + \cdots + \Theta_q \boldsymbol{\varepsilon}_{t-q}, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad (1)$$

where \mathbf{u}_t is $n \times 1$ and all other matrices conform accordingly. The autocovariances generated by this VMA(q) process are given by

$$\Gamma_j \equiv \text{E}(\mathbf{u}_t \mathbf{u}'_{t-j}) = \sum_{l=j}^q \Theta_l \boldsymbol{\Sigma} \Theta'_{l-j}, \quad j = 0, \dots, q.$$

In what follows, it will also be useful to consider a VMA(1) representation of the general VMA(q), defined as

$$\underbrace{\begin{pmatrix} \mathbf{u}_t \\ \mathbf{u}_{t-1} \\ \vdots \\ \mathbf{u}_{t-q+1} \end{pmatrix}}_{\tilde{\mathbf{u}}_\tau} = \underbrace{\begin{pmatrix} \Theta_0 & \Theta_1 & \cdots & \Theta_{q-1} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \Theta_1 \\ & & & \Theta_0 \end{pmatrix}}_{\tilde{\Theta}_0} \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\varepsilon}_{t-1} \\ \vdots \\ \boldsymbol{\varepsilon}_{t-q+1} \end{pmatrix}}_{\tilde{\boldsymbol{\varepsilon}}_\tau} + \underbrace{\begin{pmatrix} \Theta_q & & & \\ \Theta_{q-1} & \ddots & & \\ \vdots & \ddots & \ddots & \\ \Theta_1 & \cdots & \Theta_{q-1} & \Theta_q \end{pmatrix}}_{\tilde{\Theta}_1} \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_{t-q} \\ \boldsymbol{\varepsilon}_{t-q-1} \\ \vdots \\ \boldsymbol{\varepsilon}_{t-2q+1} \end{pmatrix}}_{\tilde{\boldsymbol{\varepsilon}}_{\tau-1}} \quad (2)$$

with $\tilde{\boldsymbol{\varepsilon}}_\tau \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q \otimes \boldsymbol{\Sigma})$. In this form, the corresponding autovariances may be denoted by

$$\tilde{\Gamma}_0 = \begin{pmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{q-1} \\ \Gamma'_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \Gamma_1 \\ \Gamma'_{q-1} & \cdots & \Gamma'_1 & \Gamma_0 \end{pmatrix} \quad \text{and} \quad \tilde{\Gamma}_1 = \begin{pmatrix} \Gamma_q & & & \\ \Gamma_{q-1} & \ddots & & \\ \vdots & \ddots & \ddots & \\ \Gamma_1 & \cdots & \Gamma_{q-1} & \Gamma_q \end{pmatrix}. \quad (3)$$

Writing the VMA(q) this way allows us to work directly with the simplest case, $q = 1$, and generalize the developed concepts and methods to any q through (2)-(3).

Following Metaxoglou and Smith (2007), consider now the decomposition of \mathbf{u}_t :

$$\mathbf{u}_t = \Phi_0 \mathbf{f}_t + \cdots + \Phi_q \mathbf{f}_{t-q} + \boldsymbol{\eta}_t, \quad \mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}).$$

We shall refer to this as the *expanded VMA form*. A theoretical justification for decomposing \mathbf{u}_t this way is provided by Peiris (1988, Theorem 2); the autovariances implied by this decomposition are

$$\overset{\circ}{\Gamma}_j = \sum_{l=j}^q \Phi_l \boldsymbol{\Omega} \Phi'_{l-j} + \mathbf{1}(j=0) \boldsymbol{\Lambda}, \quad j = 0, \dots, q, \quad (4)$$

and the mapping between $(\Theta_0, \dots, \Theta_q, \Sigma)$ and $(\Phi_0, \dots, \Phi_q, \Omega, \Lambda)$ is established by setting $\Gamma_j = \dot{\Gamma}_j$, i.e.,

$$\sum_{l=j}^q \Theta_l \Sigma \Theta'_{l-j} = \sum_{l=j}^q \Phi_l \Omega \Phi'_{l-j} + \mathbf{1}(j=0)\Lambda, \quad \text{for all } j = 0, \dots, q. \quad (5)$$

In practice, both representations require some normalization; we set $\Theta_0 = \mathbf{I}_n$, Ω and Λ both diagonal, and Φ_0 lower triangular with ones on the diagonal. We follow the convention in the classical literature and further assume that the VMA(q) process is invertible, i.e., the characteristic equation

$$\det(\mathbf{I}_n + \Theta_1 z + \dots + \Theta_q z^q)$$

has all roots outside the complex unit circle (note that this corresponds to all eigenvalues of $\tilde{\Theta}_1$ in (2) being less than one in modulus). However, as discussed in the next subsection, invertibility does not need to be imposed directly in the sampling scheme when working with the expanded VMA form, which greatly simplifies computation.

2.2 Identification in the Expanded VMA Form

It is clear that there are n additional parameters in the expanded form relative to the traditional VMA specification. Thus, to estimate the expanded form in practice, we must consider the “identification problem” generated by the expansion. Metaxoglou and Smith (2007) deal with this by fixing the elements of Λ to pre-determined values and estimate the remaining coefficients as free parameters. However, such a strategy might lead to mis-specification (e.g., it automatically imposes arbitrary lower bounds on the process variance). In fact, there seems to be no reasonable approach to “fixing parameters” in this context. Fortunately, it is not necessary either.

To clarify the basic ideas, consider the simplest case of a MA(1). The mapping between the two forms is defined by

$$\begin{aligned} \sigma^2(1 + \theta^2) &= \omega^2(1 + \phi^2) + \lambda^2, \\ \sigma^2\theta &= \omega^2\phi. \end{aligned}$$

Note further that in this case, we have $\sigma^2 > 0$, $\omega^2 > 0$, $\lambda^2 > 0$ and $-1 < \theta < 1$. It is easy to show, however, that these equalities and inequalities jointly imply

$$0 < \lambda^2 < \sigma^2(1 - |\theta|)^2. \quad (6)$$

Since θ, σ^2 are uniquely identified from data, λ^2 is always bounded within a finite interval—this interval is largest when $\theta = 0$ (and the model reduces to white noise) and shrinks towards zero as $\theta \rightarrow \pm 1$.

This fact has several important implications. Specifically, there always exists (except for the extreme case where $\theta = \pm 1$) some nonzero (positive) λ^2 such that any θ, σ^2 can be

recovered from $\phi, \omega^2, \lambda^2$. However, given a particular value of λ^2 , the reverse mapping from θ, σ^2 to ϕ, ω^2 is not well defined (i.e., it does not exist for all values of $\sigma^2 > 0$, $-1 < \theta < 1$). On the other hand, while many combinations of $\phi, \omega^2, \lambda^2$ will generally map to the same θ, σ^2 , only values of λ^2 satisfying (6) are admissible. Two implications follow:

1. arbitrarily fixing λ^2 at a particular value may lead to mis-specification;
2. $\phi, \omega^2, \lambda^2$ are all “partially” identified in the expanded form.

The first point demonstrates why a strategy such as the one employed by Metaxoglou and Smith (2007) might not be appropriate; the second suggests that leaving $\phi, \omega^2, \lambda^2$ as unrestricted parameters and employing Bayesian methods should work well in this context.

