Stochastic Model Specification Search for Time-Varying Parameter VARs

Eric Eisenstat
Faculty of Business Administration,
University of Bucharest and RIMIR

Joshua C.C. Chan
Research School of Economics,
Australian National University

Rodney W. Strachan
Research School of Economics,
Australian National University

June 2013

Abstract

This article develops a new econometric methodology for performing stochastic model specification search (SMSS) in the vast model space of time-varying parameter VARs. This is motivated by the concern of over-fitting and the typically imprecise inference in these highly parameterized models. For each VAR coefficient, this new method automatically decides whether it is constant or time-varying. Moreover, it can be used to shrink an otherwise unrestricted time-varying parameter VAR to a stationary VAR, thus providing an easy way to (probabilistically) impose stationarity in time-varying parameter models. We demonstrate the effectiveness of the approach with a topical application, where we investigate the dynamic effects of structural shocks in government spending on U.S. taxes and GDP during a period of very low interest rates.

Keywords: Bayesian Lasso, shrinkage, fiscal policy

JEL Classification: C11, C52, E37, E47

1 Introduction

Vector autoregressions (VARs) are widely used for modeling and forecasting in macroeconomics. In particular, VARs have been used to understand the interactions between macroeconomic variables, often through the estimation of impulse response functions that characterize the effects of a variety of structural shocks on key economic variables. In recent years there has been much interest in extending the traditional constant coefficient VARs to time-varying parameter VARs (TVP-VARs) where the VAR coefficients are allowed to gradually evolve over time (see, among many others, Cogley and Sargent, 2005; Cogley, Primiceri, and Sargent, 2010; Koop, Leon-Gonzalez, and Strachan, 2011; Koop and Korobilis, 2013). This approach is motivated by a growing body of empirical evidence that demonstrates the importance of accommodating time-varying structures for model fitting and forecasting.

However, even for a moderate size VAR, the number of parameters in the model can be quite large relative to the number of observations. This can lead to imprecise estimation of impulse response functions and poor forecast performance. This problem is exacerbated in TVP-VARs, which are much higher dimensional than constant coefficient VARs. The potential problems associated with parameter proliferation have led many researchers to use Bayesian shrinkage in VARs to improve estimates precision and forecast performance (e.g., Banbura, Giannone, and Reichlin, 2010; Carriero, Clark, and Marcellino, 2011; Koop, 2011; Korobilis, 2011). Applying shrinkage to time-varying parameter models is less straightforward and it often requires computationally demanding algorithms (Chan, Koop, Leon-Gonzalez, and Strachan, 2012; Nakajima and West, 2013) or approximate inference (Koop and Korobilis, 2012). A related issue is model specification and model selection; although empirically a TVP-VAR that allows *all* VAR coefficients to change over time typically performs better than a constant coefficient VAR, it is plausible and even likely that a TVP-VAR where only *some* coefficients are time-varying while others are time-invariant will perform better than both alternatives. Reducing the number of time-varying parameters thereby achieves a degree of parsimony. However, it is unclear how one can decide a priori which coefficients are fixed and which are time-varying.

The main goal of this paper, following the interesting work of Frühwirth-Schnatter and Wagner (2010) and Belmonte, Koop, and Korobolis (2011) (hereafter FSW and BKK, respectively), among others, is to develop a new methodology to nest both time-varying and time-invariant VARs that is applicable to high-dimensional settings. Our starting

point is the stochastic model specification search (SMSS) framework introduced in FSW. Specifically, for each VAR coefficient, we introduce an indicator that chooses between a time-varying against a time-invariant parameter. This allows the model to automatically switch to a more parsimonious specification when the time-variation of the coefficient is "small". One computational challenge in this setup is that in a typical VAR, the number of values the indicators can take can be very large — e.g., for a small VAR with 3 variables and 4 lags, the number of combinations is $2^{39} \approx 5.5 \times 10^{11}$. To circumvent this computation problem, we introduce a new hierarchical prior for the indicators under which the stochastic model specification search can be performed efficiently. In addition, this new approach also incorporates the hierarchical Lasso prior in BKK that provides additional shrinkage, while maintaining the natural and intuitive framework in FSW. The proposed approach therefore adds to the growing literature on efficient methods of ensuring parsimony in potentially over-parameterized TVP-VARs.

Another advantage of the proposed approach is that it can be used to shrink the TVP-VAR to a stationary, constant coefficient VAR. In macroeconomic applications it is often necessary to impose stationarity conditions to avoid explosive impulse-response functions or forecasts. However, imposing stationarity conditions in a TVP-VAR implies inequality constraints on the time-varying coefficients, which results in a nonlinear state space model where conventional Kalman filter-based algorithms cannot be used. As pointed out in Koop and Potter (2011), the common way to impose stationarity in such settings — estimating the unconstrained TVP-VAR using Kalman filter-based algorithms and discarding any draws that do not satisfy the stationarity conditions — may lead to invalid inference. On the other hand, the single- and multi-move samplers proposed in Koop and Potter (2011), though sampling from the correct posterior distribution, can be computationally demanding. The proposed approach therefore provides a computationally feasible alternative for imposing the stationarity conditions probabilistically. By shrinking the TVP-VAR towards a stationary VAR, the model has the features of a stationary model, while still allowing some weight on nonstationarity to capture (poorly modeled) nonlinearities or extreme events.

In the second contribution of the paper we demonstrate the overall approach with a topical application. In particular, we investigate the dynamic responses of output growth and government revenues and expenditure to a government spending shock. We generalize earlier work by Blanchard and Perotti (2002) by allowing for general time variation in the model parameters. We study the dynamic effects of structural shocks in government spending on U.S. taxes and GDP during a period where the interest rate is close

to zero. In doing so we compare the inference obtained from the SMSS and the standard TVP-VAR specifications. We find that the SMSS generally offers more accurate estimates. In particular, our results from the SMSS show clearer evidence of an implication of Ricardian equivalence that the evolution of taxes generally follows that of spending. This phenomenon is much more difficult to detect with the standard TVP-VAR alone where the impulse responses are much less precisely estimated. In contrast, the SMSS leads to a distinct, clear pattern of long run government spending and taxes changing by the same amount during the period under examination. Overall, it appears that the SMSS is able to efficiently allocate time variation among the model parameters such as to strike a balance between parsimony and flexibility: most parameters are restricted to be essentially constant, while a select few are allowed to have a large degree of time variation.

The rest of the article is organized as follows. In Section 2 we first extend the stochastic model specification search approach of FSW to a multivariate setting. We then introduce a new hierarchical prior on the indicators and highlight its advantages over competing approaches. Section 3 outlines the posterior computation. Section 4 presents the empirical application that studies the effects of fiscal structural shocks on GDP growth. This section reports responses of macro variables to fiscal policy shocks and Section 5 contains some final comments.

2 SMSS and State Space Models

A popular approach for allowing for time-varying coefficients in time series models is through the state space specification. Specifically, suppose $\boldsymbol{y}_t$ is an $n \times 1$ vector of observations on the dependent variables, $\boldsymbol{X}_t$ is an $n \times m$ matrix of observations on explanatory variables and $\boldsymbol{\beta}_t$ is an $m \times 1$ vector of states. Then a generic state space model can be written as:

$$\boldsymbol{y}_t = \boldsymbol{X}_t \boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \tag{2}$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, and $\boldsymbol{\Omega}$ is typically assumed to be a diagonal matrix $\boldsymbol{\Omega} = \text{diag}(\omega_1^2, \ldots, \omega_m^2)$. The errors $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are assumed to be independent at all leads and lags. Finally, the state equation (2) is initialized with $\boldsymbol{\beta}_0 = \boldsymbol{\alpha}$, where

$\boldsymbol{\alpha} \sim \mathcal{N}\left(\boldsymbol{\alpha}_0, \boldsymbol{A}_0^{-1}\right)$. In a TVP-VAR setting, we have $\boldsymbol{X}_t = \mathbf{I}_n \otimes \boldsymbol{x}_t'$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\boldsymbol{x}_t$ is a $k \times 1$ vector of deterministic terms and lagged observations (hence $m = nk$).

This general state space framework encompasses a wide variety of commonly used time-varying parameter (TVP) regression models in macroeconomics and has become a standard framework for analyzing macroeconomic data. As discussed previously, recent research has raised the concern that over-fitting might be a problem for these highly parameterized models. Moreover, these high-dimensional models typically give imprecise estimates, making any form of inference more difficult. Motivated by these concerns, researchers might wish to have a more parsimonious specification that reduces the potential problem of over-parameterization, while maintaining the flexibility of the state space framework that allows time-variation in the coefficients. For example, one might wish to have a default model with time-invariant coefficients, but where each of these coefficients can switch to being time-varying when there is strong evidence for time-variation. In this way, one can maintain a parsimonious specification that leads to more precise estimates, while minimizing the risk of model misspecification.

In what follows, we first outline two existing methods for performing a stochastic model specification search with the aim of shrinking the model towards a more parsimonious specification. We then present a new method to undertake the specification search, which has theoretical and computational advantages over existing approaches.

