Multivariate Stochastic Volatility with Co-Heteroscedasticity

Joshua Chan
Arnaud Doucet
Roberto León-González
Rodney W. Strachan

October 2018

National Graduate Institute for Policy Studies
7-22-1 Roppongi, Minato-ku,
Tokyo, Japan 106-8677

Multivariate Stochastic Volatility with Co-Heteroscedasticity[1]

Joshua Chan

Purdue University

Arnaud Doucet

University of Oxford

Roberto León-González

National Graduate Institute for Policy Studies (GRIPS)

and

Rodney W. Strachan

University of Queensland

October 2018

ABSTRACT

This paper develops a new methodology that decomposes shocks into homoscedastic and heteroscedastic components. This specification implies there exist linear combinations of heteroscedastic variables that eliminate heteroscedasticity. That is, these linear combinations are homoscedastic; a property we call co-heteroscedasticity. The heteroscedastic part of the model uses a multivariate stochastic volatility inverse Wishart process. The resulting model is invariant to the ordering of the variables, which we show is important for impulse response analysis but is generally important for, e.g., volatility estimation and variance decompositions. The specification allows estimation in moderately high-dimensions. The computational strategy uses a novel particle filter algorithm, a reparameterization that substantially improves algorithmic convergence and an alternating-order particle Gibbs that reduces the amount of particles needed for accurate estimation. We provide two empirical applications; one to exchange rate data and another to a large Vector Autoregression (VAR) of US macroeconomic variables. We find strong evidence for co-heteroscedasticity and, in the second application, estimate the impact of monetary policy on the homoscedastic and heteroscedastic components of macroeconomic variables.

Keywords: Markov Chain Monte Carlo, Gibbs Sampling, Flexible Parametric Model, Particle Filter, Co-heteroscedasticity, state-space, reparameterization, alternating-order.

JEL Codes: C11, C15

1   Introduction

It is now well recognised that the variance for many macroeconomic and financial variables change over time. Many approaches have been proposed and employed to model this behaviour including Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized ARCH (GARCH) models (Engle (1982), Bollerslev (1986)) in which the variance of the reduced form errors is a deterministic function of past residuals and variances. Another important class of models of the variance are the stochastic volatility (SV) models. These differ from ARCH and GARCH models in that the conditional variance is an unobserved Markov Process.

For univariate models, it has been found that SV models forecast better than GARCH in terms of root mean squared forecast error and log-predictive scores for macro data (e.g. Chan (2013)) and in terms of log Bayes factors for financial data (e.g. Kim , Shephard and Chib (1998)). Similar evidence is found in multivariate models (e.g. Clark and Ravazzolo (2015)) where the SV specification provides the advantage of more flexible models since the GARCH approach requires many simplifying restrictions to ensure that the covariance matrix is positive definite.

In many applications, it is important to have an accurate estimate of the reduced form covariance matrix, Σ_t, and of its stationary value $E(\Sigma_t)$. Impulse response functions in VAR analysis of macroeconomic data, for example, rely on decompositions of the reduced form covariance matrix (e.g. Hamilton (1994, p.320)). To accurately report the impact of structural shocks requires a good estimate of the covariance matrix. Another example is in portfolio allocation based on financial data.

If the order of the variables in a vector affects the estimates of its covariance matrix then we can say that the estimation method is not invariant to ordering. An issue with the approaches employed in most of the literature on multivariate stochastic volatility models (e.g. Primiceri (2005), Chib *et al.* (2009)) is that they are not, in fact, invariant to ordering. A survey of the applied literature suggests that a common specification is to use a Cholesky decomposition of Σ_t, and let each element of this decomposition follow a Markov Process. However, this implies that the distribution of Σ_t and the predictive distribution of y_t depend on the ordering. An alternative approach to modelling multivariate volatility uses factor models. Factor models also require restrictions to permit estimation and, again, the most commonly employed restrictions on factor models are not invariant (Chan *et al.*, (2018)).

Another feature of the multivariate SV models used in the literature is that they use a log-normal distribution for the volatility. This specification implies that all moments of the volatility exist. In many applications, particularly in finance, features of the data suggest that the distributions have heavy tails and possibly few moments. To allow for the possibility of heavy tails, it is common to use Student's-t errors.

Another line of literature in modelling multivariate SV uses Wishart or inverted Wishart processes (see, for example, Uhlig (1997), Philipov and Glickman (2006), Casarin and Sartore (2007), Asai and McAleer (2009), Fox and West (2011), Triantafyllopoulos (2011), and Karapanagiotidis (2012)). A feature of this approach is that the estimates are invariant to ordering. Inverted Wishart models also allow for non-existence of higher moments and heavier tails. In a recent paper, León-González (2018) shows, in the univariate case, that Student's-t errors are often not needed to model heavy tails when an inverse gamma volatility process is used, but are very important when a log-normal volatility process is used. The approach in this paper uses an inverse Wishart process but differs from the above approaches in a number of respects. For example, the model is relatively unrestricted, computation is more efficient, the stationary covariance matrix is estimated and the specification permits an interesting decomposition of the errors into homoscedastic and heteroscedastic parts.

In this paper we propose inference on multivariate volatility in an Inverted Wishart SV model. The resulting inference is invariant to ordering of the variables and allows for heavy tails and even non-existence of higher order moments. We present a novel particle filter that samples all volatilities jointly conditionally on the unknown parameters using Particle Gibbs (Andrieu *et al.* (2010)) with backward sampling (Whiteley (2010)). The particle filter uses an approximation of the posterior distribution as a proposal density, which allows us to substantially reduce the computation time in models of higher dimension, and it is related to the ψ-Auxiliary Particle Filters studied in Guarniero et al. (2017). We additionaly use the approximation of the posterior to find a reparameterization of the model that substantially reduces the correlation between the latent states and the fixed parameters, which we empirically find to speed up computation and thus contributes to the literature on parameterization in state space and hierarchical models (e.g. Pitt and Shephard (1999), Papaspiliopoulos et al. (2007)). We also propose to alternate the order in which the volatilities are generated, finding that this reduces the number of particles needed for accurate estimation. Another important feature of our approach is that the specification results in a reduction in the dimension of the latent variables, greatly improving computation efficiency and permitting inference on potential co-heteroscedasticity. This allows us to obtain new variance decompositions as well as new insights into the characteristics of the structural shock and its impact on the variables of interest. Co-heteroscedasticity might also be relevant to find portfolio allocations with thinner tails and constant variance returns.

The structure of the paper is as follows. Section 2 describes the model and the identification strategy. Section 3 clarifies the difference between identifying assumptions and ordering of variables and presents decompositions of impulse responses and the variance. Section 4 presents the particle filter to estimate the likelihood while in Section 5 the particle Gibbs algorithm and tools for model comparison are presented. Section 6 discusses two applications to real data, one to a large VAR of macroeconomic

variables and another one to returns of exchange rates.

2 Model and Identification Strategy

We consider the following model of stochastic volatility:

$$y_t = \beta x_t + e_t, \quad \text{with} \ \ e_t \sim N(0, \Sigma_t)$$

where $\beta : r \times k_x$, $x_t : k_x \times 1$ and $e_t : r \times 1$. The vector e_t can be decomposed into r_1 heteroscedastic errors ($u_{1t} : r_1 \times 1$) and r_2 homoscedastic errors ($u_{2t} : r_2 \times 1$), with $r = r_1 + r_2$:

$$e_t = A_1 u_{1t} + A_2 u_{2t} = \left(\begin{array}{cc} A_1 & A_2 \end{array} \right) \left(\begin{array}{c} u_{1t} \\ u_{2t} \end{array} \right) = A u_t \tag{1}$$

We can then decompose Σ_t as:

$$\Sigma_t = var(e_t) = \underbrace{var(A_1 u_{1t})}_{\text{changing}} + \underbrace{var(A_2 u_{2t})}_{\text{constant}}$$

where as a normalization we fix $cov(u_{1t}, u_{2t}) = 0$ and $var(u_{2t}) = I_{r_2}$[2]. Note that this implies that there are r_2 linear combinations of e_t which are homoscedastic. In particular, if the $(r \times r_2)$ matrix $A_{1\perp}$ lies in the space of the orthogonal complement of A_1, then the $r_2 \times 1$ vector $A_{1\perp}' e_t$ has a constant variance. We say then that the elements of the vector y_t share heteroscedastic shocks, a property that we define as co-heteroscedasticity in the following definition.