Before discussing further details, note that the above intuition generalizes in a straightforward way. For the VMA(1) case, the mapping is defined by

$$\begin{aligned}\Sigma + \Theta_1 \Sigma \Theta_1' &= \Lambda + \Phi_0 \Omega \Phi_0' + \Phi_1 \Omega \Phi_1', \\ \Theta_1 \Sigma &= \Phi_1 \Omega \Phi_0',\end{aligned}$$

where Σ is positive semi-definite (p.s.d.), $\omega_i^2 > 0$, $\lambda_i^2 > 0$ for all i , and Θ_1 restricted such that all eigenvalues are less than one in modulus. Using a result in linear algebra (Engwerda, Ran, and Rijkeboer, 1993), it follows that a necessary condition for the existence of a reverse mapping (given Λ) from (Θ_1, Σ) to (Φ_0, Φ_1, Ω) is that Λ simultaneously satisfies

$$\begin{aligned}(\mathbf{I}_n + \Theta_1) \Sigma (\mathbf{I}_n + \Theta_1)' - \Lambda &\text{ is p.s.d.}, \\ (\mathbf{I}_n - \Theta_1) \Sigma (\mathbf{I}_n - \Theta_1)' - \Lambda &\text{ is p.s.d.}\end{aligned}$$

Therefore, $\Theta_1 = \mathbf{0}$ corresponds to $\Sigma - \Lambda$ being positive semi-definite. As the modulus of any eigenvalue of Θ_1 approaches one, a corresponding eigenvalue of $(\mathbf{I}_n \pm \Theta_1) \Sigma (\mathbf{I}_n \pm \Theta_1)'$ approaches zero, and (by Weyl’s inequality) the interval on some λ_i^2 shrinks to zero as well. The extension to the VMA(q) is straightforward by employing the corresponding VMA(1) representation in (2).

Since $\lambda_1^2, \dots, \lambda_n^2$ are always finitely bounded, a Bayesian approach may generally proceed by

1. sampling $\Phi_0, \dots, \Phi_q, \Omega, \Lambda$ using the expanded form;
2. recovering $\Theta_1, \dots, \Theta_q, \Sigma$ ex-post from the simulated draws.

In this context, the above analysis suggests that working with the expanded form does not necessitate imposing invertibility restrictions on $\Phi_0, \dots, \Phi_q, \Omega, \Lambda$. Instead, these are

imposed *ex-post* in the Step 2. Furthermore, this approach does not necessitate restrictive priors on $\Phi_0, \dots, \Phi_q, \Omega, \Lambda$ either. In our empirical work, we set weakly informative priors on the (free) elements of $\Phi_0, \dots, \Phi_q, \Omega$ and the generic inverse-gamma prior on λ_i^2 :

$$\lambda_i^2 \sim \mathcal{IG}(\nu_{\lambda,0}, S_{\lambda,0}),$$

with $\nu_{\lambda,0}, S_{\lambda,0}$ set to low values. In consequence, λ_i^2 is generally allowed to move freely over a wide range of values, and therefore, imposes no *a priori* restrictions on the VMA coefficients. Moreover, in our extensive experimentation (with a variety of data sets and specifications) with this approach we have consistently found that mixing improves as $\nu_{\lambda,0}$ is decreased and the prior is flattened.

2.3 Recovering $(\Theta_1, \dots, \Theta_q, \Sigma)$ from $(\Phi_0, \dots, \Phi_q, \Omega, \Lambda)$

For the VMA(1) case, a draw of Θ_1, Σ can be easily recovered given a draw of $\Phi_0, \Phi_1, \Omega, \Lambda$ as follows. Proceed by first constructing $\Gamma_0 \equiv \dot{\Gamma}_0$ and $\Gamma_1 \equiv \dot{\Gamma}_1$ according to (4). Next, recover Θ_1 by solving the quadratic matrix equation

$$\Theta_1^2 \Gamma_1' - \Theta_1 \Gamma_0 + \Gamma_1 = \mathbf{0}.$$

Solving this equation is straightforward:

1. Solve the *generalized eigenvalue problem*

$$\begin{pmatrix} \Gamma_0 & -\Gamma_1' \\ \mathbf{I}_n & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \end{pmatrix} = \begin{pmatrix} \Gamma_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n \end{pmatrix} \begin{pmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{pmatrix}.$$

A solution to this will consist of $2n$ eigenvalues and the associated $2n$ eigenvectors.² Moreover, the matrices involving eigenvalues and eigenvectors can be rotated freely.

2. Choose \mathbf{D}_1 to contain n of the generated eigenvalues and $(\mathbf{X}'_{11}, \mathbf{X}'_{21})'$ to contain the corresponding eigenvectors. Note that by construction, exactly n of the eigenvalues will be less than one in modulus, so selecting the n smallest (in modulus) eigenvalues corresponds to enforcing invertibility (e.g., *ex-post*).
3. Compute

$$\begin{aligned} \Theta_1' &= \mathbf{X}_{11} \mathbf{X}_{21}^{-1}, \\ \Sigma &= \Gamma_0 - \Gamma_1 \Theta_1'. \end{aligned}$$

To recover $\Theta_1, \dots, \Theta_q, \Sigma$ in the general case, we appeal once again to the corresponding VMA(1) representation. Accordingly, the algorithm is as follows.

²Many modern statistical packages have built-in functions for computing generalized eigenvalues. In our empirical work, we use the command `eig` in MATLAB.

1. Compute $\Gamma_0, \dots, \Gamma_q$ from draws of $\Phi_0, \dots, \Phi_q, \Omega, \Lambda$.
2. Construct $\tilde{\Gamma}_0, \tilde{\Gamma}_1$ according to (3).
3. Use the VMA(1) procedure described above to compute Θ_1^*, Σ^* (i.e., the resulting Θ_1^*, Σ^* correspond to $\Theta_1^* = \tilde{\Theta}_1 \tilde{\Theta}_0^{-1}$ and $\Sigma^* = \tilde{\Theta}_0 (\mathbf{I}_q \otimes \Sigma) \tilde{\Theta}_0'$, respectively, with $\tilde{\Theta}_0, \tilde{\Theta}_1$ defined in (2)).
4. Recover $\Theta_1, \dots, \Theta_q$ from the first n columns of Θ_1^* and Σ from the bottom-right $n \times n$ block of Σ^* .

Note again that solving the generalized eigenvalue problem in Step 3 above leads to $2nq$ eigenvalues, of which nq are less than one in modulus and correspond an invertible VMA(q) system. Thus, the idea that invertibility can be enforced ex-post extends to the general case as well.

2.4 Extension to the VARMA

It is straightforward to generalize the above setup to VARMA. Specifically we add p autoregressive components to \mathbf{u}_t in (1) to obtain

$$\mathbf{y}_t = \sum_{j=1}^p \mathbf{A}_j \mathbf{y}_{t-j} + \sum_{j=1}^q \Theta_j \boldsymbol{\varepsilon}_{t-j} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma), \quad (7)$$

where the intercept is suppressed for notational convenience. Following the same procedure as before, we derive the *expanded VARMA form* by reparameterizing \mathbf{u}_t such that

$$\mathbf{y}_t = \sum_{j=1}^p \mathbf{A}_j \mathbf{y}_{t-j} + \sum_{j=0}^p \Phi_j \mathbf{f}_{t-j} + \boldsymbol{\eta}_t, \quad \mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \Omega) \text{ and } \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \Lambda). \quad (8)$$

Since we can easily recover $\Theta_1, \dots, \Theta_q, \Sigma$ from $\Phi_0, \dots, \Phi_q, \Omega, \Lambda$, it is clear that estimating (8) is sufficient to estimate (7).

It is important to point out, however, that introducing autoregressive components in this context leads to new complications. Specifically, it is well known that a VARMA(p, q) specified as in (7) may not be identified (see, for example, Lütkepohl, 2005). A number of ways have been proposed in the literature to deal with this. For example, one may impose the canonical echelon form by reformulating (7) as

$$\mathbf{B}_0 \mathbf{y}_t = \sum_{j=1}^p \mathbf{B}_j \mathbf{y}_{t-j} + \sum_{j=1}^q \Theta_j^* \boldsymbol{\varepsilon}_{t-j} + \mathbf{B}_0 \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma), \quad (9)$$

where \mathbf{B}_0 is lower-triangular with ones on the diagonal, $\mathbf{B}_j = \mathbf{B}_0 \mathbf{A}_j$ and $\Theta_j^* = \mathbf{B}_0 \Theta_j$. Conditional on the system's Kronecker indices $\kappa_1, \dots, \kappa_n$, with $0 \leq \kappa_i \leq p$, (9) can be estimated by imposing exclusion restrictions on the coefficients in $\mathbf{B}_0, \{\mathbf{B}_j\}$ and $\{\Theta_j^*\}$.