2.1 Existing Approaches

In an important paper Frühwirth-Schnatter and Wagner (2010) (FSW) propose the following framework for nesting both time-varying and time-invariant coefficient specifications. To set the stage, consider again the generic state space model in (1)–(2), and recall that $\boldsymbol{\beta}_t = (\beta_{1,t}, \ldots, \beta_{m,t})'$ is the state vector at time t, $\boldsymbol{\beta}_0 = \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)'$ is the vector of initial states, and $\boldsymbol{\Omega} = \text{diag}(\omega_1^2, \ldots, \omega_m^2)$ is the covariance matrix for the state transition in (2). To proceed, reparameterize $\gamma_{j,t} = (\beta_{j,t} - \alpha_j)/\omega_j$ for $j = 1, \ldots, m$, so that (1)–(2) can be rewritten as

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\alpha} + \boldsymbol{X}_t\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t, \tag{3}$$

$$\boldsymbol{\gamma}_t = \boldsymbol{\gamma}_{t-1} + \widetilde{\boldsymbol{\eta}}_t, \tag{4}$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, $\widetilde{\boldsymbol{\eta}}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$, and $\boldsymbol{\Omega}^{\frac{1}{2}} = \text{diag}(\omega_1, \ldots, \omega_m)$. It is clear that when $\omega_1 = \cdots = \omega_m = 0$, the above specification reduces to a standard time-invariant regression model. By allowing the ω_j's to have different values, the model (3)–(4) can therefore accommodate the possibility that some coefficients are time-varying, whereas others are constant over time. More importantly, by specifying a suitable prior for the ω_j's, this framework provides a convenient way to shrink the TVP state space model to a more parsimonious constant coefficient regression.

Specifically, FSW propose the following independent normal priors each with a point mass at 0:

$$p(\omega_j) = \pi_{0j}\mathbf{1}(\omega_j = 0) + (1 - \pi_{0j})\phi(\omega_j; \mu_j, \tau_j^2), \tag{5}$$

where π_{0j} is the *a priori* probability that ω_j equals 0, and $\phi(\cdot; \mu, \sigma^2)$ denotes the normal density with mean μ and variance σ^2. These priors can be equivalently specified by introducing the indicators

$$d_j = \begin{cases} 0 & \text{with probability} \quad \pi_{0j}, \\ 1 & \text{with probability} \quad 1 - \pi_{0j}, \end{cases}$$

and $\widetilde{\omega}_j \sim \mathcal{N}(\mu_j, \tau_j^2)$. Then, $\omega_j = d_j\widetilde{\omega}_j$ has the desired distribution in (5).

Each vector $\boldsymbol{d} = (d_1, \ldots, d_m)'$ corresponds to a model in which some coefficients — specifically those with $d_j = 1$ — are time-varying while others are not. Hence, the stochastic variable selection approach of George and McCulloch (1993, 1997) can be adopted to perform the stochastic model specification search. Particularly, FSW derive a Markov Chain Monte Carlo (MCMC) algorithm for sampling the indicators simultaneously with the models parameters. However, the main drawback of this approach is that it becomes computationally infeasible when m is large. Specifically, in one step of the MCMC algorithm it is required to compute the joint probability that $d_1 = i_1, \ldots, d_m = i_m$ given the data and all the other parameters except the ω_j's, where $(i_1, \ldots, i_m) \in \{0, 1\}^m$. This step becomes computationally infeasible when m is large as there are altogether 2^m combinations.

In view of this difficulty, Belmonte, Koop, and Korobolis (2011) (BKK) specify an

alternative hierarchical *Lasso* prior directly on the ω_j's as follows:

$$(\omega_j \mid \tau_j^2) \sim \mathcal{N}\left(0, \tau_j^2\right),$$
$$\tau_j^2 \sim \mathcal{E}\left(\frac{\lambda^2}{2}\right), \tag{6}$$
$$\lambda^2 \sim \mathcal{G}\left(\lambda_{01}, \lambda_{02}\right), \tag{7}$$

where $\mathcal{E}$ and $\mathcal{G}$ denote respectively the Exponential and the Gamma distributions. This Lasso prior circumvents the model proliferation problem by removing the indicators, and at the same time provides additional shrinkage towards the time-invariant model.

While the specification of BKK addresses the computation issue, this comes at the cost of losing two attractive features of the FSW specification. First, while it is true that "if ω_j is *shrunk* to 0... then we have a model with a constant parameter on predictor j", the probability of such shrinking ever occurring is in fact zero, i.e., $\Pr(\omega_j = 0) = 0$. Second, the probability that the indicator $d_j = 0$ has an intuitive interpretation: it is the probability that the corresponding coefficient does not change over time. As such, this is useful for prior elicitation — e.g., some coefficients might be more likely to vary over time than others apriori. On the other hand, prior elicitation in BKK's specification is more difficult. Further, given the posterior draws of the indicators in FSW's specification, it is easy to compute posterior model probabilities for model selection or other probabilities to assess various hypotheses — e.g., output persistence does not change over time — whereas these probabilities are more difficult to compute under the specification of BKK.

2.2 A Tobit Prior

Both the FSW and BKK prior specifications have attractive features: the former has a natural, intuitive interpretation that is useful for prior elicitation and computing a variety of interesting posterior quantities, whereas the latter provides additional shrinkage and leads to a computationally feasible MCMC sampler even when the dimension of the states is large. We therefore aim to have a setup that has the advantages of both specifications, while circumventing their drawbacks. More specifically, we propose the following censored or Tobit prior on the ω_j's. For $j = 1, \ldots, m$, introduce the latent variable

$$\omega_j^* \sim \mathcal{N}\left(\mu_j, \tau_j^2\right),$$

and set

$$\omega_j = \begin{cases} 0 & \text{if} \quad \omega_j^* \leqslant 0 \\ \omega_j^* & \text{if} \quad \omega_j^* > 0. \end{cases}$$

It is easy to see that the marginal density of ω_j unconditional of ω_j^* is given by

$$p(\omega_j \mid \tau_j^2) = \Phi\left(-\frac{\mu_j}{\tau_j}\right)\mathbf{1}(\omega_j = 0) + \phi(\omega_j; \mu_j, \tau_j^2)\mathbf{1}(\omega_j > 0), \tag{8}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Additionally, we assume τ_j^2 has the prior as in (6)-(7) to incorporate the Lasso structure.

This proposed prior specification has several appealing features. First, it incorporates the essential elements of both the priors in FSW and BKK. To see the connection to FSW, let $\pi_{0j} = \Phi\left(-\frac{\mu_j}{\tau_j}\right)$ and rewrite the proposed prior in (8) as

$$p(\omega_j \mid \tau_j^2) = \pi_{0j}\mathbf{1}(\omega_j = 0) + (1 - \pi_{0j})\phi_{(0,\infty)}(\omega_j; \mu_j, \tau_j^2),$$

where $\phi_{\mathbb{X}}(\cdot; \mu, \sigma^2)$ denotes the density of the truncated $\mathcal{N}(\mu, \sigma^2)$ distribution with support in $\mathbb{X}$. Therefore, it allows for both a non-trivial probability of a certain parameter to be time-invariant and simultaneously the *hierarchical shrinking* that BKK so strongly argue for in forecasting applications.

Second, the proposed specification imposes a sensible relationship between the probability of time-invariance π_{0j} and the distribution of ω_j in the time-varying case. Specifically, the farther the truncated mean (normalized by standard deviation) is away from zero, the lower the probability that ω_j is zero, and vice-versa. This means that for values of ω_j that are close to zero, the prior elicits a strong preference for a time-invariant specification, while at the same time, penalizing situations where a high degree of time variation (i.e., a large ω_j) would be associated with a high probability of time invariance. In doing so, it imposes an important and desirable element of shrinkage on the model space. On the other hand, setting $\mu_j = 0$ implies that π_{0j} is constant and equal to 0.5. Consequently, tuning the hyper-parameters λ_{01}, λ_{02} such that τ_j^2 is "large" with probability ≈ 1 leads to the FSW specification where d_j is zero with probability 0.5 and $\widetilde{\omega}_j$ follows a half-normal, weakly-informative distribution.

Lastly, this prior leads to a straightforward Gibbs sampler. Importantly, in contrast to the MCMC sampler in FSW that becomes computationally intensive in high-dimensional state space models, this Gibbs sampler is fast and is applicable to such settings.

3 Priors and Posterior Computation

To complete the model specification, we assume the following standard independent priors for $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$:

$$\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\alpha}_0, \boldsymbol{A}_0^{-1}), \quad \boldsymbol{\Sigma} \sim \mathcal{IW}(\nu_0, \boldsymbol{\Sigma}_0),$$

where $\mathcal{IW}$ denotes the inverse Wishart distribution. In what follows, we outline the posterior computation using MCMC, and we give the details in Appendix B. First, define the $m \times 1$ vectors $\boldsymbol{\omega}^* = (\omega_1^*, \ldots, \omega_m^*)'$, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)'$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)'$, and stack

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_T \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_T \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \vdots \\ \boldsymbol{\gamma}_T \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_T \end{pmatrix}.$$

Then, posterior draws are obtained by sequentially sampling from:

1. $p(\boldsymbol{\alpha} \,|\, \boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\omega}^*, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \lambda)$;
2. $p(\boldsymbol{\gamma} \,|\, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\omega}^*, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \lambda)$;
3. $p(\boldsymbol{\Sigma} \,|\, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\omega}^*, \boldsymbol{\tau}, \lambda)$;
4. $p(\boldsymbol{\omega}^* \,|\, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \lambda)$;
5. $p(\boldsymbol{\tau} \,|\, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\omega}^*, \boldsymbol{\Sigma}, \lambda)$;
6. $p(\lambda \,|\, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\omega}^*, \boldsymbol{\Sigma}, \boldsymbol{\tau})$.