Definition 1 *Let $\Im_t = (y_{t-1}, y_{t-2}, ..., x_t, x_{t-1}, ...,)$ be all the information available up to period t. An $r \times 1$ vector y_t with time variant conditional var-cov matrix (i.e. $var(y_t | \Im_t) = \Omega_{\Im_t}$) is said to be co-heteroscedastic if there exists a full rank matrix $A_{1\perp} : r \times r_2$ such that the vector $A_{1\perp}' y_t$ has a constant conditional var-cov matrix (i.e. $var(A_{1\perp}' y_t | \Im_t) = \Omega_0$ for all t), with $0 < r_2 < r$. The largest possible value of r_2 for which this property holds is the co-heteroscedasticity order.*

In our framework the heteroscedastic errors u_{1t} have $var(u_{1t}) = K_t^{-1}$, where $K_t = Z_t' Z_t$, and Z_t is a $n \times r_1$ matrix distributed as a Gaussian AR(1) process[3]:

$$Z_t = Z_{t-1} \rho + \varepsilon_t \qquad vec(\varepsilon_t) \sim N(0, I_{r_1} \otimes I_n) \tag{2}$$

---

[2]Because $var(e_t) = A var(u_t) A'$, it is not possible to identify $var(u_t)$ and A at the same time, hence the normalization.
[3]This representation implies that n is an integer, but in Section 5 we will use an equivalent representation that allows n to be continuous when $n \geq 2r_1$.

where ρ is diagonal $r_1 \times r_1$ (with diagonal elements smaller than one in absolute value) and we assume that $vec(Z_1)$ is drawn from the stationary distribution $N(0, (I_{r_1} - \rho^2)^{-1} \otimes I_n)$. So $var(e_t) = \Sigma_t$ can be written as:

$$\Sigma_t = A_1 K_t^{-1} A_1' + A_2 A_2' = A \begin{pmatrix} K_t^{-1} & 0 \\ 0 & I_{r_2} \end{pmatrix} A', \quad \text{for } A = \begin{pmatrix} A_1 & A_2 \end{pmatrix} \tag{3}$$

Because $var(y_t|x_t) = E(\Sigma_t)$, it is possible to identify $G = E(\Sigma_t)$, but restrictions are needed to identify $A = (A_1, A_2)$. For the purpose of identification, we assume that the time-varying component of Σ_t has on average bigger singular values than the constant component. The reason for this assumption is that it is well-known that stochastic volatility is empirically important, and therefore we introduce K_t^{-1} into the model in the way that it will have the greatest impact, which is by interacting K_t^{-1} with the largest singular values of G. To impose this assumption we use the singular value decomposition of $E(\Sigma_t)$ (SVD, e.g. Magnus and Neudecker (1999, p. 18)), which is given by:

$$G = E(\Sigma_t) = USU' = U_1 S_1 U_1' + U_2 S_2 U_2' \tag{4}$$

where U is a $r \times r$ orthogonal matrix such that $U'U = I_r$, $S = diag(s_1, ..., s_r)$ with $s_1 \geq s_2 \geq ... \geq s_r \geq 0$, and S_1 contains the r_1 biggest singular values, such that:

$$U = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \quad S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \tag{5}$$
$$U_1 : r \times r_1, \quad U_2 : r \times r_2, \ S_1 : r_1 \times r_1, \ S_2 : r_2 \times r_2$$

Let us standardize K_t^{-1} by defining $\Upsilon_t^{-1} = (n - r_1 - 1)(I - \rho^2)^{-1/2} K_t^{-1} (I - \rho^2)^{-1/2}$, such that $E(\Upsilon_t^{-1}) = I_{r_1}$. We impose the identifying assumption by writing:

$$\Sigma_t = U_1 S_1^{1/2} \Upsilon_t^{-1} S_1^{1/2} U_1' + U_2 S_2 U_2' \tag{6}$$

where we are also assuming that $n > r_1 + 1$, so that $E(\Sigma_t)$ is finite[4]. Comparing (3) and (6) we are able to identify A_1 and A_2 as follows:

---

[4]Note that the variance of any element of K_t^{-1} will only be finite when $n > r_1 + 3$ (e.g. Gupta and Nagar (2000, p.113)), and so the restriction $n > r_1 + 1$ still allows for very fat tails. Also note that the marginal of any diagonal element of K_t^{-1} is an inverted Gamma-2 distribution (Bauwens et al. (1999, p.p. 292, 305)) with $(n - r_1 + 1)$ degrees of freedom. Therefore, if $n = r_1 + 2$ (which is the minimum value that we allow in this paper), the marginal of a diagonal element of K_t^{-1} will have only 3 degrees of freedom, implying that it has finite mean but not finite variance.

$$
\begin{aligned}
A_1 &= \sqrt{(n - r_1 - 1)} U_1 S_1^{1/2} (I - \rho^2)^{-1/2} \qquad (7)\\
A_2 &= U_2 S_2^{1/2}
\end{aligned}
$$

We follow a Bayesian approach and put a prior directly on (G, ρ, n, β), where G enters the model through (A_1, A_2). This framework allows for a Normal-Wishart prior on (β, G), which simplifies the calculations in the context of estimating a large VAR. This prior allows for shrinkage, and implies that the conditional posterior of β is normal with a mean and var-cov matrix which can be calculated without inverting large matrices. In particular, the posterior mean and var-cov matrix of β can be calculated by inverting matrices of order $r_1 k_x$. This implies a reduction from $r k_x$ to $r_1 k_x$.

The distribution for K_t implied by (2) is known in the literature as Wishart Autoregressive process of order 1 (WAR(1)) and its properties are well established (e.g. Gourieroux et al. (2009), Koop et al. (2011)). For example, the stationary distribution of K_t is a Wishart distribution with n degrees of freedom and $E(K_t) = n(I - \rho^2)^{-1}$, and the conditional expectation of $K_t | K_{t-1}$ is equal to a weighted average of K_{t-1} and $E(K_t)$:

$$
E(K_t | K_{t-1}) = \rho K_{t-1} \rho + (I - \rho^2)^{1/2} E(K_t)(I - \rho^2)^{1/2}
$$

3 Invariance, Identification of Structural Shocks and Decompositions

When we refer to invariance to the ordering of the variables, we mean that the ordering of the variables in the vector y_t should not have any impact on inference. The ordering of the variables, understood in this sense, is a concept that should be distinguished from the identifying assumption to obtain the structural shocks, as any particular identifying assumption can be imposed regardless of the ordering of the variables in the vector y_t. For example, consider the identifying assumption that y_{1t} is a slow variable, such that it does not react contemporaneously to a shock to y_{2t}. This assumption is normally imposed by writing the reduced form shocks $e_t = (e_{1t}, e_{2t})$ as a function of orthogonal structural shocks as follows:

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} = \underbrace{\begin{pmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{pmatrix}}_{P} \begin{pmatrix} \xi_{1t} \\ \xi_{2t} \end{pmatrix}, \text{ with } PP' = var\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} = \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

The matrix P contains the parameters of interest and it is common to put a relatively non-informative prior on them, say $\pi(P)$. However, the same identifying assumption can also be used with the reverse ordering, first defining the following observationally-equivalent model with orthogonal shocks denoted by $(\widetilde{\xi}_{2t}, \widetilde{\xi}_{1t})$:

$$\begin{pmatrix} e_{2t} \\ e_{1t} \end{pmatrix} = \underbrace{\begin{pmatrix} \widetilde{p}_{11} & 0 \\ \widetilde{p}_{21} & \widetilde{p}_{22} \end{pmatrix}}_{\widetilde{P}} \begin{pmatrix} \widetilde{\xi}_{2t} \\ \widetilde{\xi}_{1t} \end{pmatrix}, \text{ with } \widetilde{P}\widetilde{P}' = var\begin{pmatrix} e_{2t} \\ e_{1t} \end{pmatrix} = \Sigma_o = \begin{pmatrix} \sigma_{22} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{pmatrix}$$

where Σ_o is equal to Σ after reordering the rows and columns. After specifying a prior on $\widetilde{P}$, say $\widetilde{\pi}(\widetilde{P})$, the model can be estimated, and the identifying assumption can be imposed ex-post by first reordering the rows and columns of the estimated $\Sigma_o = \widetilde{P}\widetilde{P}'$ to obtain an estimate of Σ, and then calculating its Cholesky decomposition to obtain estimates of the structural parameters in P. However, unless both $\pi(P)$ and $\widetilde{\pi}(\widetilde{P})$ imply the same prior on Σ, these two approaches will give different results (Sims and Zha (1998, eqns. 24-25, p. 961)), with the difference being very pronounced in the context of a large VAR with stochastic volatility, as we illustrate in the empirical section. Whenever precise prior information is available about P, it seems reasonable to put directly a prior on P. However, in most cases there is not such prior information, and this is specially the case when a block identifying assumption is used (as we do in Section 6.1, see also Zha (1999)), where a block of variables is assumed to react slowly, without making any assumption about which variables are slower within the block.