Observe that conditional on the Kronecker indices, it is straightforward to impose echelon form restrictions on the expanded VARMA as well, by rewriting (8) as

$$\mathbf{B}_0 \mathbf{y}_t = \sum_{j=1}^p \mathbf{B}_j \mathbf{y}_{t-j} + \sum_{j=0}^p \mathbf{\Phi}_j \mathbf{f}_{t-j} + \boldsymbol{\eta}_t, \quad \mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}) \text{ and } \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}). \quad (10)$$

Clearly, exclusion restrictions can be easily imposed on the elements of \mathbf{B}_0 , $\{\mathbf{B}_j\}$ and $\{\mathbf{\Phi}_j\}$ for sampling purposes (see the Appendix for more details). More importantly, the echelon form requires that only full-row restrictions are imposed on $\{\boldsymbol{\Theta}_j^*\}$. Because restricting any row of $\mathbf{\Phi}_j$ to zero will correspond to restricting the same row of $\boldsymbol{\Theta}_j^*$ to zero under (5), we can once again work directly with (10) and recover the parameters of (9) *ex post*.

In practice, of course, Kronecker indices will be unknown, and in our related paper (Chan, Eisenstat, and Koop, 2015), we construct efficient sampling procedures to jointly estimate Kronecker indices and model parameters using the expanded VARMA form. For the remainder of the present paper, however, we will rely on the identifying assumption that the concatenated matrix $[\mathbf{A}_p : \boldsymbol{\Theta}_q]$ has full rank n (e.g., Hannan, 1976). This assumption will generally hold when q is relatively small. The drawback of this approach is that the resulting representation may not be canonical (Lütkepohl and Poskitt, 1996). In our application, however, the main interest is to assess whether adding moving average terms to a VAR with stochastic volatility improves forecasts of macroeconomic variables. In this context, we find the above identification strategy to be suitable, and as further discussed in Section 5, adding even a small number of moving average components yields substantial gains in forecast accuracy.

3 Estimation

In this section, we introduce the conjugate priors for the parameters of the VARMA(p, q) in (8), followed by a discussion of an efficient Gibbs sampler. We note that throughout, the analysis is performed conditional on the initial observations $\mathbf{y}_0, \dots, \mathbf{y}_{1-p}$. One can extend the posterior sampler to the case where the initial observations are modeled explicitly. Moreover, we will assume that the VARMA is specified with an intercept term $\boldsymbol{\mu}$. Again, the ensuing algorithm is easily extended to include additional exogenous variables.

To facilitate estimation, we first rewrite (8) as

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta} + \mathbf{\Phi}_0 \mathbf{f}_t + \mathbf{\Phi}_1 \mathbf{f}_{t-1} + \dots + \mathbf{\Phi}_q \mathbf{f}_{t-q} + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{X}_t = \mathbf{I}_n \otimes (1, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})$, $\boldsymbol{\beta} = \text{vec}((\boldsymbol{\mu}, \mathbf{A}_1, \dots, \mathbf{A}_p)')$, $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})$, and $\mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$. Now, stacking the observations over t , we have

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{\Phi} \mathbf{f} + \boldsymbol{\varepsilon}, \quad (11)$$

where

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_T \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_T \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_T \end{pmatrix},$$

and Φ is a $Tn \times Tn$ lower triangular matrix with Φ_0 on the main diagonal block, Φ_1 on first lower diagonal block, Φ_2 on second lower diagonal block, and so forth. For example, for $q = 2$, we have

$$\Phi = \begin{pmatrix} \Phi_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \Phi_1 & \Phi_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \Phi_2 & \Phi_1 & \Phi_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Phi_2 & \Phi_1 & \Phi_0 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Phi_2 & \Phi_1 & \Phi_0 \end{pmatrix}.$$

Note that in general Φ is a band $Tn \times Tn$ matrix—i.e., its nonzero elements are confined in a narrow band along the main diagonal—that contains at most

$$n^2 \left((q+1)T - \frac{q(q+1)}{2} \right) < n^2(q+1)T$$

nonzero elements, which grows *linearly* in T and is substantially less than the total $(Tn)^2$ elements for typical applications where $T \gg q$. This special structure can be exploited to speed up computation, e.g., by using block-banded or sparse matrix algorithms (see, e.g., Kroese, Taimre, and Botev, 2011, p. 220).

To complete the model specification, we assume independent priors for β , Φ_0 , Φ_1, \dots, Φ_q , Ω , and Λ as follows. For β , we consider the multivariate normal prior $\mathcal{N}(\beta_0, \mathbf{V}_\beta)$. Let ϕ_i be a column vector that contains the free parameters in the i -th row of $(\Phi_0, \Phi_1, \dots, \Phi_q)$ —recall that Φ_0 is restricted to be lower triangular with ones on the diagonal. We assume the normal prior $\mathcal{N}(\phi_{0i}, \mathbf{V}_{\phi_i})$ for ϕ_i . For later reference we further let $\phi = (\phi'_1, \dots, \phi'_n)'$. Finally, for the diagonal elements of $\Omega = \text{diag}(\omega_1^2, \dots, \omega_n^2)$ and $\Lambda = \text{diag}(\lambda_1^2, \dots, \lambda_n^2)$, we assume the following independent inverse-gamma priors:

$$\omega_i^2 \sim \mathcal{IG}(\nu_{\omega_0}, S_{\omega_0}), \quad \lambda_i^2 \sim \mathcal{IG}(\nu_{\lambda_0}, S_{\lambda_0}).$$

Our approach to constructing the sampling algorithm is based on treating (11) as a latent factor model. Consequently, the Gibbs sampler proceeds by sequentially drawing from

1. $p(\beta, \mathbf{f} \mid \mathbf{y}, \phi, \Lambda, \Omega)$;
2. $p(\phi \mid \mathbf{y}, \beta, \mathbf{f}, \Lambda, \Omega)$;
3. $p(\Lambda, \Omega \mid \mathbf{y}, \beta, \mathbf{f}, \phi) = p(\Lambda \mid \mathbf{y}, \beta, \mathbf{f}, \phi)p(\Omega \mid \mathbf{y}, \beta, \mathbf{f}, \phi)$.

Steps 2 and 3 are standard, and we leave the details to the Appendix; here we focus on Step 1. Although it might be straightforward to sample β and \mathbf{f} separately by simulating $(\beta \mid \mathbf{y}, \mathbf{f}, \phi, \Lambda, \Omega)$ followed by $(\mathbf{f} \mid \mathbf{y}, \beta, \phi, \Lambda, \Omega)$, such an approach would potentially induce high autocorrelation and slow mixing in the constructed Markov chain as β and \mathbf{f} enter (11) additively. Instead, we aim to sample β and \mathbf{f} jointly—by first drawing $(\beta \mid \mathbf{y}, \phi, \Lambda, \Omega)$ marginally of \mathbf{f} , followed by $(\mathbf{f} \mid \mathbf{y}, \beta, \phi, \Lambda, \Omega)$.