We note that each ω_j is completely determined by ω_j^*. In fact, under the Tobit prior discussed in Section 2.2, we have $\omega_j = 0$ if $\omega_j^* \leq 0$ and $\omega_j = \omega_j^*$ otherwise. Now, given $\boldsymbol{\omega}$, the model in (3)–(4) is a conventional TVP-VAR. Hence, Steps 1-3 are standard, and we leave the details to Appendix B. Here we discuss the implementations of Steps 4-6.

One feasible approach to sample $\boldsymbol{\omega}^*$ from its full conditional distribution is to simulate each element ω_j^* at a time. To that end, first recall that $\boldsymbol{X}_t = \mathbf{I}_n \otimes \boldsymbol{x}_t'$, and $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{k,t})'$ is a $k \times 1$ vector of deterministic terms and lagged observations. Now,

defining $\boldsymbol{G}_t = \boldsymbol{X}_t \operatorname{diag}(\boldsymbol{\gamma}_t)$, let

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{G}_1 \\ \vdots \\ \boldsymbol{G}_T \end{pmatrix}.$$

Stack (3) over t and rewrite the measurement equation as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{G}\boldsymbol{\omega} + \boldsymbol{\varepsilon}. \tag{9}$$

Further, let $\boldsymbol{g}_j$ denote the j-th column of $\boldsymbol{G}$, and accordingly, let $\boldsymbol{G}_{\setminus j}$ be the $Tn \times (m-1)$ matrix obtained by deleting the j-th column in $\boldsymbol{G}$. Similarly, let $\boldsymbol{\omega}_{\setminus j}$ represent $\boldsymbol{\omega}$ with the j-th row removed. Then,

$$\boldsymbol{v}_j \equiv \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{G}_{\setminus j}\boldsymbol{\omega}_{\setminus j} = \boldsymbol{g}_j\omega_j + \boldsymbol{\varepsilon}.$$

Compute further the following posterior quantities:

$$\begin{aligned}
\widehat{\tau}_j^2 &= \left(\tau_j^{-2} + \boldsymbol{g}_j'(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{g}_j\right)^{-1} \\
\widehat{\mu}_j &= \widehat{\tau}_j^2 \left(\mu_j/\tau_j^2 + \boldsymbol{g}_j'(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{v}_j\right) \\
\widehat{\psi}_j &= \frac{\Phi\left(\widehat{\mu}_j/\widehat{\tau}_j\right)}{\Phi\left(-\mu_j/\tau_j\right)} \frac{\widehat{\tau}_j}{\tau_j} \exp\left\{\frac{1}{2}\left(\frac{\widehat{\mu}_j^2}{\widehat{\tau}_j^2} - \frac{\mu_j^2}{\tau_j^2}\right)\right\} \\
\widehat{\pi}_j &= (1 + \widehat{\psi}_j)^{-1}.
\end{aligned}$$

Then, the conditional density of ω_j^* is the following 2-component mixture of truncated normals

$$p(\omega_j^* \mid \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\omega}_{\setminus j}^*, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \lambda) = \widehat{\pi}_j \phi_{(-\infty,0)}\left(\omega_j^* \mid \mu_j, \tau_j^2\right) + (1 - \widehat{\pi}_j)\phi_{(0,\infty)}\left(\omega_j^* \mid \widehat{\mu}_j, \widehat{\tau}_j^2\right).$$

The derivations can be found in Appendix B. A draw from the above mixture distribution can be obtained as follows. First, get a Bernoulli draw Z with $\mathbb{P}(Z = 0) = \widehat{\pi}_j$. If $Z = 0$, sample ω_j^* from the truncated $\mathcal{N}(\omega_j^*; \mu_j, \tau_j^2)$ distribution with support $(-\infty, 0)$; if $Z = 1$, sample ω_j^* instead from the truncated $\mathcal{N}(\omega_j^*; \widehat{\mu}_j, \widehat{\tau}_j^2)$ distribution with support $(0, \infty)$.

Finally, τ_j^2 and λ conditional on ω_j^* may be sampled exactly as discussed in BKK:

$$(\tau_j^{-2} \mid \lambda, \omega_j^*) \sim \mathcal{IG}\left(\sqrt{\frac{\lambda^2}{(\omega_j^* - \mu_j)^2}}, \lambda^2\right),$$
$$(\lambda^2 \mid \boldsymbol{\tau}) \sim \mathcal{G}\left(\lambda_{01} + m, \lambda_{02} + \frac{1}{2}\sum_{j=1}^{m} \tau_j^2\right),$$

where $\mathcal{IG}$ denotes the inverse Gaussian distribution.

4 Application

In this section we investigate the effects of a fiscal shock on output, taxes and government spending. We estimate the impulse responses of these variables to a shock to government spending. Perotti (2005) and Galí, Vallés and López-Salido (2007) point out that models based upon different theories — neoclassical, Keynesian or neo-Keynesian — can make conflicting predictions about the responses of macro variables to fiscal shocks. When the empirical evidence is imprecise little can be learned about the support for alternative theories and therefore the likely effects of fiscal policy. The results from this section provide improved inference by producing more precise estimates of impulse responses of output, government spending and receipts to a spending shock. We estimate a VAR for the vector of US variables $\boldsymbol{w}_t = (t_t, g_t, y_t)'$, where t_t is a measure of government revenue, g_t is government expenditures, and y_t is output. All variables are in logs of real per capita values. Government expenditure consists of government consumption and investment while government revenue is less government transfers. Following Blanchard and Perotti (2002), we use real per capita figures and deflate the variables using the GDP deflator.

As all of these variables display strong trending behaviors, and any test for a unit root will support this conclusion, the temptation is to model them in differences. However differencing removes information on the relationships among the levels of the variables. It is reasonable to conclude that government expenditure should not exceed GDP nor be negative. We can say something more than this as, in the US at least, government expenditure tends to remain fairly stable as a proportion of GDP. There is also reason to believe that government expenditures cannot wander too far from government receipts for too long. We therefore respecify the VAR into a VECM in which we impose station-

arity of the difference $t_t - g_t$ and $g_t - y_t$. Imposing the first of these relations expresses a prior belief in Ricardian equivalence. This seems to be a reasonable restriction as evidence on these constraints suggests they are $I(0)$, although the exact form of the process may have structural breaks (see discussion in, for example, Martin, 2009).

Setting $\boldsymbol{z}_t = (t_t - g_t, g_t - y_t)'$, and assuming four autoregressive lags (following Blanchard and Perotti, 2002), the VECM has the form

$$\Delta \boldsymbol{w}_t = \boldsymbol{\beta}_{0,t} + \boldsymbol{\beta}_{1,t}\boldsymbol{z}_{t-1} + \sum_{l=1}^{4} \boldsymbol{\beta}_{1+l,t}\Delta \boldsymbol{w}_{t-l} + \boldsymbol{\varepsilon}_t, \tag{10}$$

where $\boldsymbol{\beta}_{0,t}$ is an $n \times 1$ vector of intercepts, $\boldsymbol{\beta}_{1,t}$ an $n \times 2$ matrix of coefficients, and $\boldsymbol{\beta}_{2,t}, \ldots, \boldsymbol{\beta}_{5,t}$ are $n \times 3$ matrices of coefficients. In terms of the model in (1), we have

$$\boldsymbol{y}_t = \Delta \boldsymbol{w}_t = \begin{pmatrix} \Delta t_t \\ \Delta g_t \\ \Delta y_t \end{pmatrix}, \qquad \boldsymbol{x}_t = \begin{pmatrix} 1 \\ \boldsymbol{z}_{t-1} \\ \Delta \boldsymbol{w}_{t-1} \\ \vdots \\ \Delta \boldsymbol{w}_{t-4} \end{pmatrix}, \qquad \boldsymbol{\beta}_t = \text{vec}\left((\boldsymbol{\beta}_{0,t}, \cdots, \boldsymbol{\beta}_{5,t})'\right).$$

This setup is further motivated by the well documented importance of allowing parameters to vary over time when analyzing macroeconomic data (see for example the survey in Koop and Korobilis, 2010). Indeed, Blanchard and Perotti (2002) allow for limited time-variation in the transmission mechanism through various trending specifications along with a quarterly dependence of the parameters. Nevertheless, while they acknowledge that allowing for more general time variation would be appropriate, they conclude that doing so "would have quickly exhausted all degrees of freedom."

Accordingly, we utilize this framework to compare the inference obtained from the SMSS specification proposed in this article to the benchmark TVP-VAR. In particular, we seek to ascertain whether the Tobit prior can provide the necessary degree of parsimony such as to yield sufficiently precise inference when the parameter-rich TVP-VAR fails to do so, while allowing for certain parameters to vary over time with some (nonzero) probability. To this end, we complete the SMSS specification of the VECM outlined above by setting

the following hyper-parameters on the priors discussed in Section 3:

$$\begin{aligned} \boldsymbol{\alpha}_0 &= \mathbf{0}, & \boldsymbol{A}_0 &= \frac{1}{100}\mathbf{I}_m, \\ \nu_0 &= n+3, & \boldsymbol{\Sigma}_0 &= \mathbf{I}_n, \\ \lambda_{01} &= 1+\frac{1}{100}, & \lambda_{02} &= \frac{1}{100}. \end{aligned}$$

To estimate the standard TVP-VAR, we employ the algorithm given in Chan and Jeliazkov (2009), with the priors chosen to match as closely as possible the SMSS prior specification above. Finally, our data essentially coincide with that used by Blanchard and Perotti (2002) but are extended to cover the sample period 1958Q3 to 2011Q4.