Equations (1) and (7) imply that the heteroscedastic component of e_t can be obtained as $e_t^{het} = U_1 U_1' e_t = H_1 e_t = A_1 u_{1t}$, such that y_t can be decomposed into its heteroscedastic (y_t^{het}) and homoscedastic components (y_t^{hom}). In the case of a VAR in which x_t contains a constant and lags of y_t, the decomposition can be obtained by using the moving average representation:

$$\begin{aligned} y_t &= y_t^{het} + y_t^{hom} = \mu + e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + ... & (8) \\ y_t^{het} &= H_1\mu + H_1 e_t + \Psi_1 H_1 e_{t-1} + \Psi_2 H_1 e_{t-2} + ... + \\ y_t^{hom} &= H_2\mu + H_2 e_t + \Psi_1 H_2 e_{t-1} + \Psi_2 H_2 e_{t-2} + ... \end{aligned}$$

where $H_2 = I_r - H_1 = U_2 U_2'$. Note that y_t^{het} is a sum of only heteroscedastic shocks, and y_t^{hom} is a sum of only homoscedastic shocks. After writing e_t in terms of orthogonal shocks $e_t = P\xi_t$, and using some assumption to identify P, the impulse response function can be decomposed as:

$$IR(s; i, j) = \frac{\partial y_{i,t+s}}{\partial \xi_{jt}} = \frac{\partial y_{i,t+s}^{het}}{\partial \xi_{jt}} + \frac{\partial y_{i,t+s}^{hom}}{\partial \xi_{jt}} = IR^{het}(s; i, j) + IR^{hom}(s; i, j)$$

where the partial derivatives correspond to the (i, j) elements of $\Psi_s P$, $\Psi_s H_1 P$ or $\Psi_s H_2 P$, respectively, with $\Psi_0 = I_r$. $IR^{het}(s; i, j)$ can be interpreted as the impact of ξ_{jt} on y_t^{het} at horizon s, and $IR^{hom}(s; i, j)$ can be interpreted in an analogous way. Equation (8) can also be used to obtain variance decompositions, calculating the proportion of $var(y_t|y_{t-h})$, $var(y_t^{het}|y_{t-h})$ or $var(y_t^{hom}|y_{t-h})$ caused by the structural shocks ξ_t or by the heteroscedastic shocks (u_{1t}), as we illustrate in the empirical section.

Importantly, we can also calculate the proportion of the variance of the structural shock ξ_t caused by the heteroscedastic shocks, by noting that $\xi_t = P^{-1} e_t = P^{-1}(A_1 u_{1t} + A_2 u_{2t})$, and therefore the variance of the heteroscedastic component of ξ_t is $var(P^{-1}(A_1 u_{1t})) = var(P^{-1}(H_1 e_t)) = P^{-1} H_1 G H_1' \left(P^{-1} \right)'$.

Finally, note that by decomposing y_t into y_t^{het} and y_t^{hom} using equation (8), we can estimate how y_t would behave if there were only heteroscedastic shocks (i.e. $u_{2t} = 0$ and $y_t = y_t^{het}$) or there were only homoscedastic shocks (i.e. $u_{1t} = 0$ and $y_t = y_t^{hom}$), as we illustrate in the empirical section.

4 Likelihood and Particle Filter

The value of the likelihood evaluated at the posterior means of the parameters can be used to calculate the Bayesian Information Criterion (BIC) or the marginal likelihood (Chib and Jeliazkov (2001)) for selecting r_1. However, because the likelihood cannot be calculated in analytical form, we propose a particle filter that provides a numerical approximation. This particle filter will also be a key ingredient of the particle Gibbs algorithm that we use to sample from the posterior distribution, and that we describe in Section 5.

Although one could in principle use a bootstrap particle filter (Gordon et al. (1994)), such approach would require too much computation time, especially as r_1 increases, and could become impractical when the data contains extreme observations, which are often abundant in economic or financial datasets[5].

The bootstrap particle filter uses the prior distribution of $K_{1:T}$, with density $\pi(K_{1:T}|\theta)$ and $\theta = (G, \rho, n)$, as a proposal density, and then the resampling weights are given by the likelihood $L(Y|\Sigma_{1:T}, \beta)$, where Y represents the observed data and $\Sigma_{1:T} = (\Sigma_1, ..., \Sigma_T)$. In order to define a more efficient particle filter, we first find a convenient approximation of the posterior $\pi(K_{1:T}|Y, \theta)$, denoted as the pseudo

---

[5]León-González (2018) proposes methods of inference for the univariate version of the model proposed in this paper, and finds that a particle Metropolis-Hasting algorithm that uses the bootstrap particle filter would have an effective sample size for n of only 0.29 per minute when $T = 2000$, when using only one core.

posterior $\widetilde{\pi}(K_{1:T}|Y,\theta)$, such that $\pi(K_{1:T}|\theta)L(Y|\Sigma_{1:T},\beta) \propto \widetilde{\pi}(K_{1:T}|Y,\theta)R(K_{1:T})$, for some function $R(.)$. Then we use $\widetilde{\pi}(K_{1:T}|Y,\theta)$ as a proposal density, such that the resampling weights will be determined by $R(K_{1:T})$. We can expect that the particle filter will be more efficient when $\widetilde{\pi}(K_{1:T}|Y,\theta)$ approximates $\pi(K_{1:T}|Y,\theta)$ well, and we can expect an improvement over the bootstrap particle filter whenever $\widetilde{\pi}(K_{1:T}|Y,\theta)$ is better than $\pi(K_{1:T}|\theta)$ as an approximation of $\pi(K_{1:T}|Y,\theta)$[6].

In order to define the pseudo posterior, note that because of the Gaussian assumption about e_t, the density of Y given $\Sigma_{1:T}$ is given by:

$$L(Y|\Sigma_{1:T},\beta) = \left(\prod_{t=1}^{T}|\Sigma_t|^{-1/2}\right)\widetilde{L}(Y|\Sigma,\beta)$$

where $\widetilde{L}(Y|\Sigma,\beta)$ is a pseudo-likelihood defined as:

$$\widetilde{L}(Y|\Sigma_{1:T},\beta) = (2\pi)^{-Tr/2}\exp\left(-\frac{1}{2}\sum_{t=1}^{T}tr\left(\Sigma_t^{-1}e_t e_t'\right)\right), \text{ where } e_t = y_t - \beta x_t \tag{9}$$

Exploiting the fact that the prior is conjugate for the pseudo likelihood, we define the pseudo-posterior as $\widetilde{\pi}(K_{1:T}|Y,\theta) \propto \pi(K_{1:T}|\theta)\widetilde{L}(Y|\Sigma_{1:T},\beta)$, which turns out to be (as shown in Appendix I) also a WAR(1) process, and can be represented as $K_t = Z_t'Z_t$ with:

$$
\begin{aligned}
Z_t &= Z_{t-1}\rho V_t + \varepsilon_t & vec(\varepsilon_t) &\sim N(0, V_t \otimes I_n), \text{ for } t > 1 \\
Z_1 &= \varepsilon_1 & vec(\varepsilon_1) &\sim N(0, V_1 \otimes I_n)
\end{aligned}
\tag{10}
$$

where V_t is given by the following recursion:

$$
\begin{aligned}
V_T &= (I + B_1'e_T e_T'B_1)^{-1} \\
V_t &= (I + B_1'e_t e_t'B_1 + \rho(I - V_{t+1})\rho)^{-1} & t > 1 \\
V_1 &= (I + B_1'e_1 e_1'B_1 - \rho V_2\rho)^{-1}
\end{aligned}
\tag{11}
$$

where $B = (B_1, B_2) = (A^{-1})'$, with $B_1 : r \times r_1$. Appendix I shows that the true likelihood, after integrating out the latent K, can be compactly written as:

$$
\begin{aligned}
L(Y|\beta,\theta) &= \widetilde{c}_L E_{\widetilde{\pi}}\left(\prod_{t=1}^{T}|K_t|^{1/2}\right) \tag{12} \\
\widetilde{c}_L &= |I - \rho^2|^{n/2}|A_1 A_1' + A_2 A_2'|^{-T/2}\exp\left(-\frac{1}{2}\sum_{t=1}^{T}tr\left(B_2 B_2'e_t e_t'\right)\right)\left(\prod_{t=1}^{T}|V_t|^{n/2}\right) \tag{13}
\end{aligned}
$$

---

[6]The particle filter that we propose can be viewed as a bootstrap particle filter on a 'twisted model', and falls into the class of ψ-Auxiliary Particle Filters discussed in Guarniero et al. (2017).

where the expectation is taken with respect to the pseudo-posterior $\widetilde{\pi}(K_{1:T}|Y,\theta)$. Because this expectation cannot be calculated in analytical form, we propose a particle filter that provides a numerical approximation.