To implement the first step, we note that since $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_T \otimes \mathbf{\Omega})$, the joint density of \mathbf{y} marginal of \mathbf{f} is

$$(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\phi}, \mathbf{\Lambda}) \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{S}_y),$$

where $\mathbf{S}_y = \mathbf{I}_T \otimes \mathbf{\Lambda} + \boldsymbol{\Phi}(\mathbf{I}_T \otimes \mathbf{\Omega})\boldsymbol{\Phi}'$. Since both $\mathbf{\Lambda}$ and $\mathbf{\Omega}$ are diagonal matrices, and $\boldsymbol{\Phi}$ is a lower triangular sparse matrix, the covariance matrix is also banded. Using standard results from linear regression (see, e.g., Kroese and Chan, 2014, p. 239-240), we have

$$(\boldsymbol{\beta} | \mathbf{y}, \boldsymbol{\phi}, \mathbf{\Lambda}, \mathbf{\Omega}) \sim \mathcal{N}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_\beta),$$

where

$$\mathbf{D}_\beta = (\mathbf{V}_\beta^{-1} + \mathbf{X}'\mathbf{S}_y^{-1}\mathbf{X})^{-1}, \quad \widehat{\boldsymbol{\beta}} = \mathbf{D}_\beta (\mathbf{V}_\beta^{-1}\boldsymbol{\beta}_0 + \mathbf{X}'\mathbf{S}_y^{-1}\mathbf{y}).$$

It is important to realize that in order to compute $\mathbf{X}'\mathbf{S}_y^{-1}\mathbf{X}$ or $\mathbf{X}'\mathbf{S}_y^{-1}\mathbf{y}$, one needs not obtain the $Tn \times Tn$ matrix \mathbf{S}_y^{-1} —otherwise it would involve $\mathcal{O}(T^3)$ operations. Instead, we exploit the fact that the covariance matrix \mathbf{S}_y is banded. For instance, to obtain $\mathbf{X}'\mathbf{S}_y^{-1}\mathbf{y}$, we can first solve the system $\mathbf{S}_y\mathbf{z} = \mathbf{y}$ for \mathbf{z} , which can be computed in $\mathcal{O}(T)$ operations. The solution is $\mathbf{z} = \mathbf{S}_y^{-1}\mathbf{y}$ and we return $\mathbf{X}'\mathbf{z} = \mathbf{X}'\mathbf{S}_y^{-1}\mathbf{y}$, which is the desired quantity. Similarly, $\mathbf{X}'\mathbf{S}_y^{-1}\mathbf{X}$ can be computed quickly without inverting any big matrices.

Next, we sample all the latent factors \mathbf{f} jointly. Note that even though *a priori* the latent factors are independent, they are no longer independent given \mathbf{y} . As such, sampling each \mathbf{f}_t sequentially would potentially induce high autocorrelation and slow mixing in the Markov chain. One could sample \mathbf{f} using Kalman filter-based algorithms, but they would involve redefining the states so that only the state at time t enters the measurement equation at time t . As such, each (new) state vector would be of much higher dimension, which in turn results in slower algorithms. Instead, we avoid the Kalman filter and instead implement the precision-based sampler developed in Chan and Jeliazkov (2009) to sample the latent factors jointly. To that end, recall that *a priori* $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_T \otimes \mathbf{\Omega})$. Using (11) and standard linear regression results again, we have

$$(\mathbf{f} | \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\phi}, \mathbf{\Lambda}, \mathbf{\Omega}) \sim \mathcal{N}(\widehat{\mathbf{f}}, \mathbf{K}_f^{-1}),$$

where

$$\mathbf{K}_f = \mathbf{I}_T \otimes \mathbf{\Omega}^{-1} + \boldsymbol{\Phi}'(\mathbf{I}_T \otimes \mathbf{\Lambda}^{-1})\boldsymbol{\Phi}, \quad \widehat{\mathbf{f}} = \mathbf{K}_f^{-1}\boldsymbol{\Phi}'(\mathbf{I}_T \otimes \mathbf{\Lambda}^{-1})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

The challenge, of course, is that the covariance matrix \mathbf{K}_f^{-1} is a $Tn \times Tn$ full matrix, and sampling $(\mathbf{f} | \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\phi}, \mathbf{\Lambda}, \mathbf{\Omega})$ using brute force is infeasible. However, the precision matrix \mathbf{K}_f is banded (recall that both $\mathbf{\Lambda}^{-1}$ and $\mathbf{\Omega}^{-1}$ are diagonal, whereas $\boldsymbol{\Phi}$ is also banded), which can be exploited to speed up computation. As before, we first obtain $\widehat{\mathbf{f}}$ by solving

$$\mathbf{K}_f\mathbf{z} = \boldsymbol{\Phi}'(\mathbf{I}_T \otimes \mathbf{\Lambda}^{-1})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

for \mathbf{z} . Next, obtain the Cholesky decomposition of \mathbf{K}_f such that $\mathbf{C}_f\mathbf{C}_f' = \mathbf{K}_f$. Solve $\mathbf{C}_f'\mathbf{z} = \mathbf{u}$ for \mathbf{z} , where $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{Tn})$. Finally, return $\mathbf{f} = \widehat{\mathbf{f}} + \mathbf{z}$, which follows the desired distribution. We refer the readers to Chan and Jeliazkov (2009) for details.

4 Extensions: TVP and SV

As highlighted in the preceding section, the key advantage of working directly with the expanded VARMA form is that it is conditionally linear, and therefore, leads to straightforward computation. This in turn opens the door to a wealth of extensions that have already been well developed for linear models, but have thus far been inaccessible for even the simplest of VMA specifications. Our particular interest is to enhance the basic VARMA(p, q) with two extensions particularly relevant for empirical macroeconomic applications: stochastic volatility and time-varying parameters.

First, we relax the assumption of constant variance and incorporate stochastic volatility into the model. Specifically, consider again the VARMA(p, q) specified in (11). Now, we allow the latent factors to have time-varying volatilities $\mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_t)$, where $\mathbf{\Omega}_t = \text{diag}(e^{h_{1t}}, \dots, e^{h_{nt}})$, and each of the log-volatility follows an independent random walk:

$$h_{it} = h_{i,t-1} + \zeta_{it}, \quad (12)$$

where $\zeta_{it} \sim \mathcal{N}(0, \psi_{h_i}^2)$ for $i = 1, \dots, n, t = 2, \dots, T$. The log-volatilities are initialized with $h_{i1} \sim \mathcal{N}(h_{i0}, V_{h_{i1}})$, where h_{i0} and $V_{h_{i1}}$ are known constants. For notational convenience, let $\mathbf{h}_t = (h_{1t}, \dots, h_{nt})'$, $\mathbf{h} = (\mathbf{h}'_1, \dots, \mathbf{h}'_T)'$ and $\boldsymbol{\psi}_h^2 = (\psi_{h_1}^2, \dots, \psi_{h_n}^2)'$. Note that in terms of (7), this will induce time variation in both the covariance $\boldsymbol{\Sigma}$ and the VMA coefficients $\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_q$, albeit in a restricted way.