4.1 Identification Restrictions

Eliciting impulse response functions necessitates the estimation of *structural* shocks. Blanchard and Perotti (2002) provide the restrictions needed to identify the structural shocks in the VAR framework. It turns out that under these restrictions, structural parameters can be readily recovered from the reduced form covariance matrix. Consequently, we follow this strategy in simulating the posterior of the VECM given in (10) directly, then recovering draws of the necessary structural parameters by solving a system of nonlinear equations for each draw. We briefly discuss the utilized approach in this section, with the details provided in Appendix A.

Specifically, equations (2)-(4) in Blanchard and Perotti (2002) may be represented as

$$\begin{pmatrix} 1 & 0 & -a_1 \\ 0 & 1 & -b_1 \\ -c_1 & -c_2 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_t^t \\ \varepsilon_t^g \\ \varepsilon_t^y \end{pmatrix} = \begin{pmatrix} 1 & a_2 & 0 \\ b_2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{pmatrix} \begin{pmatrix} u_t^t \\ u_t^g \\ u_t^y \end{pmatrix}, \tag{11}$$

where $u_t^j \sim \mathcal{N}(0,1)$ for $j = t, g, y$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_t^t, \varepsilon_t^g, \varepsilon_t^y)' \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is the reduced form residual in (10). Given a draw of $\boldsymbol{\Sigma}$, therefore, it is possible to recover draws of the parameters in (11) by solving an appropriate system of equations. There are two key components in the identification strategy set forth by Blanchard and Perotti (2002) that make solving this system relatively straightforward:

1. a_1 and b_1 are given;

2. either (i) $a_2 = 0, b_2 \neq 0$, or (ii) $a_2 \neq 0, b_2 = 0$.

Blanchard and Perotti (2002) set by assumption $b_1 = 0$ and compute a value for $a_1 = 2.08$ (as the average *within-quarter elasticity of net taxes with respect to output*). Because we consider our data to be sufficiently similar, and since our central concern is to compare the precision in inferences obtained from the SMSS and TVP-VAR specifications, these values are retained for the purpose of the impulse response analysis reported below. Alternatively, one could specify non-degenerate distributions over a_1 and b_1, to reflect less dogmatic prior information. Combining samples of a_1, b_1 from such distributions with posterior draws of the model parameters (in simulating impulse response functions) would essentially lead to an exercise in averaging over prior beliefs.

4.2 Results

We implement the Gibbs sampler described in Section 3 to obtain 25,000 posterior draws (after a 2,500 burn-in) of the parameters in the VECM model (10). In doing so, we re-scale and re-center the three "first-differenced" series (i.e., $\Delta \boldsymbol{w}_t$) to match the naive priors. The two integration relations in $\boldsymbol{z}_t$ are left untouched. Accordingly, all impulse response functions reported below are appropriately adjusted such as to reverse these transformations and match the original series.[1] Our implementation of the algorithm also utilizes the three Generalized Gibbs steps (Liu and Sabatti, 2000) discussed in Appendix B. The stability of the algorithm was verified by tracking the inefficiency factors of all sampled variables and replicating the simulation run several times.

To demonstrate the effect of the Tobit prior in controlling the time variation of the parameters, we report two types of posterior quantities: (i) the *time-invariance probability* (TIP), which represents the estimated probability that a particular $\beta_{j,t}$ is constant — e.g., $\Pr(\omega_{j,t} = 0 \,|\, \boldsymbol{y})$ — and (ii) the *maximum time variation* (MTV) in posterior means, which is computed as the difference $\max\{\mathrm{E}(\beta_{j,t} \,|\, \boldsymbol{y})\}_{t=1}^{T} - \min\{\mathrm{E}(\beta_{j,t} \,|\, \boldsymbol{y})\}_{t=1}^{T}$. As illustrated in Table 1, estimated TIPs under the SMSS specification vary from 0.192 (effect of first lag in GDP on spending) to 0.769 (effect of first lag in spending on taxes). Consequently, both the magnitude of and variation in the probabilities of time invari-

[1] A similar approach to "standardizing" the series is undertaken by BKK and affects only the prior specification, as long as the transformations are appropriately accounted for in posterior computations. Alternatively, one may work with the original series and employ a *training sample* to specify the priors (for example, as in Primiceri, 2005), although this is operationally more involved. It is worthwhile noting that we apply the same standardization in both the SMSS and the TVP-VAR specifications.

Table 1: Estimated Time-Invariance Probabilities under the SMSS Specification

		Tax	Spend	GDP
	Intercept	0.592	0.332	0.528
	$t_{t-1} - g_{t-1}$	0.523	0.503	0.491
	$g_{t-1} - y_{t-1}$	0.542	0.235	0.506
1st lag	Tax	0.332	0.559	0.565
	Spend	0.769	0.690	0.699
	GDP	0.339	0.192	0.676
2nd lag	Tax	0.717	0.734	0.606
	Spend	0.537	0.668	0.685
	GDP	0.553	0.565	0.695
3rd lag	Tax	0.405	0.721	0.697
	Spend	0.727	0.715	0.669
	GDP	0.643	0.719	0.752
4th lag	Tax	0.730	0.513	0.738
	Spend	0.578	0.613	0.616
	GDP	0.657	0.630	0.740

ance give strong support to this approach over the highly parameterized TVP-VAR. This is further reinforced by comparing the MTVs reported in Table 2, across the two specifications. Note that these in general follow the same pattern for both SMSS and TVP-VAR in the sense that parameter estimates that exhibit relatively more time variation under the TVP-VAR also vary relatively more under the SMSS. Nevertheless, the SMSS specification evidently leads to an overall reduction in time variation across the parameters.

There is likewise a strong link between the TIPs and MTVs (under SMSS): as expected, higher TIPs generally correspond to lower MTVs. More interestingly, higher TIPs appear to be associated with more substantial reductions in MTVs. However, at the lower end of the TIPs, time variation actually increases. For example, the TIP related to the effect of the co-integration term $g_{t-1} - y_{t-1}$ on spending is estimated at 0.235, and its posterior mean varies from -6.47 to -5.33 (a difference of 1.14) while under the TVP-VAR the maximum variation is less than half that at 0.503.

The above results suggest that the Tobit prior does not only produce the effect of reducing "unnecessary" time variation, but it also reallocates time variation across the model parameters by reducing such variation in some parameters, while increasing it in others. As made evident in Tables 1 and 2, the SMSS specification (relative to the standard TVP-VAR) tends to remove time variation from a majority of the model

Table 2: Comparison of Maximum Time Variation in Posterior Means under the SMSS and the TVP-VAR Specifications

		SMSS			TVP-VAR		
		Tax	Spend	GDP	Tax	Spend	GDP
	Intercept	0.232	0.780	0.176	0.372	0.478	0.270
	$t_{t-1} - g_{t-1}$	0.064	0.062	0.241	0.157	0.077	0.215
	$g_{t-1} - y_{t-1}$	0.254	1.140	0.261	0.390	0.503	0.325
1st lag	Tax	0.304	0.130	0.113	0.440	0.311	0.367
	Spend	0.016	0.040	0.030	0.365	0.226	0.202
	GDP	0.428	0.512	0.049	0.546	0.575	0.157
2nd lag	Tax	0.024	0.014	0.085	0.141	0.159	0.441
	Spend	0.102	0.048	0.036	0.418	0.238	0.190
	GDP	0.145	0.118	0.026	0.484	0.312	0.174
3rd lag	Tax	0.172	0.028	0.040	0.590	0.206	0.451
	Spend	0.028	0.034	0.048	0.243	0.228	0.305
	GDP	0.072	0.015	0.012	0.670	0.129	0.367
4th lag	Tax	0.022	0.153	0.022	0.232	0.433	0.202
	Spend	0.088	0.069	0.083	0.297	0.276	0.345
	GDP	0.043	0.086	0.022	0.277	0.298	0.209

parameters and concentrate it on a select few. Indeed, this highlights the mechanism by which the Tobit prior induces parsimony in a time-varying parameter model.

To further illustrate the effectiveness of the SMSS approach, we compute impulse responses to spending shocks for taxes, spending and GDP, in the spirit of Blanchard and Perotti (2002). In particular, we employ the identification restrictions discussed in Section 4.1, along with the procedure outlined in Appendix A, to decompose the reduced form covariance matrix $\boldsymbol{\Sigma}$. We then use the the resulting posterior draws of the structural parameters (together with draws of $\boldsymbol{\beta}$) to compute the percent changes in taxes, spending and GDP to a one percent increase in spending. Therefore, in contrast to Blanchard and Perotti (2002), the impulse responses are expressed in terms of elasticities rather than absolute levels. Moreover, because we are working in a time-varying parameter context, we choose a specific (within-sample) time period — 2000Q1 to 2010Q4 (10 years) — to conduct the analysis.