An unbiased estimate of the expectation in (12) can be obtained using a particle filter in which the proposal density is given by the pseudo-posterior $\widetilde{\pi}(K_t|K_{1:t-1},Y,\theta)$. Using N particles, denoted as $\{K_{1:T}^k = (K_1^k, ..., K_T^k)\}$ for $k = 1, ..., N$, the particle filter can be described as follows (e.g. Andrieu et al. (2010, p. 272)):

Algorithm 2 *Particle filter.*

Step 1: at time $t = 1$,

    (a) sample K_1^k from a Wishart $W_{r_1}(n, V_1)$ for every $k = 1, ..., N$, and

    (b) compute the weights:

$$C_1 := \frac{1}{N}\sum_{m=1}^{N}\left|K_1^m\right|^{0.5}, \quad W_1^k := \frac{\left|K_1^k\right|^{0.5}}{NC_1}$$

    Step 2: at times $t = 2, ..., T$,

    (a) sample the indices A_{t-1}^k, for every $k = 1, ..., N$, from a multinomial distribution on $(1, ..., N)$ with probabilities $\mathbf{W}_{t-1} = (W_{t-1}^1, ..., W_{t-1}^N)$

    (b) sample K_t^k from a non-central Wishart $W_{r_1}(n, V_t, \rho K_{t-1}^{A_{t-1}^k}\rho V_t)$ and

    (c) compute the weights

$$C_t := \frac{1}{N}\sum_{m=1}^{N}\left|K_t^m\right|^{0.5}, \quad W_t^k := \frac{\left|K_t^k\right|^{0.5}}{NC_t}$$

    Step 3: Estimate the Likelihood value as:

$$\widehat{L}(Y|\beta,\theta) := \widetilde{c}_L \prod_{t=1}^{T} C_t$$

where $W_{r_1}(n, V_t, \Omega)$ denotes a non-central Wishart distribution with noncentrality parameters Ω (e.g. Muirhead 2005, p. 442).

When $n \geq 2r_1$ a draw K_t from $W_{r_1}(n, V_t, \rho K_{t-1}\rho V_t)$ can be obtained (e.g. Anderson and Girshick (1944, pp. 347-349) or Appendix II) by drawing a matrix $L_{1t} : r_1 \times r_1$ from a normal ($vec(L_{1t}) \sim N(vec((K_{t-1})^{1/2}\rho V_t), I_{r_1} \otimes V_t)$), K_{2t} from a Wishart $W_{r_1}(n - r_1, V_t)$ and calculating $K_t = (L_{1t})'L_{1t} + K_{2t}$, where $(K_{t-1})^{1/2}$ is any matrix such that $((K_{t-1})^{1/2})'(K_{t-1})^{1/2} = K_{t-1}$ (for example the upper triangular Cholesky factor). When $r_1 \leq n < 2r_1$ and n is an integer, a draw K_t from $W_{r_1}(n, V_t, \rho K_{t-1}\rho V_t)$ can be obtained by drawing $Z_t : n \times r_1$ from a Normal ($vec(Z_t) \sim N(vec(\mu_Z), I_n \otimes V_t)$), with

$\mu_Z = ((K_{t-1})^{1/2} \rho V_t)', 0_{r_1 \times (n-r_1)})'$ and calculating $K_t = Z_t' Z_t$.

5 Particle Gibbs and Model Comparison

In order to define the Gibbs algorithm, we rewrite the WAR(1) process in (2) using the representation of a non-central Wishart in Anderson and Girshick (1944, p.p. 347-349), which has been used for constructing simulation algorithms (e.g. Gleser (1976)). This representation writes a non-central Wishart matrix with n degrees of freedom K_t as $K_t = L_{1t}' L_{1t} + K_{2t}$, where L_{1t} is a $r_1 \times r_1$ normally distributed matrix and K_{2t} is a Wishart density with $n - r_1$ degrees of freedom. Applying this representation to the WAR(1) process in (2), we get the following:

$$
\begin{aligned}
K_t &\sim W(n, (I - \rho^2)^{-1}), \text{ for } t = 1 &(14)\\
L_{1t} &= (K_{t-1})^{1/2}\rho + \varepsilon_{1t}, \quad \varepsilon_{1t} : r_1 \times r_1, \quad vec(\varepsilon_{1t}) \sim N(0, I_{r_1} \otimes I_{r_1}), \text{ for } t > 1\\
K_{2t} &\sim W(n - r_1, I_{r_1}), \text{ for } t > 1
\end{aligned}
$$

where $(K_{t-1})^{1/2}$ is any matrix such that $((K_{t-1})^{1/2})'(K_{t-1})^{1/2} = K_{t-1}$ (for example the upper triangular Cholesky factor). This representation allows n to be a continuous parameter but requires that $n \geq 2r_1$ (otherwise K_{2t} would be singular). When $n \leq 2r_1$ we assume that n is an integer and write $K_{2t} = L_{2t}' L_{2t}$, where L_{2t} is a $(n - r_1) \times r_1$ matrix such that $vec(L_{2t}) \sim N(0, I_{r_1} \otimes I_{n-r_1})$ for $t > 1$ and $vec(L_{2t}) \sim N(0, (I_{r_1} - \rho^2)^{-1} \otimes I_{n-r_1})$ for $t = 1$.

As a prior for n we assume a discrete probability distribution in the interval $[r_1 + 2, 2r_1]$ and a continuous density on $(2r_1, \infty)$. The continuous density is specified with a normal prior on $\widetilde{n} = \log(n - 2r_1)$.

Our algorithm for simulating from the posterior distribution groups the parameters and latent states in 3 main blocks: $(L_{1,1:T}, K_{2,1:T})$, β and (G, ρ, n). The latent states $(L_{1,1:T}, K_{2,1:T}) = (L_{11}, ..., L_{1T}, K_{21}, ..., K_{2T})$ are drawn using a Particle Gibbs algorithm with Backward Sampling (Andrieu et al. (2010) and Whiteley (2010)), β is drawn from a normal distribution and the parameters in (G, ρ, n) are generated jointly using a reparameterization and a sequence of Metropolis steps.

The latent states can be generated starting from $t = 1$ up $t = T$ (natural order), or in the reverse order (starting at $t = T$ and continuing up to $t = 1$). In the natural order, the mixing properties of the states tend to be better for the states near $t = T$, whereas in the reverse order the mixing tends to be better for the states near $t = 1$. Although one solution to obtain equally good mixing properties for all t is to increase the number of particles, this requires extra computational cost. Here we propose a strategy

which consists in alternating between the natural order and the reverse order at different iterations of the algorithm. In this way we are using a mixture of two MCMC kernels, resulting in an algorithm that we find performs empirically better than the use of only one of them at no extra computational cost[7]. In particular we find that using the Macroeconomic data described in Section 6.1 with $r_1 = 7$, the alternating order algorithm delivers an Effective Sample Size (ESS) for $\overline{K} = (K_1 + ... + K_T)/T$ that is 37% (33%) higher than the ESS of the natural order (reverse order) algorithm, respectively, when the number of particles is 80 (see appendix II for more details).

5.1 Drawing the latent states $(L_{1,1:T}, K_{2,1:T})$ in natural order

Let $L_t = (L_{1t}, K_{2t})$ for $t > 1$ and $L_1 = K_1$, and $L_{1:T} = (L_1, L_2, ..., L_T)$. Let the N particles be denoted as $L_t^k = (L_{1t}^k, K_{2t}^k)$ for $t > 1$ and $L_1^k = (K_1^k)$, for $k = 1, ..., N$. Define $K_t^k = (L_{1t}^k)'L_{1t}^k + K_{2t}^k$ for $t > 1$. The value of $L_{1:T}$ at iteration i, denoted as $L_{1:T}(i) = (L_1(i), ..., L_T(i))$, can be generated given the previous value $L_{1:T}(i-1)$ using a conditional Particle Filter with Backward sampling (cPFBS).

Algorithm 3 *cPFBS.*

Step 1: Fix the last particle equal to $L_{1:T}(i-1)$, that is, $L_{1:T}^N = L_{1:T}(i-1)$

Step 2: at time $t = 1$,

(a) sample $K_t^k \sim W_{r_1}(n, V_1)$ for $k = 1, ..., N-1$, and

(b) compute and normalize the weights:

$$w_1^k := \left|K_1^k\right|^{0.5}, \ W_1^k := w_1^k / (\sum_{m=1}^{N} w_1^m), \ k = 1, ..., N$$

Step 3: at times $t = 2, ..., T$,

(a) sample the indices A_{t-1}^k, for $k = 1, ..., N-1$, from a multinomial distribution on $(1, ..., N)$ with probabilities $\mathbf{W}_{t-1} = (W_{t-1}^1, ..., W_{t-1}^N)$

(b) sample $vec(L_{1t}^k) \sim N((K_{t-1}^{A_{t-1}^k})^{1/2} \rho V_t, I_{r_1} \otimes V_t)$ and $K_{2t}^k \sim W_{r_1}(n - r_1, V_t)$, calculate $K_t^k = (L_{1t}^k)'L_{1t}^k + K_{2t}^k$ for $k = 1, ..., N-1$ and

(c) compute and normalize the weights

$$w_t^k := \left|K_t^k\right|^{0.5}, \quad W_t^k := w_t^k / (\sum_{m=1}^{N} w_t^m), \ k = 1, ..., N$$

Step 4: at time $t = T$, sample b_T from a multinomial distribution on $(1, ..., N)$ with probabilities $\mathbf{W}_T$, and set $L_T(i) = L_T^{b_T}$.