Estimation of this VARMA(p, q) with stochastic volatility would only require minor modifications of the basic Gibbs sampler developed above. More specifically, posterior draws can be obtained by sequentially sampling from

1. $p(\boldsymbol{\beta}, \mathbf{f} \mid \mathbf{y}, \boldsymbol{\phi}, \mathbf{h}, \boldsymbol{\psi}_h^2, \boldsymbol{\Lambda});$
2. $p(\boldsymbol{\phi} \mid \mathbf{y}, \boldsymbol{\beta}, \mathbf{f}, \mathbf{h}, \boldsymbol{\psi}_h^2, \boldsymbol{\Lambda});$
3. $p(\mathbf{h} \mid \mathbf{y}, \boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\phi}, \boldsymbol{\psi}_h^2, \boldsymbol{\Lambda});$
4. $p(\boldsymbol{\Lambda}, \boldsymbol{\psi}_h^2 \mid \mathbf{y}, \boldsymbol{\beta}, \mathbf{f}, \mathbf{h}, \boldsymbol{\phi}) = p(\boldsymbol{\Lambda} \mid \mathbf{y}, \boldsymbol{\beta}, \mathbf{f}, \mathbf{h}, \boldsymbol{\phi})p(\boldsymbol{\psi}_h^2 \mid \mathbf{y}, \boldsymbol{\beta}, \mathbf{f}, \mathbf{h}, \boldsymbol{\phi}).$

Steps 1 and 2 proceed as before with minor modifications—e.g., the prior covariance matrix for \mathbf{f} is now $\text{diag}(e^{h_{11}}, \dots, e^{h_{n1}}, \dots, e^{h_{1T}}, \dots, e^{h_{nT}})$ instead of $\mathbf{I}_T \otimes \boldsymbol{\Omega}$. But since the covariance matrix remains diagonal, all the band matrix routines are as fast as before. Step 3 can be carried out, for example, using the auxiliary mixture sampling approach of Kim, Shepherd, and Chib (1998). Lastly, Step 4 is also standard if we assume independent inverse-gamma priors for $\lambda_{h_1}^2, \dots, \lambda_{h_n}^2$.

We can further enrich this specification by allowing the matrices $\boldsymbol{\Phi}_0, \dots, \boldsymbol{\Phi}_q$ to be time-varying:

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\Phi}_{0t} \mathbf{f}_t + \boldsymbol{\Phi}_{1t} \mathbf{f}_{t-1} + \dots + \boldsymbol{\Phi}_{qt} \mathbf{f}_{t-q} + \boldsymbol{\varepsilon}_t, \quad (13)$$

where \mathbf{X}_t and $\boldsymbol{\beta}$ are defined as before, $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})$ and $\boldsymbol{\Lambda}$ is a diagonal matrix. Let $\boldsymbol{\phi}_{it}$ denote the vector of free parameters in the i -th row of $(\boldsymbol{\Phi}_{0t}, \boldsymbol{\Phi}_{1t}, \dots, \boldsymbol{\Phi}_{qt})$. Then, consider the transition equation

$$\boldsymbol{\phi}_{it} = \boldsymbol{\phi}_{i,t-1} + \boldsymbol{\eta}_{it},$$

where $\boldsymbol{\eta}_{it} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i)$ for $i = 1, \dots, n, t = 2, \dots, T$. The initial conditions are specified as $\boldsymbol{\phi}_{it} \sim \mathcal{N}(\boldsymbol{\phi}_{i0}, \mathbf{Q}_{i0})$, where $\boldsymbol{\phi}_{i0}$ and \mathbf{Q}_{i0} are known constant matrices. This additional generalization will correspond to fully time-varying $\boldsymbol{\Theta}_{1,t}, \dots, \boldsymbol{\Theta}_{j,t}$ and $\boldsymbol{\Sigma}_t$.

Since $\boldsymbol{\Lambda}$ is a diagonal matrix and each $(\boldsymbol{\phi}'_{i1}, \dots, \boldsymbol{\phi}'_{iT})'$ is defined by a linear Gaussian state space model, sampling from the relevant conditional distribution can be done quickly by the precision sampler. In addition, sampling all the other parameters can be carried out as in the basic sampler with minor modifications. For example, to sample $(\boldsymbol{\beta}, \mathbf{f})$ jointly, note that we can stack (13) over t and write

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \tilde{\boldsymbol{\Phi}}\mathbf{f} + \boldsymbol{\varepsilon},$$

where

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_T \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_T \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_T \end{pmatrix},$$

and $\tilde{\boldsymbol{\Phi}}$ is a $Tn \times Tn$ lower triangular matrix with $\boldsymbol{\Phi}_{01}, \dots, \boldsymbol{\Phi}_{0T}$ on the main diagonal block, $\boldsymbol{\Phi}_{12}, \dots, \boldsymbol{\Phi}_{1T}$ on first lower diagonal block, $\boldsymbol{\Phi}_{23}, \dots, \boldsymbol{\Phi}_{2T}$ on second lower diagonal block, and so forth. The algorithms discussed before can then be applied directly.

5 Empirical Application

In this section we illustrate the proposed approach and estimation methods with a recursive forecasting exercise that involves US CPI inflation and real GDP growth. These two variables are commonly used in forecasting (e.g., Banbura, Giannone, and Reichlin, 2010; Koop, 2011) and small DSGE models (e.g., An and Schorfheide, 2007). We first outline the set of competing models in Section 5.1, followed by a brief description of the data and the priors. The results of the density forecasting exercise are reported in Section 5.2.

5.1 Competing Models, Data and Prior

The main goal of this forecasting exercise is to illustrate the methodology and investigate how VARMA and the variants with stochastic volatility compare with standard VARs. We consider three sets of VARMA: VARMA($p, 1$) and two versions with different time-varying VMA coefficients and volatility. The VARMA($p, 1$) is the same as given in (8). In the first version with time-varying volatility, we allow the latent factors in (8) to have a stochastic volatility component: $\mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_t)$, where $\boldsymbol{\Omega}_t = \text{diag}(e^{h_{1t}}, \dots, e^{h_{nt}})$, and

each of the log-volatility follows an independent random walk as in (12). We call this version VARMA($p, 1$)-SV1.

In the second version, we further allow the matrices Φ_0, \dots, Φ_q to be time-varying as specified in (13). This more general version is denoted as VARMA($p, 1$)-SV2. For comparison we also include standard VAR(p) and VAR(p) with stochastic volatility, where the latter is denoted as VAR(p)-SV.

The data consist of US quarterly CPI inflation and real GDP growth from 1959:Q1 to 2011:Q4. More specifically, given the quarterly real GDP series w_{1t} , we transform it via $y_{1t} = 400 \log(w_{1t}/w_{1,t-1})$ to obtain the growth rate. We perform a similar transformation to the CPI index to get the inflation rate. For easy comparison, we choose broadly similar priors across models. For instance, the priors for the VAR coefficients in VARMA specifications are exactly the same as those of the corresponding VAR.

As discussed in Section 3, we assume the following independent priors: $\beta \sim \mathcal{N}(\beta_0, \mathbf{V}_\beta)$, $\phi_i \sim \mathcal{N}(\phi_{0i}, \mathbf{V}_{\phi_i})$, $\omega_i^2 \sim \mathcal{IG}(\nu_{\omega_0}, S_{\omega_0})$ and $\lambda_i^2 \sim \mathcal{IG}(\nu_{\lambda_0}, S_{\lambda_0})$, $i = 1, \dots, n$. We set $\beta_0 = \mathbf{0}$ and set the prior covariance \mathbf{V}_β to be diagonal, where the variances associated with the intercepts are 100 and those corresponding to the VAR coefficients are 1. For ϕ_i , we set $\phi_{0i} = \mathbf{0}$, and set \mathbf{V}_{ϕ_i} to be the identity matrix. We choose relatively small values for the degrees of freedom and scale hyperparameters for ω_i^2 , which imply weakly informative priors: $\nu_{\omega_0} = 3$, $S_{\omega_0} = 2$. These values imply $\mathbb{E}\omega_i^2 = 1$, $i = 1, \dots, n$. For VARMA with stochastic volatility, we also need to specify priors for ψ_{hi}^2 . We assume $\psi_{hi}^2 \sim \mathcal{IG}(\nu_{hi}, S_{hi})$, where $\nu_{hi} = 5$ and $S_{hi} = 0.04$, which implies $\mathbb{E}\psi_{hi}^2 = 0.01$. Finally for each λ_i^2 , we specify the noninformative prior $\nu_{\lambda_0} = 0$, $S_{\lambda_0} = 0.1$.