The resulting impulse response functions are summarized in Figure 1, where the the HPD intervals (shaded regions) are constructed as [16%, 84%] while the *estimated* impulse responses (solid lines) are posterior medians. In the figure, the top row contains the impulse responses generated by the SMSS specification, while in the bottom row are those corresponding to the standard TVP-VAR. It is clear that that the TVP-VAR

generally yields wider HPD intervals relative to SMSS. This is particularly evident for taxes and GDP (where incidentally SMSS tends to assign higher probabilities of time-invariance). The more accurate inference delivered by SMSS is important since, as pointed out by Perotti (2005), the effects of fiscal policy have become weaker over time and so it is important to have a method that can distinguish when the effect is different from zero.

The TVP-VAR yields somewhat peculiar (median) impulse responses for taxes and GDP, in contrast to SMSS. For example, the TVP-VAR predicts that both taxes and GDP responses are initially negative and require about two years to cross into positive territory, where they remain in the long run. This is counter-intuitive and appears to be an artifact of the imprecision with which the TVP-VAR estimates the impulse responses. The reported HPD intervals contain zero for all of this initial period. Under the SMSS specification, on the other hand, the responses of taxes and GDP are estimated to be positive for the entire period in examination — a much more plausible result and one consistent with the conclusions drawn in Blanchard and Perotti (2002).

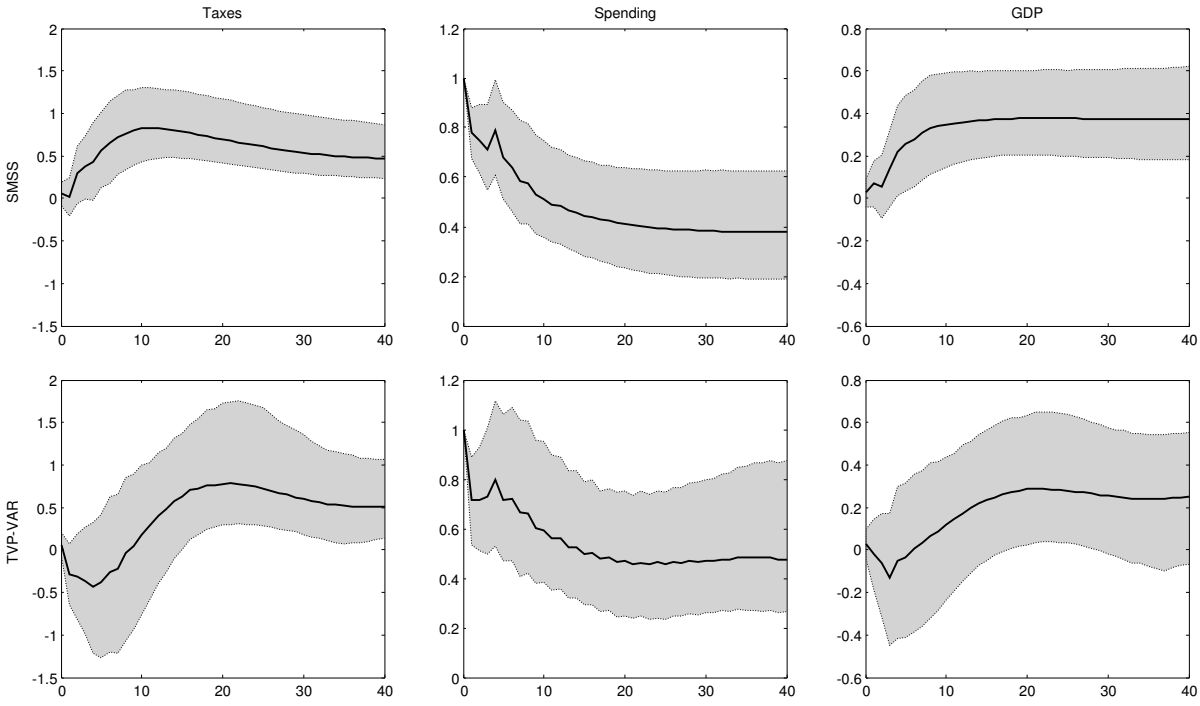

Figure 1: Comparison of IRFs generated by the SMSS and TVP-VAR for 40 quarters following a 1% Shock in Spending in 2000Q1

Another interesting feature of the results is the clear effect of imposing of Ricardian equivalence in the responses. This feature is imposed upon both models and the long

run effect is obviously present in both models, but the SMSS gives much tighter bounds. To highlight the comparison, Figure 2 plots the median responses for taxes and spending (without the HPD limits), as generated by the SMSS and standard TVP-VAR specifications. Indeed, it seems that taxes are always chasing spending with a delayed response, until spending stabilizes and they converge. This is very clear in the SMSS, but noticeably more distorted with the TVP-VAR, where taxes dip into the negative over the first two years. Based on the TVP-VAR results alone, therefore, distinguishing this effect is made difficult by the imprecision of the estimates related to the parameter proliferation present in this specification. In contrast, employing the Tobit prior leads to a much clearer representation and greatly facilitates economic interpretation of the simulation results.

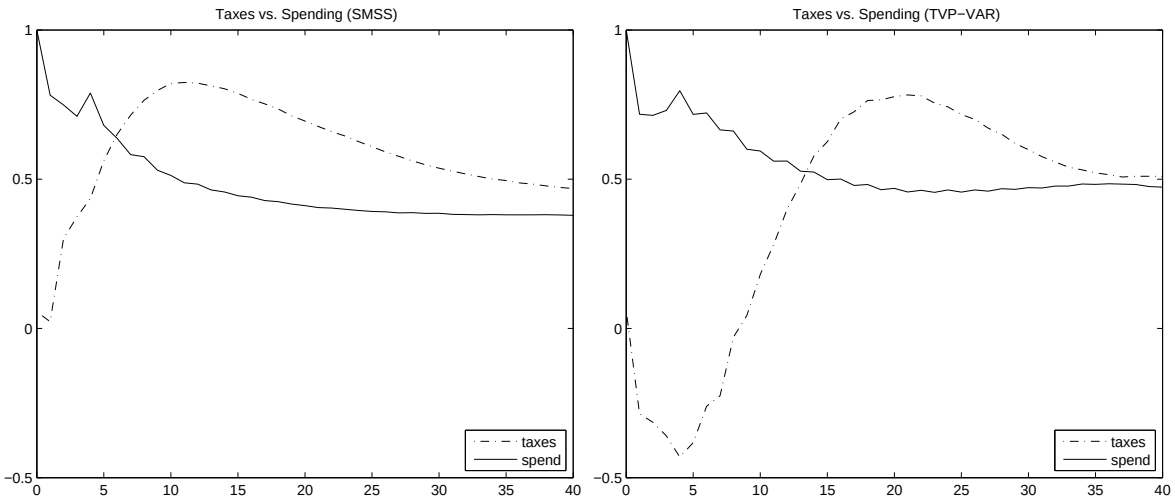

Figure 2: Median Impulse Response of Taxes and Spending to a 1% Shock in Spending in 2000Q1

5 Concluding Remarks

Time-varying VARs are widely used for studying the dynamic effects of structural shocks on key economic variables through the estimation of impulse response functions. However, since these models are highly parameterized, inference is typically imprecise and conclusions are often difficult to draw. In this paper we present a new method that allows the data to decide whether parameters are time varying or time invariant in a VAR, thus allowing the model to automatically switch to a more parsimonious specification

when the time-variation of the coefficient is small. By introducing a Tobit prior on the variances in the state equations, the task of computing the many indicators is greatly simplified.

We apply the new methodology and the computation scheme to a brief study of responses in US government receipts, government expenditure and GDP to a fiscal policy shock. Compared to those results obtained under an unrestricted TVP-VAR, we find differences in the median response paths but the most significant improvement is in the precision of estimation of the responses. This is an important result since these effects have become weaker over time (Perotti, 2005) so greater precision in estimation is necessary for accurate inference.