---

[7]There is an ample literature on the use of mixture of kernels to improve MCMC algorithms, see for example Andrieu et al. (2003, section 3.3) for a review.

Step 5: at times $t = T - 1, ..., 1$

(a) compute the updated weights

$$\widetilde{w}_t^k = w_t^k f(K_{t+1}^{b_{t+1}}|K_t^k),..., \widetilde{W}_t^k := \widetilde{w}_t^k / (\sum_{m=1}^{N} \widetilde{w}_t^m), \ k = 1, ..., N, \ where$$

$$\mu_t^k = (K_t^k)^{1/2} \rho V_{t+1}, \quad f(K_{t+1}^{b_{t+1}}|K_t^k) = \exp(-\frac{1}{2} tr(V_{t+1}^{-1}(L_{1(t+1)}^{b_{t+1}} - \mu_t^k)'(L_{1(t+1)}^{b_{t+1}} - \mu_t^k))$$

(b) sample b_t from a multinomial distribution on $(1, ..., N)$ with probabilities $\widetilde{\mathbf{W}}_t = (\widetilde{W}_t^1, ..., \widetilde{W}_t^N)$, and set $L_t(i) = L_t^{b_t}$.

5.2 Drawing the latent states $(L_{1,1:T}, K_{2,1:T})$ in reverse order

Because we assume that Z_1 is drawn from the stationary distribution, the WAR(1) process in equation (2) can be equivalently written in reverse order as $Z_t = Z_{t+1}\rho + \varepsilon_t$, with Z_T drawn from the stationary distribution (see Appendix II for a proof). However, we show in the Appendix II that to use the representation in (14) we need to define first $(\widetilde{L}_{1t}, \widetilde{K}_{2t})$ as follows:

$$
\begin{align}
\widetilde{L}_{1t} &= (K_{t+1}^{-1/2})' L'_{1,t+1} K_t^{1/2}, \quad \text{for } t = 1, ..., (T-1) \tag{15} \\
\widetilde{K}_{2t} &= K_t - \widetilde{L}'_{1t} \widetilde{L}_{1t}
\end{align}
$$

Using this definition we can write the transition equation in reverse order as:

$$
\begin{align}
K_t &\sim W(n, (I - \rho^2)^{-1}), \text{ for } t = T \tag{16} \\
\widetilde{L}_{1t} &= (K_{t+1})^{1/2}\rho + \varepsilon_{1t}, \quad \varepsilon_{1t} : r_1 \times r_1, \quad vec(\varepsilon_{1t}) \sim N(0, I_{r_1} \otimes I_{r_1}), \text{ for } t < T \\
\widetilde{K}_{2t} &\sim W(n - r_1, I_{r_1}), \text{ for } t < T
\end{align}
$$

Then the pseudo-posterior in (10) can be written in reverse order by first adapting the recursion in (11) as follows:

$$
\begin{align}
\widetilde{V}_1 &= (I + B'_1 e_1 e'_1 B_1)^{-1} \tag{17} \\
\widetilde{V}_t &= (I + B'_1 e_t e'_t B_1 + \rho(I - \widetilde{V}_{t-1})\rho)^{-1} \quad t < T \\
\widetilde{V}_T &= (I + B'_1 e_T e'_T B_1 - \rho\widetilde{V}_{T-1}\rho)^{-1}
\end{align}
$$

and then writing the pseudo-posterior in reverse order as:

$$
\begin{aligned}
K_t &\sim W(n, \widetilde{V}_t), \text{ for } t = T \\
\widetilde{L}_{1t} &= (K_{t+1})^{1/2} \rho \widetilde{V}_t + \varepsilon_{1t}, \quad \varepsilon_{1t} : r_1 \times r_1, \quad vec(\varepsilon_{1t}) \sim N(0, \widetilde{V}_t \otimes I_{r_1}), \text{ for } t < T \\
\widetilde{K}_{2t} &\sim W(n - r_1, \widetilde{V}_t), \text{ for } t < T
\end{aligned}
$$
(18)

A similar conditional particle filter to that defined in Section 5.1 can be defined to draw $\widetilde{L}_{1:T}$ (see Algorithm 4 in Appendix II for details). A draw of $\widetilde{L}_{1:T}$ can be then converted into a draw of $L_{1:T}$ using the inverse transformation of (15):

$$
\begin{aligned}
L_{1t} &= (K_{t-1}^{-1/2})' \widetilde{L}'_{1,t-1} K_t^{1/2}, \quad \text{for } t = 2, ..., T \\
K_{2t} &= K_t - L'_{1t} L_{1t}
\end{aligned}
$$
(19)

5.3 Drawing $\theta = (G, \rho, n)$

It is well known that the choice of parameterization can have an important impact on the computational efficiency of MCMC algorithms in state space and hierarchical models (e.g. Pitt and Shephard (1999), Papaspiliopoulos et al. (2007)). In line with this literature, we compare two algorithms, one generates θ conditional on the latent states $L_{1:T}$ and the other one conditional on a one-to-one differentiable transformation of the states $f_\theta(L_{1:T})$, where the transformation depends on θ. In the first case we use a Metropolis step that targets the conditional posterior $\pi(\theta|L_{1:T}, \beta)$ and in the other we target $\pi^t(\theta|\epsilon, \beta)$, where $\epsilon = f_\theta(L_{1:T})$ such that the conditional density $\pi^t(\theta|\epsilon, \beta)$ can be obtained by the change of variables theorem as:

$$
\pi^t(\theta|\epsilon, \beta) \propto \pi(\theta|f_\theta^{-1}(\epsilon), \beta) J
$$

where J is the Jacobian of the transformation. Note that although we keep ϵ constant when we generate θ, the latent states in the original parameterization might change. That is, when we condition on ϵ, a new value generated for θ (say θ^*) implies that the latent states have to updated as[8] $L_{1:T}^* = f_{\theta^*}^{-1}(\epsilon)$. Using the transformation will be more efficient when θ is less correlated with ϵ than with $L_{1:T}$.

To find an efficient parameterization, we propose to obtain first a convenient approximation to the distribution of the conditional posterior of the states given parameters $\pi^a(L_{1:T}|\theta)$, in which it is possible to find ϵ such that in the approximated density, ϵ becomes independent of θ: $\pi^a(\epsilon|\theta) = \pi^a(\epsilon)$. This strategy could be used in conjunction with existing linear Gaussian approximation methods (e.g.

---

[8]In Section 5.5 we give the overall summary of the MCMC algorithm, and we use the so-called State-Space expansion of the Gibbs sampler described in Papaspiliopoulos et al. (2007).

Durbin and Koopman (ch. 11)), given that in linear Gaussian state-models the standardized residuals are independent of θ. We can expect that the better the approximation, the weaker the dependence of ϵ and θ in the true posterior, and hence the better the reparameterization[9]. In our context, we can apply this strategy by rewriting the pseudo-posterior in (10) using the decomposition of the non-central Wishart:

$$K_t \quad \sim \quad W(n, V_t), \text{ for } t = 1 \tag{20}$$

$$L_{1t} \quad = \quad (K_{t-1})^{1/2}\rho V_t + \varepsilon_{1t}, \quad \varepsilon_{1t} : r_1 \times r_1, \quad vec(\varepsilon_{1t}) \sim N(0, V_t \otimes I_{r_1}), \text{ for } t > 1$$

$$K_{2t} \quad \sim \quad W(n - r_1, V_t), \text{ for } t > 1$$

Define the standardized residuals as $a_{1t} = \varepsilon_{1t}(V_t)^{-1/2}$. To standardize $\widehat{K}_{2t}$ we first define $\widehat{K}_{2t} = (V_t^{-1/2})'K_{2t}(V_t)^{-1/2}$, and then, to eliminate the dependence on the degrees of freedom parameter n, we use the Bartlett decomposition. For this purpose define the Cholesky factor Δ_t such that $\widehat{K}_{2t} = \Delta_t'\Delta_t$, and let c_t be the off-diagonal elements of Δ_t. Let p_t be the result of evaluating the distribution function of a χ^2 distribution at the squared of the diagonal elements of Δ_t. Then we define the reparameterization as $\epsilon = (a_{1t}, p_t, c_t : t = 1, ..., T)$, which is independent of θ in the pseudo-posterior. The details of the conditional density $\pi^t(\theta|\epsilon, \beta)$ are given in Appendix III, as well as a simulation that shows that, accounting for computation time, this reparameterization is 16 times more efficient to sample n and 4.7 times more efficient to sample ρ in terms of Effective Sample Size (ESS) when using the data in Section 6.1 with $r_1 = 7$.