5.2 Forecasting Results

To compare the performance of the competing models in producing density forecasts, we consider a recursive out-of-sample forecasting exercise at various forecast horizons as follows. At the t -th iteration, for each of the model we use data up to time t , denoted as $\mathbf{y}_{1:t}$, to construct the joint predictive density $p(\mathbf{y}_{t+k} | \mathbf{y}_{1:t})$ under the model, and use it as the k -step-ahead density forecast for \mathbf{y}_{t+k} . We then expand the sample using data up to time $t + 1$, and repeat the whole exercise. We continue this procedure until time $T - k$. At the end of the iterations, we obtain density forecasts under the competing models for $t = t_0, \dots, T - k$, where we set t_0 to be 1975Q1.

The joint predictive density $p(\mathbf{y}_{t+k} | \mathbf{y}_{1:t})$ is not available analytically, but it can be estimated using MCMC methods. For VARs and VARMA, the conditional density of \mathbf{y}_{t+k} given the data and the model parameters—denoted as $p(\mathbf{y}_{t+k} | \mathbf{y}_{1:t}, \boldsymbol{\theta})$ —is Gaussian with known mean vector and covariance matrix. Hence, the predictive density can be estimated by averaging $p(\mathbf{y}_{t+k} | \mathbf{y}_{1:t}, \boldsymbol{\theta})$ over the MCMC draws of $\boldsymbol{\theta}$. For VARMA with stochastic volatility, at every MCMC iteration given the model parameters and all the states up to time t , we simulate future log-volatilities from time $t + 1$ to $t + k$ using the transition equation. Given these draws, \mathbf{y}_{t+k} has a Gaussian density. Finally, these

Gaussian densities are averaged over the MCMC iterations to obtain the joint predictive density $p(\mathbf{y}_{t+k} | \mathbf{y}_{1:t})$.

To evaluate the quality of the joint density forecast, consider the predictive likelihood $p(\mathbf{y}_{t+k} = \mathbf{y}_{t+k}^o | \mathbf{y}_{1:t})$, i.e., the joint predictive density of \mathbf{y}_{t+k} evaluated at the observed value \mathbf{y}_{t+k}^o . Intuitively, if the actual outcome \mathbf{y}_{t+k}^o is unlikely under the density forecast, the value of the predictive likelihood will be small, and vice versa. We then evaluate the joint density forecasts using the sum of log predictive likelihoods as a metric:

$$\sum_{t=t_0}^{T-k} \log p(\mathbf{y}_{t+k} = \mathbf{y}_{t+k}^o | \mathbf{y}_{1:t}).$$

This measure can also be viewed as an approximation of the log marginal likelihood; see, e.g., Geweke and Amisano (2011) for a more detailed discussion. In addition to assessing the joint density forecasts, we also evaluate the performance of the models in terms of the forecasts of each individual series. For example, to evaluate the performance for forecasting the i -th component of \mathbf{y}_{t+k} , we simply replace the joint density $p(\mathbf{y}_{t+k} = \mathbf{y}_{t+k}^o | \mathbf{y}_{1:t})$ by the marginal density $p(y_{i,t+k} = y_{i,t+k}^o | \mathbf{y}_{1:t})$, and report the corresponding sum.

Table 1: Relative log predictive likelihoods for 1-quarter-ahead density forecasts (compared to the random walk model).

	Inflation	GDP	Joint
VAR(1)	73.7	24.6	99.9
VARMA(1,1)	85.4	23.4	109.0
VAR(2)	89.2	24.7	114.5
VAR(2)-SV	114.9	23.8	143.6
VARMA(2,1)	94.0	22.2	116.4
VARMA(2,1)-SV1	121.4	34.0	158.4
VARMA(2,1)-SV2	103.2	23.1	133.0
VAR(3)	86.0	22.3	111.0
VAR(3)-SV	113.7	21.9	139.3
VARMA(3,1)	92.0	21.6	114.6
VARMA(3,1)-SV1	119.6	32.9	153.3
VARMA(3,1)-SV2	100.0	21.3	123.7

Table 1 reports the performance of the twelve competing models for producing 1-quarter-ahead joint density forecasts, as well as the marginal density forecasts for the two components. These values are presented relative to those of the random walk benchmark: $\mathbf{y}_t = \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$, where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Hence, any positive values indicate better forecast performance than the benchmark and vice versa.

A few broad conclusions can be drawn from the results. For joint density forecasts, adding a moving average component improves the forecast performance of standard VARs

for all lag lengths considered. For example, the difference between the log predictive likelihoods of VAR(3) and VARMA(3,1) is 3.6, which may be interpreted as a Bayes factor of about 37 in favor of the VARMA(3,1) model. In addition, adding stochastic volatility to the models further improves their forecast performance substantially. This is in line with the large literature on inflation forecasting that shows the considerable benefits of allowing for stochastic volatility for both point and density forecasts (see, e.g., Stock and Watson, 2007; Chan, Koop, Leon-Gonzalez, and Strachan, 2012; Clark and Doh, 2014). It is also interesting to note that VARMA($p,1$)-SV1 always outperform the more general VARMA($p,1$)-SV2, indicating that allowing for time-variation in $\mathbf{\Omega}_t$ is sufficient. Overall, VARMA(2,1)-SV1 forecasts better than all other models.

Next, to investigate the source of differences in forecast performance, we look at the log predictive likelihoods for each series. The results suggest that the gain in adding the moving average and stochastic volatility components comes mainly from forecasting inflation better. In fact, all the models perform very similarly in forecasting GDP growth—with the exception that VARMA($p,1$)-SV1 does better than the rest.

Tables 2–3 present the results for 2- and 3-quarter-ahead density forecasts, respectively. For longer horizons, the advantage of VARMA over VARs is less obvious—it is perhaps not surprising as the VMA has only one lag. What remains to be important is allowing for stochastic volatility. Both VAR(p)-SV and VARMA($p,1$)-SV1 outperform the rest of the models. Among the two, the former models perform slightly better for 2-quarter-ahead forecasts, whereas the latter are better for 3-quarter-ahead forecasts. Similar to 1-quarter-ahead results, most of the gains seem to have come from forecasting inflation better; all models perform quite similarly in forecasting GDP growth.

Table 2: Relative log predictive likelihoods for 2-quarter-ahead density forecasts (compared to the random walk model).

	Inflation	GDP	Joint
VAR(1)	48.0	52.1	101.2
VARMA(1,1)	57.4	52.4	110.5
VAR(2)	63.8	53.1	117.7
VAR(2)-SV	83.1	53.6	140.6
VARMA(2,1)	66.1	51.3	118.5
VARMA(2,1)-SV1	80.8	54.6	138.9
VARMA(2,1)-SV2	64.0	50.0	118.6
VAR(3)	60.1	52.2	114.6
VAR(3)-SV	82.1	52.4	138.5
VARMA(3,1)	62.6	51.2	116.4
VARMA(3,1)-SV1	79.5	55.0	135.7
VARMA(3,1)-SV2	52.1	48.6	105.5

Table 3: Relative log predictive likelihoods for 3-quarter-ahead density forecasts (compared to the random walk model).

	Inflation	GDP	Joint
VAR(1)	43.1	75.0	117.9
VARMA(1,1)	42.8	75.2	118.3
VAR(2)	43.5	74.7	118.0
VAR(2)-SV	67.6	73.1	143.5
VARMA(2,1)	46.8	72.9	119.8
VARMA(2,1)-SV1	70.0	77.6	147.9
VARMA(2,1)-SV1	57.9	70.2	127.2
VAR(3)	41.6	73.5	115.0
VAR(3)-SV	68.0	71.2	141.5
VARMA(3,1)	48.4	70.7	119.3
VARMA(3,1)-SV1	70.2	71.0	141.7
VARMA(3,1)-SV2	55.4	68.4	120.4

Figure 1 allows us to investigate the forecast performance of the competing models in more detail by plotting the cumulative sums of log predictive likelihoods over the whole evaluation period (relative to the random walk model). For 1-quarter-ahead forecasts, it is clear that overall VARMA—*with or without stochastic volatility*—consistently perform better than VARs.