Appendix A: Recovering Structural Parameters from the Precision Matrix

In terms of (11), define $\boldsymbol{\psi} = -\begin{pmatrix} a_1 \\ b_1 \end{pmatrix}$, $\boldsymbol{\gamma} = -\begin{pmatrix} c_1 & c_2 \end{pmatrix}$, $\boldsymbol{\Delta} = \begin{pmatrix} 1 & a_2 \\ b_2 & 1 \end{pmatrix}$, $\boldsymbol{\Xi}^{\frac{1}{2}} = \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix}$ and rewrite the system as

$$\begin{pmatrix} \mathbf{I} & \boldsymbol{\psi} \\ \boldsymbol{\gamma} & 1 \end{pmatrix} \boldsymbol{\varepsilon}_t = \begin{pmatrix} \boldsymbol{\Delta} & \\ & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Xi}^{\frac{1}{2}} & \\ & s_3 \end{pmatrix} \boldsymbol{u}_t, \quad \boldsymbol{u}_t \overset{iid}{\sim} \mathcal{N}(0, \mathbf{I}_3).$$

Hence,

$$\boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{N}\left(0, (\boldsymbol{K}'\boldsymbol{K})^{-1}\right),$$

where

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{\Xi}^{-\frac{1}{2}} & \\ & \frac{1}{s_3} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Delta}^{-1} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{I} & \boldsymbol{\psi} \\ \boldsymbol{\gamma} & 1 \end{pmatrix}.$$

If we estimate a reduced-form covariance matrix $\boldsymbol{\Sigma}$, therefore, posterior draws of the structural parameters and variances (e.g. the six free parameters in $\boldsymbol{K}$) can be obtained by solving $\boldsymbol{K}'\boldsymbol{K} = \boldsymbol{Q} \equiv \boldsymbol{\Sigma}^{-1}$. More precisely, partitioning

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{q}_{12} \\ \boldsymbol{q}_{12}' & q_3 \end{pmatrix},$$

and denoting $\boldsymbol{\Theta} = (\boldsymbol{\Delta}\boldsymbol{\Xi}\boldsymbol{\Delta}')^{-1}$ leads to the system of nonlinear equations:

$$\boldsymbol{\Theta} + \frac{1}{s_3^2}\boldsymbol{\gamma}'\boldsymbol{\gamma} = \boldsymbol{Q}_{11} \tag{12}$$

$$\boldsymbol{\Theta}\boldsymbol{\psi} + \frac{1}{s_3^2}\boldsymbol{\gamma}' = \boldsymbol{q}_{12} \tag{13}$$

$$\boldsymbol{\psi}'\boldsymbol{\Theta}\boldsymbol{\psi} + \frac{1}{s_3^2} = q_3 \tag{14}$$

As discussed in the text, the following two conditions make this system easy to solve:

1. a_1 and b_1 are given (more specifically, $b_1 = 0$), and therefore, $\boldsymbol{\psi}$ is known
2. either (i) $a_2 = 0, b_2 \neq 0$, which implies that $\boldsymbol{\Delta}$ is *lower* triangular with 1's on the diagonal, or (ii) $a_2 \neq 0, b_2 = 0$, which implies that $\boldsymbol{\Delta}$ is *upper* triangular with 1's on the diagonal.

Given this, $\boldsymbol{\gamma}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Xi}^{\frac{1}{2}}$, and s_3 can be solved for in three steps:

1. Solve for s_3:

$$\boldsymbol{\psi}'\boldsymbol{\Theta}\boldsymbol{\psi} + \frac{1}{s_3^2}\boldsymbol{\psi}'\boldsymbol{\gamma}'\boldsymbol{\gamma}\boldsymbol{\psi} = \boldsymbol{\psi}'\boldsymbol{Q}_{11}\boldsymbol{\psi}$$
$$\boldsymbol{\psi}'\boldsymbol{\Theta}\boldsymbol{\psi} + \frac{1}{s_3^2}\boldsymbol{\psi}'\boldsymbol{\gamma}' = \boldsymbol{\psi}'\boldsymbol{q}_{12}$$
$$\boldsymbol{\psi}'\boldsymbol{\Theta}\boldsymbol{\psi} + \frac{1}{s_3^2} = q_3,$$

   implies that

$$q_3 - \frac{1}{s_3^2} + \frac{1}{s_3^2}\left[s_3^2\left(\boldsymbol{\psi}'\boldsymbol{q}_{12} - q_3\right) - 1\right]^2 = \boldsymbol{\psi}'\boldsymbol{Q}_{11}\boldsymbol{\psi},$$

   and therefore,

$$s_3 = \frac{\sqrt{\boldsymbol{\psi}'\boldsymbol{Q}_{11}\boldsymbol{\psi} - 2\boldsymbol{\psi}'\boldsymbol{q}_{12} + q_3}}{\boldsymbol{\psi}'\boldsymbol{q}_{12} - q_3}. \tag{15}$$

2. Solve for $\boldsymbol{\gamma}$, given s_3:

   Substituting (12) into (13) leads to

$$\boldsymbol{\gamma}'\boldsymbol{\gamma}\boldsymbol{\psi} - \boldsymbol{\gamma}' + s_3^2\left(\boldsymbol{q}_{12} - \boldsymbol{Q}_{11}\boldsymbol{\psi}\right) = 0. \tag{16}$$

   It can be readily verified that the solution to (16) is of the form

$$\boldsymbol{\gamma}' = \frac{1 \pm \sqrt{1 - 4s_3^2\boldsymbol{\psi}'\left(\boldsymbol{q}_{12} - \boldsymbol{Q}_{11}\boldsymbol{\psi}\right)}}{2\boldsymbol{\psi}'\left(\boldsymbol{q}_{12} - \boldsymbol{Q}_{11}\boldsymbol{\psi}\right)}\left(\boldsymbol{q}_{12} - \boldsymbol{Q}_{11}\boldsymbol{\psi}\right). \tag{17}$$

3. Solve for s_1, s_2 and either a_2 or b_2, given s_3, $\boldsymbol{\gamma}$:

   From (12), we have

$$\boldsymbol{\Delta}\boldsymbol{\Xi}\boldsymbol{\Delta}' = \left(\boldsymbol{Q}_{11} - \frac{1}{s_3^2}\boldsymbol{\gamma}'\boldsymbol{\gamma}\right)^{-1}. \tag{18}$$

   Therefore,

   (a) $a_2 = 0, b_2 \neq 0$: s_1, s_2, b_2 are found by the LDL decomposition of the 2×2 matrix $\left(\boldsymbol{Q}_{11} - \frac{1}{s_3^2}\boldsymbol{\gamma}'\boldsymbol{\gamma}\right)^{-1}$.

   (b) $a_2 \neq 0, b_2 = 0$: s_1, s_2, a_2 are found by the UDL decomposition of the 2×2 matrix $\left(\boldsymbol{Q}_{11} - \frac{1}{s_3^2}\boldsymbol{\gamma}'\boldsymbol{\gamma}\right)^{-1}$.

Of course, one caveat here is that the $\pm$ in (17) generates two possible values for $\boldsymbol{\gamma}$. Indeed, there are two solutions to (16). However, only one of these solutions leads to a correct decomposition of $\left(\boldsymbol{Q}_{11} - \frac{1}{s_3^2}\boldsymbol{\gamma}'\boldsymbol{\gamma}\right)^{-1}$ in Step 3. To solve this problem, we write a script that computes both solutions from (17) and checks which one leads to a correct solution of the entire system.

Appendix B: MCMC Sampler

In this appendix we provide the details of the MCMC sampler outlined in Section 3. First, we derive the expression for $\widehat{\psi}_j$. Note that the joint conditional density for (ω_j^*, ω_j) is given by

$$\begin{aligned} p\left(\omega_j^*, \omega_j \,|\, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\omega}_{\setminus j}^*, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \lambda\right) \propto\ & \phi\left(\boldsymbol{v}; \boldsymbol{g}_j\boldsymbol{\omega}_j, \mathbf{I}_T \otimes \boldsymbol{\Sigma}\right) \phi\left(\omega_j^*; \mu_j, \tau_j^2\right) \mathbf{1}(\omega_j^* \leq 0)\mathbf{1}(\omega_j = 0) \\ & + \phi\left(\boldsymbol{v}; \boldsymbol{g}_j\boldsymbol{\omega}_j, \mathbf{I}_T \otimes \boldsymbol{\Sigma}\right) \phi\left(\omega_j^*; \mu_j, \tau_j^2\right) \mathbf{1}(\omega_j^* > 0)\mathbf{1}(\omega_j = \omega_j^*). \end{aligned}$$

Hence, marginalizing over ω_j yields

$$\begin{aligned} p\left(\omega_j^* \,|\, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\omega}_{\setminus j}^*, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \lambda\right) &\propto \phi\left(\boldsymbol{v}_j; \mathbf{0}, \mathbf{I}_T \otimes \boldsymbol{\Sigma}\right) \phi\left(\omega_j^*; \mu_j, \tau_j^2\right) \mathbf{1}(\omega_j^* \leq 0) \\ &\quad + \phi\left(\boldsymbol{v}_j; \boldsymbol{g}_j\boldsymbol{\omega}_j^*, \mathbf{I}_T \otimes \boldsymbol{\Sigma}\right) \phi\left(\omega_j^*; \mu_j, \tau_j^2\right) \mathbf{1}(\omega_j^* > 0) \\ &\quad \propto \Phi\left(-\frac{\mu_j}{\tau_j}\right) \phi\left(\boldsymbol{v}_j; \mathbf{0}, \mathbf{I}_T \otimes \boldsymbol{\Sigma}\right) \phi_{(-\infty,0)}\left(\omega_j^*; \mu_j, \tau_j^2\right) \\ &\quad + \Phi\left(\frac{\widehat{\mu}_j}{\widehat{\tau}_j}\right) \frac{\phi\left(\boldsymbol{v}_j; \boldsymbol{g}_j\boldsymbol{\omega}_j^*, \mathbf{I}_T \otimes \boldsymbol{\Sigma}\right) \phi\left(\omega_j^*; \mu_j, \tau_j^2\right)}{\phi\left(\omega_j^*; \widehat{\mu}_j, \widehat{\tau}_j^2\right)} \phi_{(0,\infty)}\left(\omega_j^*; \widehat{\mu}_j, \widehat{\tau}_j^2\right) \\ &\quad \propto \phi_{(-\infty,0)}\left(\omega_j^*; \mu_j, \tau_j^2\right) + \widehat{\psi}_j \phi_{(0,\infty)}\left(\omega_j^*; \widehat{\mu}_j, \widehat{\tau}_j^2\right), \end{aligned}$$