To generate θ we use a Metropolis step repeated a number of times with an AR(1) proposal, using an inverse Wishart for (G), and a normal for $\widetilde{n} = \log(n - 2r_1)$ and $\widetilde{\rho} = \ln(-\ln(1 - \rho^2))$ (details in Appendix III).

5.4   Drawing β

The matrix of coefficients β has dimension $r \times k$, where the k is the number of rows of x_t. The conditional posterior of β is a normal, so when rk is not large, it is easy to generate β. However, in a large VAR the value of rk is large (equal to 1620 in our application, with $r = 20$ and $k = 81$) and so the time to calculate the posterior var-cov matrix of β and its Cholesky decomposition is not negligible. Fortunately, as shown in Appendix IV, we only need to invert matrices of the order r_1k to generate β. For example, in our empirical analysis with a large VAR and $r = 20$ we conclude that the best value of

---

[9]This approach is slightly different from that in Papaspiliopoulos et al. (2007) and Pitt and Shephard (1999), who compare the efficiency of centred versus uncentred parameterizations. Papaspiliopoulos et al. (2007) defines centred parameterizations as those in which Y is indepedent of θ given $L_{1:T}$, and uncentred when ϵ and θ are independent in the prior. Our approach is instead to make ϵ and θ independent in an approximation of the posterior.

r_1 is 7, and therefore we only need to invert matrices of order 567 (as opposed to 1620). The details of the conditional posterior of β are given in Appendix IV.

5.5   Summary of the Algorithm

Let $(L_{1:T}(i-1), \theta(i-1), \beta(i-1))$ be the values of $(L_{1:T}, \theta, \beta)$ at the $(i-1)^{th}$ iteration. The values at the ith iteration are generated as follows:

- If i is even:

    - Generate a value $L_{1:T}^*$ for $L_{1:T}$, using the natural order Algorithm 3 in section 5.1.

- If i is odd: Calculate $\widetilde{L}_{1:T}(i-1)$ from $L_{1:T}(i-1)$ using the inverse transformation (15).

    - Generate a value $\widetilde{L}_{1:T}^*$ of $\widetilde{L}_{1:T}$ in reverse order using Algorithm 4 in Appendix II.

    - Calculate $L_{1:T}^*$ from $\widetilde{L}_{1:T}^*$ using transformation (19).

- Calculate the transformation $\epsilon = f_{\theta(i-1)}(L_{1:T}^*)$ using Algorithm 5 in Appendix III.

- Generate $\theta(i)|\epsilon$ using a Metropolis step targeting the conditional distribution $\pi^t(\theta|\epsilon, \beta)$ in expression (26) of Appendix III.

- Fix $L_{1:T}(i)$ as $L_{1:T}(i) = f_{\theta(i)}^{-1}(\epsilon)$ using the inverse transformation outlined in Algorithm 6 in Appendix III.

- Draw $\beta(i)$ from a multivariate Normal Density with mean $\overline{\mu}_\beta$ and variance $\overline{V}_\beta$ described in Proposition 8 in Appendix IV.

5.6   Model Comparison

In order to select the value of r_1 we can use the BIC, which can be calculated by evaluating the log likelihood at the posterior mean of the parameters using the particle filter. Note that when r_1 increases by 1, the number of parameters only increases by 1, so the BIC easily allows us to assess whether it is worthwhile increasing r_1 in terms of the likelihood gains. We also calculate the marginal likelihood using the approach of Chib and Jeliazkov (2001), but using an approximation that allows us to calculate the value of the posterior density of (G, β) at the posterior mean without doing any additional simulations. For this we use the value of a Normal-Wishart density calibrated with the posterior mean of β and G, which are calculated with the MCMC algorithm. For the posterior density ordinate of $\widetilde{n} = \log(n - 2r_1)$ and of $\widetilde{\rho} = \ln(-\ln(1-\rho^2))$ we use a normal approximation. We also use the BIC and marginal likelihood to compare the model that assumes $n \geq 2r_1$ with the model that assumes $n = r_1 + 2$.

6 Empirical Application

6.1 Macroeconomic Data

We use 20 macroeconomic quarterly variables (defined in Table 1) with 4 lags in the VAR, and a normal-inverse-Wishart prior with shrinkage for (β, G)[10]. In order to compare models, we calculate the log predictive likelihood value for observations $t = 51...214$. This is calculated as the difference of two log marginal likelihoods (one with $T = 214$ and the other with $T = 50$). Table 2 shows model selection criteria when the restriction $n \geq 2r_1$ is imposed, indicating that the best value for r_1 is 7 according to the predictive likelihood. We also estimated models that impose the restriction $n = r_1 + 2$, finding that their predictive likelihoods, not shown here for brevity, suggested a value of r_1 equal to 5, but were smaller than for models with the restriction $n \geq 2r_1$, leading us to choose $r_1 = 7$. Estimating the preferred model took 3 hours in a Intel Xeon CPU (E5-2690) with 2.9 GHz, using 10 cores, and obtaining 10000 iterations with no thinning, a burn-in of 20 iterations and 80 particles[11]. The evaluation of the loglikelihood, using the average of 100 independent replications of the particle filter, each one with $Tr_1 = 1498$ particles, took 1.5 minutes, with a numerical standard error of 0.7. For the preferred model, Figure 1 shows the trace plot and autocorrelations for the seventh diagonal element of K_{107} (107 is the middle period in the sample), for the 4th element of ρ, for n and for the (1,1) element of G. In all cases the autocorrelations decrease below 0.2 after 20 lags, indicating good convergence. For each of these parameters, and also for the elements of $\overline{K} = (K_1 + ... + K_T)$, the Monte Carlo Standard Error (MCSE) was less than 4% of the posterior standard deviation[12].

Figure 2 shows the posterior estimates of the reduced form volatility of the GDP growth (GDP) and the real stock returns (SP) using the non-invariant method, prevalent in the econometrics literature, that specifies random walks as transition equations for the log of the diagonal elements of the Cholesky decomposition (e.g. Carriero et al. (2015), Chiu et al. (2017), Clark and Ravazzolo (2015)) and our invariant method, as well as the Bayesian squared residuals (which we define as the squared value of an element of the vector $(y_t - \widehat{\beta}x_t)$, with $\widehat{\beta}$ being the posterior mean). Defining ordering 1 with GDP as the first variable and SP the last, and ordering 2 with SP the first and GDP the last (the other 18 variables remaining in the same place), we see that the non-invariant method is very sensitive to the ordering. For example, the volatility of SP in 2008 Q4 is 35 under ordering 1 and 70 under ordering

---

[10]This is the same dataset used in Chan (2018), and the shrinkage prior has $k_1 = 0.6, k_2 = 100$, with (k_1, k_2) defined in Chan (2018). The degrees of freedom parameter for the inverse-Wishart prior of G is $r + 3$. The prior for each diagonal elemet of ρ is a beta distribution $B(80, 20)$ and the prior for $\widetilde{n} = \log(n - 2r_1)$ is a Normal $N(\ln(23.5 - 2r_1), 1.5)$.

[11]The code is written in C++ and runs in Rstudio, allowing for computations in parallel. It is available at http://www3.grips.ac.jp/~rlg/. It can be run in Amazon Web Services (AWS) machines using an AMI for Rstudio. The AWS machines c5d.4xlarge, using only 10 cores out of 16 to avoid memory bottlenecks, take 50% more of time than the Intel Xeon E5-2690, with a cost per hour of approximately \$0.26 per hour (spot price at 28 Sept 2018).

[12]MCSE values were calculated using the library LaplacesDemon in R (Hall, 2005) with the option of batch means (Jones et al. 2006 and Flegal et al. 2008).

2[13]. Our method, in contrast, produces a very sensible stationary pattern of volatility which closely resembles the spikes in the squared residuals.

Table 3 shows the proportion of the variance $var(y_t|y_{t-h})$ caused by the heteroscedastic shock u_{1t} for several key variables. For $h = 1$ the proportions due to u_{1t} for SP and the Federal Funds Rate (FFR) are (100%, 10%), whereas they become (72%, 17%) for $h = \infty$, respectively. In order to identify the monetary policy shock (ξ^{mon}) we use the same classification of slow and fast variables as in Banbura et al. (2010) and summarized in Table 1. We can see that ξ^{mon} affects only the homoscedastic components of y_t (y_t^{hom}), to the extent that the proportion of $var(y_t^{het}|y_{t-h})$ explained by ξ^{mon} is virtually 0 for all h. This is confirmed by the impulse response functions in Figure 3, where we can see that ξ^{mon} has the expected impact on GDP growth (GDP) and the 10 year bond yield (BOND), but only through the homoscedastic component. In contrast, Figure 3 also shows that GDP responds positively to a shock to SP[14] only through its heteroscedastic component (y_t^{het}). Interestingly, a shock to SP affects both the hetero and homo components of BOND, but with different signs, with the homo component increasing and the hetero decreasing when SP increases. The negative impact could be interpreted as the substitution effect between bonds and stocks in the highly volatile and heteroscedastic stock market, whereas the positive effect can be interpreted through the positive impact that the stock market has on real economic activity, and therefore on the interest rates.