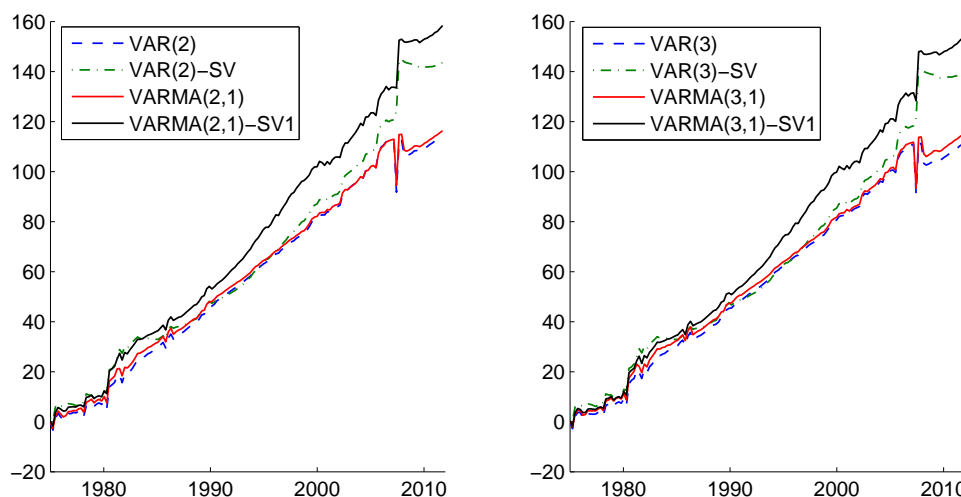


Figure 1: Cumulative sums of log predictive likelihoods for jointly forecasting inflation and GDP growth relative to the random walk model; one-quarter-ahead forecasts.

The figure also reveals some interesting patterns over time. For example, during the Great Recession, the performance of both VARs and VARMA with constant volatility

deteriorates against the random walk model, whereas models with stochastic volatility perform substantially better. This again highlights the importance of allowing for stochastic volatility, especially during turbulent times.

6 Concluding Remarks and Future Research

We build on a recently introduced latent factors representation of the VARMA model and derive some theoretical properties of the *expanded VARMA form* that justify its use in a Bayesian framework. On this foundation, we have developed a straightforward Gibbs sampler for the model and discussed how this algorithm can be extended to models with time-varying VMA coefficients and stochastic volatility. The proposed methodology was demonstrated using a density forecasting exercise, in which we also showed that VARMA with stochastic volatility forecast better standard VARs with stochastic volatility.

The methodology developed in this article leads to a few lines of inquiry in the future. In Chan, Eisenstat, and Koop (2015), we develop algorithms that facilitate exact inference from echelon form VARMA specifications (with unknown Kronecker indices) and demonstrate that these can be readily used to estimate systems with as many as twelve equations. An interesting result we obtain is that moving average components and the canonical form gain in importance as the system size increases, even when compared to parsimonious Bayesian VARs with shrinkage priors. However, in that work we do not consider extensions such as stochastic volatility, and this would be one point of interest for further research.

In addition, empirical investigation of fully time-varying parameter VARMA and regime-switching VARMA also seems beneficial. Moreover, the expanded VARMA form bears a close resemblance to the popular FAVAR. In fact, it can be shown that reducing the number of “factors” in this VARMA representation together with a particular set of restriction on the model parameters leads to exactly the FAVAR specification. It would therefore be of interest to further investigate how this relates to the recent work of Dufour and Stevanović (2013), as well as how alternative identification restrictions compare to the standard ones used in the literature.

Appendix

In this appendix we discuss the details of Steps 2 and 3 in the basic Gibbs sampler for the VARMA(p, q) model in the expanded form given by (8). To sample $(\phi | \mathbf{y}, \boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\Lambda}, \boldsymbol{\Omega})$, note that the innovation $\boldsymbol{\varepsilon}_t$ has a diagonal covariance matrix $\boldsymbol{\Lambda}$. Hence, we can estimate $\boldsymbol{\Phi}_0, \dots, \boldsymbol{\Phi}_q$ equation by equation. To that end, define $\mathbf{y}_t^* = \mathbf{y}_t - \boldsymbol{\mu} - \mathbf{A}_1 \mathbf{y}_{t-1} - \dots - \mathbf{A}_p \mathbf{y}_{t-p}$, and let y_{it}^* denote the i -th element of \mathbf{y}_t^* . As discussed in Section 2, we need to impose linear restrictions on the elements of $\boldsymbol{\Phi}_0$ and it will often be of interest to further impose

exclusion restrictions on $\mathbf{A}_1, \dots, \mathbf{A}_p$ (alternatively, $\mathbf{B}_0, \dots, \mathbf{B}_p$ for the echelon form) and Φ_1, \dots, Φ_q as well. To economize on space, we explicitly discuss implementing such restrictions on Φ_1, \dots, Φ_q ; sampling $\mathbf{A}_1, \dots, \mathbf{A}_p$ subject to similar restrictions follows analogously. Let $\phi_{j,i}$ denote the i -th column of Φ'_j , and accordingly, $\phi_{j,i}^*$ be the free elements in $\phi_{j,i}$, such that

$$\begin{aligned}\phi_{0,i} &= \mathbf{R}_{0,i} \phi_{0,i}^* + \boldsymbol{\nu}_i, \\ \phi_{j,i} &= \mathbf{R}_{j,i} \phi_{j,i}^*, \quad \text{for } j \geq 1,\end{aligned}$$

where $\boldsymbol{\nu}_i$ is an $n \times 1$ vector with the i -th element $\nu_{ii} = 1$ and all others set to zero. $\mathbf{R}_{j,i}$ in this context is a pre-determined selection matrix of appropriate dimensions, whereas our assumptions regarding Φ_0 correspond to $\mathbf{R}_{0,i} = (\mathbf{I}_{i-1}, \mathbf{0})'$ for $i > 1$ (for $i = 1$, it is clear that $\phi_{0,1} = \boldsymbol{\nu}_1$ and there are no free elements). Using this formulation, define $\mathbf{R}_i = \text{diag}(\mathbf{R}_{0,1}, \mathbf{R}_{1,1}, \dots, \mathbf{R}_{q,i})$ and $\phi_i^* = (\phi_{0,i}^{*'}, \phi_{1,i}^{*'} \dots, \phi_{q,i}^{*'})'$, such that $\phi_i = \mathbf{R}_i \phi_i^* + (\boldsymbol{\nu}_i', \mathbf{0})'$ forms the i -th column of $(\Phi_0, \Phi_1, \dots, \Phi_q)'$.