where

$$\begin{aligned} \widehat{\psi}_j &= \frac{\Phi\left(\widehat{\mu}_j/\widehat{\tau}_j\right)}{\Phi\left(-\mu_j/\tau_j\right)} \times \frac{\phi\left(\boldsymbol{v}_j; \boldsymbol{g}_j\boldsymbol{\omega}_j^*, \mathbf{I}_T \otimes \boldsymbol{\Sigma}\right) \phi\left(\omega_j^*; \mu_j, \tau_j^2\right)}{\phi\left(\boldsymbol{v}_j; \mathbf{0}, \mathbf{I}_T \otimes \boldsymbol{\Sigma}\right) \phi\left(\omega_j^*; \widehat{\mu}_j, \widehat{\tau}_j^2\right)} \\ &= \frac{\Phi\left(\widehat{\mu}_j/\widehat{\tau}_j\right)}{\Phi\left(-\mu_j/\tau_j\right)} \times \frac{\widehat{\tau}_j}{\tau_j} \times \exp\left\{-\frac{1}{2}\left(\left(\boldsymbol{v}_j - \boldsymbol{g}_j\boldsymbol{\omega}_j^*\right)'\left(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1}\right)\left(\boldsymbol{v}_j - \boldsymbol{g}_j\boldsymbol{\omega}_j^*\right)\right.\right. \\ &\quad \left.\left. -\boldsymbol{v}_j'\left(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{v}_j + \frac{1}{\tau_j^2}\left(\omega_j^* - \mu_j\right)^2 - \frac{1}{\widehat{\tau}_j^2}\left(\omega_j^* - \widehat{\mu}_j\right)^2\right)\right\}. \end{aligned} \tag{19}$$

However, since

$$\frac{\widehat{\mu}_j}{\widehat{\tau}_j^2} - \frac{\mu_j}{\tau_j^2} = \boldsymbol{g}_j' \left(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{v}_j$$
$$\frac{1}{\widehat{\tau}_j^2} - \frac{1}{\tau_j^2} = \boldsymbol{g}_j' \left(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{g}_j,$$

(19) simplifies to

$$\widehat{\psi}_j = \frac{\Phi\left(\widehat{\mu}_j/\widehat{\tau}_j\right)}{\Phi\left(-\mu_j/\tau_j\right)} \frac{\widehat{\tau}_j}{\tau_j} \exp\left\{\frac{1}{2}\left(\frac{\widehat{\mu}_j^2}{\widehat{\tau}_j^2} - \frac{\mu_j^2}{\tau_j^2}\right)\right\}.$$

Next, we provide the details of Steps 1-3 in the MCMC sampler. For Step 1, sample

$$(\boldsymbol{\alpha} \,|\, \boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\omega}^*, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \lambda) \sim \mathcal{N}\left(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{A}}^{-1}\right),$$

where

$$\widehat{\boldsymbol{A}} = \boldsymbol{A}_0 + \boldsymbol{X}'(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{X}, \quad \widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{A}}^{-1}\left(\boldsymbol{A}_0\boldsymbol{\alpha}_0 + \boldsymbol{X}'(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1})(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\gamma})\right),$$

and $\boldsymbol{W}$ is obtained by defining $\boldsymbol{W}_t = \boldsymbol{X}\boldsymbol{\Omega}^{\frac{1}{2}}$ and stacking

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}_1 \\ \vdots \\ \boldsymbol{W}_T \end{pmatrix}.$$

For Step 2, we first define

$$\boldsymbol{H} = \begin{pmatrix} \mathbf{I}_m & & & \\ -\mathbf{I}_m & \mathbf{I}_m & & \\ & \ddots & \ddots & \\ & & -\mathbf{I}_m & \mathbf{I}_m \end{pmatrix}.$$

Then, sample

$$(\boldsymbol{\gamma} \,|\, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\omega}^*, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \lambda) \sim \mathcal{N}\left(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\Gamma}}^{-1}\right)$$

using the precision-based sampler in Chan and Jeliazkov (2009), where

$$\widehat{\boldsymbol{\Gamma}} = \boldsymbol{H}'\boldsymbol{H} + \boldsymbol{W}'(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{W}, \quad \widehat{\boldsymbol{\gamma}} = \widehat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{W}'(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}).$$

Finally, obtain

$$(\boldsymbol{\Sigma} \,|\, \boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\omega}^*, \boldsymbol{\tau}, \lambda) \sim \mathcal{IW}\left(\nu_0 + T, \widehat{\boldsymbol{\Sigma}}\right),$$

where

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_0 + \sum_{t=1}^{T} \left(\boldsymbol{y}_t - \boldsymbol{X}_t\left(\boldsymbol{\alpha} + \boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\gamma}_t\right)\right)\left(\boldsymbol{y}_t - \boldsymbol{X}_t\left(\boldsymbol{\alpha} + \boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\gamma}_t\right)\right)'.$$

Recall that the details related to the conditional sampling of $\boldsymbol{\omega}^*$, $\boldsymbol{\tau}$, and λ (Steps 4-6) are provided in Section 3.

A few remarks regarding this algorithm are in order. First, one may readily observe that $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ have an analytically tractable joint distribution of the form

$$(\boldsymbol{\alpha}, \boldsymbol{\gamma} \,|\, \boldsymbol{y}, \boldsymbol{\omega}^*, \boldsymbol{\Sigma}, \boldsymbol{\tau}, \lambda) \sim \mathcal{N}\left(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\Delta}}^{-1}\right), \tag{20}$$

where

$$\widehat{\boldsymbol{\Delta}} = \begin{pmatrix} \boldsymbol{A}_0 & \\ & \boldsymbol{H}'\boldsymbol{H} \end{pmatrix} + \begin{pmatrix} \boldsymbol{X}' \\ \boldsymbol{W}' \end{pmatrix} \left(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1}\right) \begin{pmatrix} \boldsymbol{X} & \boldsymbol{W} \end{pmatrix}$$

$$\widehat{\boldsymbol{\delta}} = \widehat{\boldsymbol{\Delta}}^{-1}\left(\begin{pmatrix} \boldsymbol{A}_0\boldsymbol{a}_0 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{X}' \\ \boldsymbol{W}' \end{pmatrix} \left(\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{y}\right).$$

However, the $mT \times mT$ matrix $\widehat{\boldsymbol{\Delta}}$ presents a number of computational difficulties. In particular, while $\widehat{\boldsymbol{\Delta}}$ is indeed a sparse matrix, it does not inherit a very convenient sparsity structure; operations such as inversion and Cholesky decomposition (both crucial to sampling from (20)) are computationally intensive, and in fact, quite inhibitive in large dimensional settings.

On the other hand, sampling $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ sequentially — i.e., drawing $(\boldsymbol{\alpha} \,|\, \boldsymbol{\gamma},\ \cdot\ )$ followed by $(\boldsymbol{\gamma} \,|\, \boldsymbol{\alpha},\ \cdot\ )$ — does not appear to generate significant autocorrelation in the MCMC chain, to the extent that it would justify the additional computational burden necessary to sample $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in a single block. In fact, a battery of tests applied under various settings and with various data consistently found that the computation time required to achieve the same level of efficiency was much lower under the sequential sampling scheme, in comparison to the blocking approach.

Moreover, the efficiency of the proposed algorithm may be improved by augmenting the six steps outlined in Section 3 with one or more *Generalized Gibbs* (GG) steps, as proposed in Liu and Sabatti (2000). The basic idea of such an augmentation is

to transform the draws obtained by recursive, conditional sampling within a typical Gibbs loop in such a way as to preserve the invariant target distribution of the Markov chain. Because the transformation involves additional randomly sampled quantities, this has the potential to introduce randomness into otherwise highly autocorrelated MCMC chains, and hence, boost sampling efficiency.

To this end, we identify three potentially beneficial GG moves in our context. We present these transformations assuming that $\mu_j = 0$ — an assumption maintained in all practical implementations of the algorithm outlined in Section 3. Generalizing to cases with non-zero μ_j is conceptually straightforward, but requires more involved sampling procedures, as subsequently discussed. In particular, a combination of one or more of the following GG steps have been found to improve the mixing of the MCMC chain generated by the proposed Gibbs algorithm:

1. Given the current draws of $\boldsymbol{\gamma}, \boldsymbol{\tau}^2, \lambda^2$, sample for each $j = 1, \ldots, m$

$$(\vartheta_j \,|\, \boldsymbol{\gamma}_j, \tau_j^2, \lambda^2) \sim \mathcal{GIG}\left(\frac{T-3}{2}, \tau_j^2\lambda^2, \boldsymbol{\gamma}_j'\widetilde{\boldsymbol{H}}'\widetilde{\boldsymbol{H}}\boldsymbol{\gamma}_j\right), \tag{21}$$

   and apply the transformations

$$\begin{aligned}
\tau_j^{2(\text{new})} &= \tau_j^2/\vartheta_j \\
\omega_j^{*(\text{new})} &= \omega^*/\sqrt{\vartheta_j}, \quad \omega_j^{(\text{new})} = \omega_j^{*(\text{new})}\mathbf{1}\left(\omega_j^{*(\text{new})} > 0\right) \\
\boldsymbol{\gamma}_j^{(\text{new})} &= \boldsymbol{\gamma}_j\sqrt{\vartheta_j},
\end{aligned}$$

   where $\boldsymbol{\gamma}_j$ is a $T \times 1$ vector of the elements in $\boldsymbol{\gamma}$ that correspond to the jth covariate, and

$$\widetilde{\boldsymbol{H}} = \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}.$$

   Likewise, $\mathcal{GIG}(\,\cdot\,)$ denotes the *Generalized Inverse Gamma* distribution and may be efficiently sampled from using the rejection method of Dagpunar (1989).