Figure 4 plots the actual value of GDP growth (GDP), and its hetero (GDP^{het}) and homo (GDP^{hom}) components, showing what the economy would be if there were only hetero (u_{1t}) or homoscedastic shocks (u_{2t}). Although both GDP^{het} and GDP^{hom} have approximately the same average of 3%, we can see that GDP^{hom} has a smaller variance and becomes negative less often than GDP^{het}, indicating that an economy with only homoscedastic shocks would have higher welfare than one with only heteroscedastic shocks. Figure 4 also decomposes the structural monetary policy shock into the hetero and homo components, indicating that the hetero component became important in the early 1980s recession, which is when the FED raised significantly the interest rate following concerns on inflation. Overall, however, a variance decomposition shows that the structural monetary policy shock is homoscedastic, with only 3% of its variance caused by the heteroscedastic shocks u_{1t}.

Finally, Figure 5 plots the squared residuals (e_t^2), the squared heteroscedastic residuals ($(e_t^{het})^2$) and the squared homoscedastic residuals ($(e_t^{hom})^2$) for GDP and inflation (INF)[15]. We can see that the large spikes in the crises periods are captured only by $(e_t^{het})^2$, whereas $(e_t^{hom})^2$ does not have any noticeable abnormal behavior during the crises years.

---

[13]For the other variables that do not change position with the ordering, the estimated volatilies are the same under both orderings. The non-diagonal elements are assumed to remain constant with time as in Carriero et al (2015).

[14]The response to the SP shock is identified by assuming that SP is the fastest variable.

[15]Recall that Section 3 defined $e_t^{het} = H_1 e_t$, and $e_t^{hom} = H_2 e_t$, where $H_1 = U_1 U_1'$ and $H_2 = U_2 U_2' = I_r - H_1$

Variable	Transformation	Speed
Real gross domestic product	400 Δ log	Slow
Consumer price index	400 Δ log	Slow
Effective Federal funds rate	no transformation	
M2 money stock	400 Δ log	Fast
Personal income	400 Δ log	Slow
Real personal consumption expenditure	400 Δ log	Slow
Industrial production index	400 Δ log	Slow
Civilian unemployment rate	no transformation	Slow
Housing starts	log	Slow
Producer price index	400 Δ log	Slow
Personal consumption expenditures: chain-type price index	400 Δ log	Slow
Average hourly earnings: manufacturing	400 Δ log	Slow
M1 money stock	400 Δ log	Fast
10-Year Treasury constant maturity rate	no transformation	Fast
Real gross private domestic investment	400 Δ log	Slow
All employees: total nonfarm	400 Δ log	Slow
ISM manufacturing: PMI composite index	no transformation	Slow
ISM manufacturing: new orders index	no transformation	Slow
Business sector: real output per hour of all Persons	400 Δ log	Slow
Real stock prices (S&P 500 index divided by PCE index)	100 Δ log	Fast

Table 1: Definition of Macroeconomic variables in the large VAR and block identification assumption. Fast variables are assumed to react contemporaneously to a shock to the funds rate, whereas slow variables react only after a lag.

r_1	$p(\theta) + l(Y\|\theta)$	$l(Y\|\theta)$	$l(Y)$	BIC	Pred Lik	Pred BIC
0	-6950.3	-5974.6	-9353.6	-5975.7	-7177.7	-5660.0
1	-6989.7	-5982.5	-9303.2	-5984.8	-7169.5	-5567.0
2	-6877.9	-5938.2	-9215.9	-5941.7	-7126.2	-5531.0
3	-6814.1	-5932.3	-9168.5	-5936.9	-7120.2	-5530.4
4	-6710.4	-5904.3	-9097.4	-5910.2	-7090.3	-5507.8
5	-6637.5	-5893.4	-9045.7	-5900.4	-7078.6	-5502.5
6	-6563.7	-5892.3	-9002.3	-5900.5	-7075.9	-5508.8
7	-6501.1	-5892.7	-8961.3	-5902.0	-7073.8	-5514.0
8	-6451.2	-5898.2	-8925.2	-5908.7	-7076.9	-5526.0

Table 2: Model selection criteria for each value of r_1, with n estimated subject to $n > 2r_1$ (Macro data). $p(\theta)$ and $l(Y|\theta)$ denote the values of the log prior and log likelihood at the posterior mean, respectively. $l(Y)$ is the approximated marginal likelihood, Pred Lik is the predictive likelihood for observations 51 to 214, and Pred BIC is the BIC approximation to the predictive likelihood (calculated as the difference between the BIC with the full sample and the BIC with an initial sample). Numerical standard errors (NSE) for the log likelihood values were estimated using 100 independent replications of the particle filter, with $N = Tr_1$ particles. In all cases, the NSE values were smaller than 0.9.

	Mon Policy Shock (ξ_t)						Hetero shock (u_{1t})					
	1	2	3	4	5	∞	1	2	3	4	5	∞
GDP	0	0	4	4	4	4	88	75	64	61	60	55
GDP^{het}	0	0	0	0	0	0	100	100	100	100	100	100
GDP^{hom}	0	0	10	10	9	10	0	0	0	0	0	0
INF	0	3	3	3	3	10	62	64	61	59	59	31
INF^{het}	0	0	0	0	0	0	100	100	100	100	100	100
INF^{hom}	0	7	8	6	6	15	0	0	0	0	0	0
FFR	82	69	60	50	45	14	10	17	21	29	33	17
FFR^{het}	0	0	0	0	0	0	100	100	100	100	100	100
FFR^{hom}	90	82	75	70	67	17	0	0	0	0	0	0
BOND	15	17	16	16	17	9	23	24	24	26	27	16
BOND^{het}	0	0	0	0	0	0	100	100	100	100	100	100
BOND^{hom}	90	82	75	70	67	17	0	0	0	0	0	0
SP	0	0	1	2	2	3	100	96	91	87	85	72
SP^{het}	0	0	0	0	0	0	100	100	100	100	100	100
SP^{hom}	1	1	12	15	13	12	0	0	0	0	0	0

Table 3: Variance Decompositions. Percentage of $var(y_t|y_{t-h})$, $var(y_t^{het}|y_{t-h})$ and $var(y_t^{hom}|y_{t-h})$ caused by the monetary policy shock (ξ_t, left panel) and by the heteroscedastic shocks (u_{1t}, right panel), for $h = 1, 2, 3, 4, 5, \infty$.
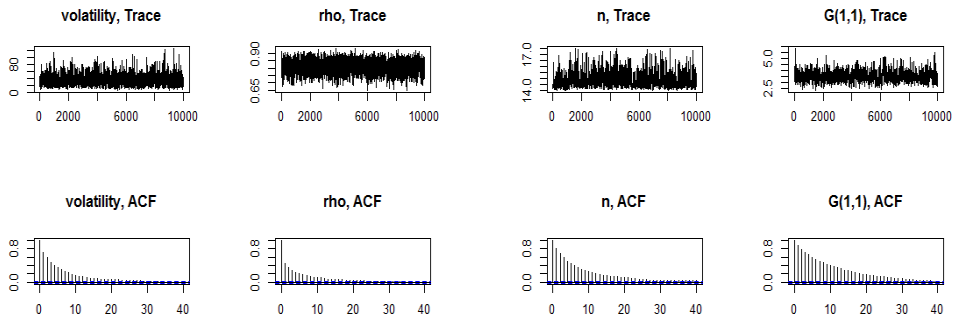