Then, the i -th equation in (8) can be written as

$$y_{it}^* = f_{it} + \tilde{\mathbf{w}}_t' \mathbf{R}_i \phi_i^* + \varepsilon_{it},$$

where $\tilde{\mathbf{w}}_t = (\mathbf{f}'_t, \mathbf{f}'_{t-1}, \dots, \mathbf{f}'_{t-q})$, f_{it} is the i -th element of \mathbf{f}_t , and ε_{it} is the i -th element of $\boldsymbol{\varepsilon}_t$. Denoting further $\mathbf{w}_{it} = \tilde{\mathbf{w}}_t' \mathbf{R}_i$, standard results dictate

$$(\phi_i^* | \mathbf{y}, \boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\Lambda}, \boldsymbol{\Omega}) \sim \mathcal{N}(\hat{\phi}_i^*, \mathbf{D}_{\phi_i^*}),$$

where

$$\mathbf{D}_{\phi_i^*} = \left(\mathbf{V}_{\phi_i^*}^{-1} + \frac{1}{\lambda_i^2} \sum_{t=1}^T \mathbf{w}'_{it} \mathbf{w}_{it} \right)^{-1}, \quad \hat{\phi}_i^* = \mathbf{D}_{\phi_i^*} \left(\mathbf{V}_{\phi_i^*}^{-1} \phi_{0i}^* + \frac{1}{\lambda_i^2} \sum_{t=1}^T \mathbf{w}'_{it} (y_{it}^* - f_{it}) \right),$$

$\mathbf{V}_{\phi_i^*}$ is the prior variance for the elements in ϕ_i^* , while ϕ_{0i}^* is the prior mean of ϕ_i^* , and λ_i^2 is the i -th diagonal element of $\boldsymbol{\Lambda}$.

For Step 2, note that both $p(\boldsymbol{\Omega} | \mathbf{y}, \boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\phi})$ and $p(\boldsymbol{\Lambda} | \mathbf{y}, \boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\phi})$ are products of inverse-gamma densities, and can therefore be sampled using standard methods. In fact, we have

$$\begin{aligned}(\omega_i^2 | \mathbf{y}, \boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\Phi}) &\sim \text{IG} \left(\nu_{\omega,0} + T/2, S_{\omega,0} + \sum_{t=1}^T f_{it}^2/2 \right), \\ (\lambda_i^2 | \mathbf{y}, \boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\Phi}) &\sim \text{IG} \left(\nu_{\lambda,0} + T/2, S_{\lambda,0} + \sum_{t=1}^T (y_{it}^* - f_{it} - \mathbf{w}_{it} \phi_i)^2/2 \right).\end{aligned}$$

References

- S. An and F. Schorfheide. Bayesian analysis of DSGE models. *Econometric Reviews*, 26(2-4):113–172, 2007.
- G. Athanasopoulos and F. Vahid. VARMA versus VAR for macroeconomic forecasting. *Journal of Business and Economic Statistics*, 26(2):237–252, 2008.
- G. Athanasopoulos, D. S. Poskitt, and F. Vahid. Two canonical VARMA forms: Scalar component models vis-à-vis the echelon form. *Econometric Reviews*, 31(1):60–83, 2012.
- M. Banbura, D. Giannone, and L. Reichlin. Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92, 2010.
- B. Bernanke, J. Boivin, and P. S. Elias. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 120(1):387–422, 2005.
- F. Canova. Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model. *Journal of Economic Dynamics and Control*, 17:233–261, 1993.
- J. C. C. Chan. Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, 176(2):162–172, 2013.
- J. C. C. Chan and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1:101–120, 2009.
- J. C. C. Chan, G. Koop, R. Leon-Gonzalez, and R. Strachan. Time varying dimension models. *Journal of Business and Economic Statistics*, 30:358–367, 2012.
- J. C. C. Chan, E. Eisenstat, and G. Koop. Large Bayesian VARMA. Working Paper, 2015.
- T. E. Clark and T. Doh. A Bayesian evaluation of alternative models of trend inflation. *International Journal of Forecasting*, 30(3):426–448, 2014.
- T. Cogley and T. J. Sargent. Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262 – 302, 2005.
- T. F. Cooley and M. Dwyer. Business cycle analysis without much theory: A look at structural VARs. *Journal of Econometrics*, 83(12):57–88, 1998.
- J.-M. Dufour and D. Pelletier. Practical methods for modelling weak VARMA processes: identification, estimation and specification with a macroeconomic application. Discussion Paper, McGill University, 2014.
- J.-M. Dufour and D. Stevanović. Factor-augmented VARMA models with macroeconomic applications. *Journal of Business & Economic Statistics*, 31(4):491–506, 2013.

- E. Eisenstat, J. C. C. Chan, and R. W. Strachan. Model specification search for time-varying parameter VARs. *Econometric Reviews*, 2014. Forthcoming.
- J. C. Engwerda, A. C. M. Ran, and A. L. Rijkeboer. Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation $X + A^*X^{-1}A = Q$. *Linear Algebra and its Applications*, 186:255–275, 1993.
- J. Fernández-Villaverde, J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson. ABCs (and Ds) of understanding VARs. *American Economic Review*, 97(3):1021–1026, 2007.
- E. I. George and R. E. McCulloch. Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.
- J. Geweke and G. Amisano. Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, 26:1–29, 2011.
- E. J. Hannan. The identification and parameterization of ARMAX and state space forms. *Econometrica*, 44(4):713–723, 1976.
- C. Kascha. A comparison of estimation methods for vector autoregressive moving-average models. *Econometric Reviews*, 31(3):297–324, 2012.
- C. Kascha and C. Trenkler. Simple identification and specification of cointegrated VARMA models. *Journal of Applied Econometrics*, 2014. Forthcoming.
- S. Kim, N. Shepherd, and S. Chib. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65(3):361–393, 1998.
- G. Koop. Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 2011. DOI: 10.1002/jae.1270.
- G. Koop and D. Korobilis. Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3(4):267–358, 2010.
- G. Koop and D. Korobilis. Large time-varying parameter VARs. *Journal of Econometrics*, 177(2):185–198, 2013.
- G. Koop and S. M. Potter. Estimation and forecasting in models with multiple breaks. *Review of Economic Studies*, 74:763–789, 2007.
- D. Korobilis. Assessing the transmission of monetary policy shocks using time-varying parameter dynamic factor models. *Oxford Bulletin of Economics and Statistics*, 2012. Forthcoming.
- D. Korobilis. VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, 28(2):204–230, 2013.
- D. P. Kroese and J. C. C. Chan. *Statistical Modeling and Computation*. Springer, New York, 2014.

- D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods*. John Wiley & Sons, New York, 2011.
- E. M. Leeper, T. B. Walker, and S.-C. S. Yang. Fiscal foresight: Analytics and econometrics. NBER Working Paper 14028, 2008.
- H. Li and R. S. Tsay. A unified approach to identifying multivariate time series models. *Journal of the American Statistical Association*, 93(442):770–782, 1998.
- H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin, 2005.
- H. Lütkepohl and H. Claessen. Analysis of cointegrated VARMA processes. *Journal of Econometrics*, 80:223–239, 1997.
- H. Lütkepohl and D. S. Poskitt. Specification of echelon-form VARMA models. *Journal of Business & Economic Statistics*, 14(1):69–79, 1996.
- K. Metaxoglou and A. Smith. Maximum likelihood estimation of VARMA models using a state-space EM algorithm. *Journal of Time Series Analysis*, 28(5):666–685, 2007.
- R. Paap and H. van Dijk. Bayes estimates of Markov trends in possibly cointegrated series: An application to us consumption and income. *Journal of Business and Economic Statistics*, 21:547–563, 2003.
- M. S. Peiris. On the study of some functions of multivariate ARMA processes. *Journal of Time Series Analysis*, 25(1):146–151, 1988.
- D. S. Poskitt. Vector autoregressive moving average identification for macroeconomic modeling: A new methodology. Working Paper, Monash University, 2011.
- G. E. Primiceri. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852, 2005.
- N.i Ravishanker and B. K. Ray. Bayesian analysis of vector ARMA models using Gibbs sampling. *Journal of Forecasting*, 16(3):177–194, 1997.
- C. A. Sims. Macroeconomics and reality. *Econometrica*, 48:1–48, 1980.
- J. H. Stock and M. W. Watson. Why has U.S. inflation become harder to forecast? *Journal of Money Credit and Banking*, 39:3–33, 2007.
- S.-C. S. Yang. Quantifying tax effects under policy foresight. *Journal of Monetary Economics*, 52(8):1557 – 1568, 2005.