2. Given the current draws of $\boldsymbol{\omega}^*, \boldsymbol{\tau}_2^2, \lambda^2$, sample

$$(\vartheta \,|\, \boldsymbol{\omega}^*, \boldsymbol{\tau}_2^2, \lambda^2) \sim \mathcal{G}\left(\lambda_{01} + \frac{m}{2}, \lambda_{02}\lambda^2 + \frac{1}{2}\sum_{j=1}^{m}\frac{\omega_j^{*2}}{\tau_j^2}\right), \tag{22}$$

and apply the transformations

$$\lambda^{2(\text{new})} = \lambda^2 \vartheta, \quad \boldsymbol{\tau}^{2(\text{new})} = \boldsymbol{\tau}^2 / \vartheta.$$

3. Given the current draws of $\boldsymbol{\alpha}, \boldsymbol{\gamma}_1, \boldsymbol{\omega}$, and defining

$$\widehat{\boldsymbol{\Theta}} = \mathbf{I}_{nT} + \boldsymbol{\Omega}^{\frac{1}{2}} \boldsymbol{A}_0 \boldsymbol{\Omega}^{\frac{1}{2}}, \quad \widehat{\boldsymbol{\vartheta}} = \widehat{\boldsymbol{\Theta}}^{-1} \left(\boldsymbol{\gamma}_1 - \boldsymbol{\Omega}^{\frac{1}{2}} \boldsymbol{A}_0 (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \right),$$

where $\boldsymbol{\gamma}_1$ a $m \times 1$ vector of the elements in $\boldsymbol{\gamma}$ that correspond to the period $t = 1$, sample

$$(\boldsymbol{\vartheta} \,|\, \boldsymbol{\alpha}, \boldsymbol{\gamma}_1, \boldsymbol{\omega}) \sim \mathcal{N} \left(\widehat{\boldsymbol{\vartheta}}, \boldsymbol{\Theta}^{-1} \right),$$

and apply the transformations

$$\boldsymbol{\alpha}^{(\text{new})} = \boldsymbol{\alpha} + \boldsymbol{\omega} \odot \boldsymbol{\vartheta}, \quad \boldsymbol{\gamma}^{(\text{new})} = \boldsymbol{\gamma} - \boldsymbol{\vartheta},$$

where $\odot$ represents element-by-element multiplication.

It is worthwhile to note that transformations 1 and 2 incorporate a straightforward extension of the *scale group* explicitly discussed in Liu and Sabatti (2000). The third GG move presented above is based on an extension of the *translation group*, and in fact, is nearly identical to the transformation applied in the context of state-space models in example 4.1 of Liu and Sabatti (2000). Therefore, their Theorem 1 guarantees that the target posterior is preserved under all three types of moves, given that $\mu_j = 0$ holds. To that end, GG move 1 can be similarly implemented in the case that $\mu_j \neq 0$, but the distributions (21) from which the scales $\{\vartheta_j\}$ are sampled would no longer be valid; the resulting distributions that ensure the invariance of the target posterior (with $\mu_j \neq 0$) are of nonstandard from, and hence, would require alternative implementations of the sampling algorithms. Moves 2 and 3, however, are not materially affected by the $\mu_j = 0$ assumption (one would only require replacing ω_j^* with $\omega_j^* - \mu_j$ in (22)).

In our experience, implementing all three moves into the Gibbs sampler of section 3 tends to increase execution time by approximately 10%. In exchange, we have observed a reduction in *inefficiency factors* — in particularly favorable cases — of up to 15 times. In other scenarios, however, the improvement in sampling efficiency is much less pronounced. Therefore, the decision regarding whether or not to include the GG moves into the sampler generally relies on trial and error. In the application of section 4, we incorporate all three moves outlined above; doing so leads to reductions in inefficiency

factors of up to four times (depending on the variable) and noticeably more stable posterior estimates (given the same number of retained and discarded simulation draws).

As a final remark, we emphasize that a certain amount of care should be taken in handling the large matrices $\boldsymbol{W}$ $(nT \times mT)$ and $\boldsymbol{G}$ $(nT \times m)$. In particular, it should be noted that it is unnecessary — and impractical — to reconstruct these matrices entirely at each iteration of the Gibbs sampler. Rather, observing that both matrices are primarily sparse in nature, a reasonable algorithm would proceed by pre-allocating space prior to the commencement of the Gibbs loop and only updating the necessary elements in the course of execution.

Specifically, a prototype $\boldsymbol{W}$ can be preconstructed (say, by letting $\boldsymbol{W} = \mathbf{I}_{nT} \otimes \boldsymbol{\iota}_m'$, where $\boldsymbol{\iota}_m$ denotes a $m \times 1$ vector of ones), with the indices of the mT nonzero elements stored in w_{nz}. Then, at each iteration, it is only required to update the nonzero elements as

$$\boldsymbol{W}(w_{\text{nz}}) = \text{vec}\left(\boldsymbol{X} \odot \begin{pmatrix} \boldsymbol{\omega} & \cdots & \boldsymbol{\omega} \end{pmatrix}\right), \tag{23}$$

where it is assumed that $\boldsymbol{W}(w_{\text{nz}})$ represents the $mT \times 1$ vector of the nonzero elements in $\boldsymbol{W}$ (e.g. this notation is consistent with the syntax of a variety of popular scripting software, such as Matlab).

A similar approach provides an efficient procedure for handling $\boldsymbol{G}$ as well, except in this case it is operationally easier to work directly with its transpose $\boldsymbol{G}'$. Accordingly, one may initialize the sampler by setting $\boldsymbol{G}' = \boldsymbol{X}'$. Storing the indices of the mT nonzero elements in g_{nz}', these elements are updated at each iteration as

$$\boldsymbol{G}'(g_{\text{nz}}') = \text{vec}(\boldsymbol{X}) \odot \boldsymbol{\gamma}. \tag{24}$$

References

M. Banbura, D. Giannone, and L. Reichlin. Large Bayesian vector auto regressions. Journal of Applied Econometrics, 25(1):71–92, 2010.

M. Belmonte, G. Koop, and D. Korobolis. Hierarchical shrinkage in time-varying parameter models. CORE Discussion Papers 2011036, 2011.

O. Blanchard and R. Perotti. An empirical characterization of the dynamic effects of changes in government spending and taxes on output. The Quarterly Journal of

Economics, 117:1329–1368, 2002.

A. Carriero, T. Clark, and M. Marcellino. Bayesian VARs: specification choices and forecast accuracy. Federal Reserve Bank of Cleveland Working Paper 1112, 2011.

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(3):358–367, 2012.

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. Cogley, G. Primiceri, and T. Sargent. Inflation-gap persistence in the U.S. American Economic Journal: Macroeconomics, 2:43–69, 2010.

J. S. Dagpunar. An easily implemented generalised inverse Gaussian generator. Communications in Statistics - Simulation and Computation, 18(2):703–710, 1989.

S. Frühwirth-Schnatter and H. Wagner. Stochastic model specification search for Gaussian and partial non-Gaussian state space models. Journal of Econometrics, 154:85–100, 2010.

J. Galí, J. Vallés and J. D. López-Salido. Understanding the effects of government spending on consumption. Journal of the European Economic Association March 5(1):227–270, 2007.

E. I. George and R. McCulloch. Variable Selection via Gibbs Sampling. Journal of the American Statistical Association, 88: 881-889, 1993.

E. I. George and R. McCulloch. Approaches for Bayesian variable selection. Statistica Sinica, 7:339–373, 1997.

G. Koop. Forecasting with medium and large Bayesian VARs. Journal of Applied Econometrics, 28(2):177–203, 2013.

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, 2013. DOI: 10.1016/j.jeconom.

G. Koop and S. M. Potter. Time varying VARs with inequality restrictions. Journal of Economic Dynamics and Control, 35:1126–1138, 2011.

G. Koop, R. Leon-Gonzalez, and R.W. Strachan. On the evolution of the monetary policy transmission mechanism. Journal of Economic Dynamics and Control, 33(4):997–

1017, 2011.

D. Korobilis. VAR forecasting using Bayesian variable selection. Journal of Applied Econometrics, 2011. DOI: 10.1002/jae.1271.

J. S. Liu and C. Sabatti. Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation. Biometrika, 87(2):353–369, 2000.

J. Nakajima and M. West. Bayesian analysis of latent threshold dynamic models. Journal of Business and Economic Statistics, 31(2):151-164, 2013.

R. Perotti. Estimating the effects of fiscal policy in OECD countries. Proceedings, Federal Reserve Bank of San Francisco, 2005.

G. E. Primiceri. Time varying structural vector autoregressions and monetary policy. Review of Economic Studies, 72(3):821–852, 2005.