Figure 1: Trace plot and autocorrelations for the (7,7) element of K_{107}, the 4th diagonal element of ρ, for n and the (1,1) element of G (Macro data).
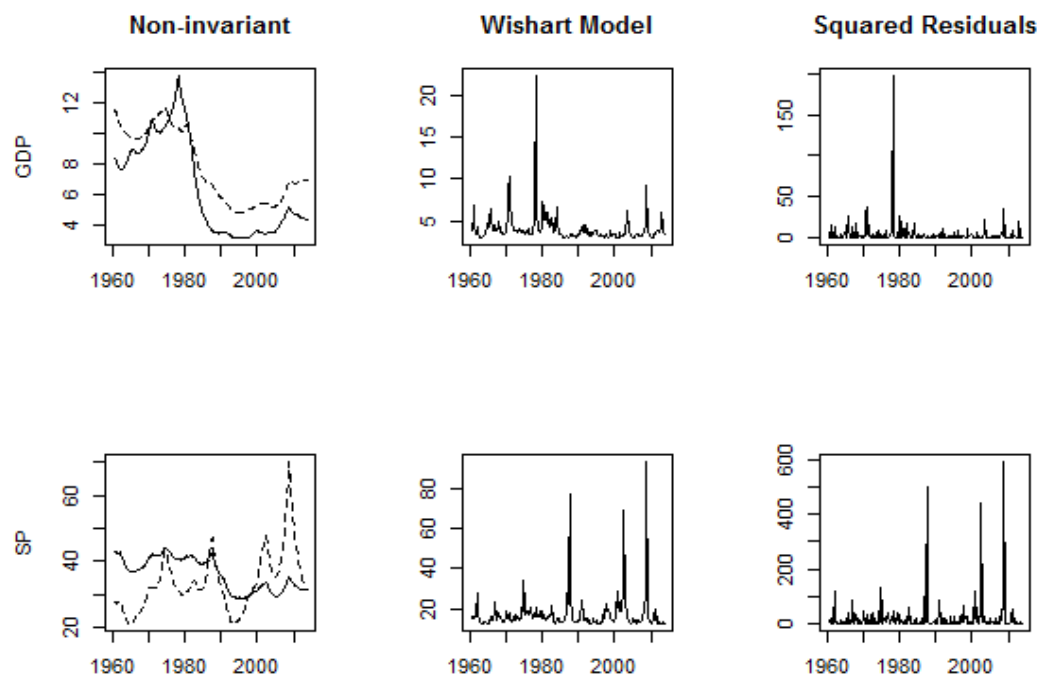
Figure 2: Estimated reduced form volatilities and squared Bayesian residuals for the real GDP growth rate (GDP) and the real stock returns (SP). The left panel corresponds to the method of Carriero et al. (2016) under two different orderings: the solid line is for ordering 1 and the dotted line is for ordering 2. The central panel corresponds to our method with $r_1 = 7$, and the right panel are the squared of the Bayesian residuals.
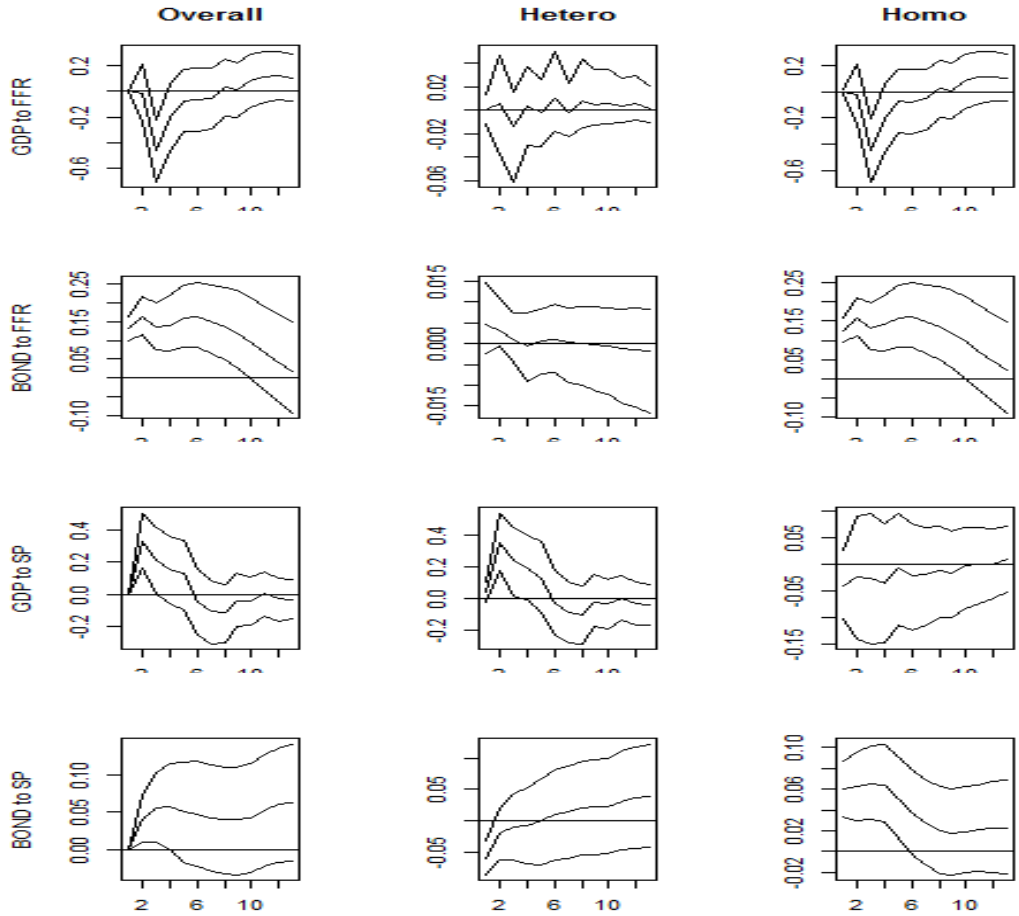
Figure 3: Impulse Responses of GDP growth (GDP) and the 10 year bond yield (BOND) to a monetary policy shock (FFR) and to a stock return shock (SP). The left panel plots the overall response, the central panel plots the reponse of the heteroscedastic component (y^{het}), and the right panel the homoscedastic component (y^{hom}). Posterior median and 90% credible interval.
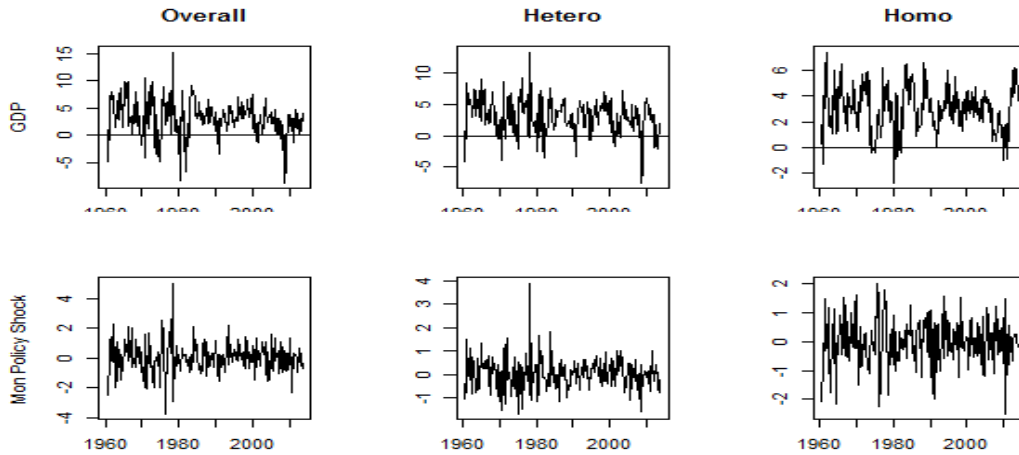

Figure 4: Decomposition of GDP growth and the estimated monetary policy shock into their heteroscedastic and homoscedastic components.
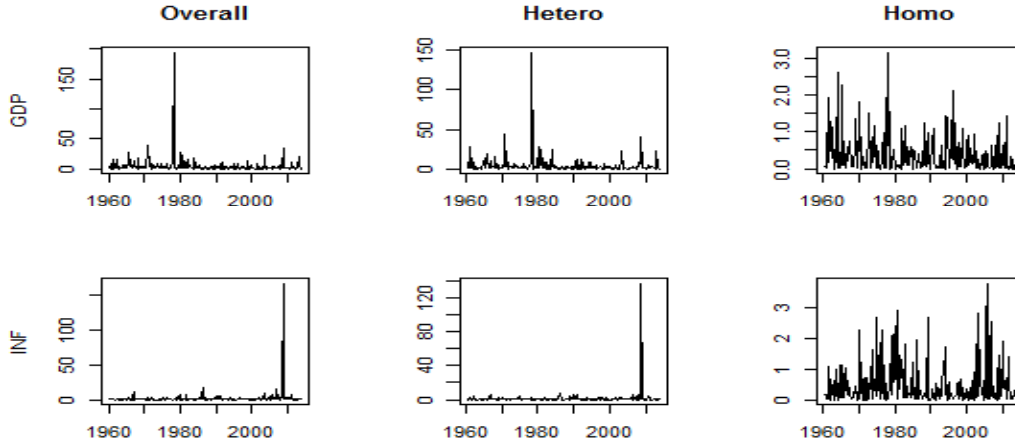
Figure 5: Plot of squared residuals e_t^2, squared heteroscedastic residuals $(e_t^{het})^2$ and squared homoscedastic residuals $(e_t^{hom})^2$ for GDP growth (GDP) and the inflation rate (INF).

6.2 Exchange Rates Data

We use data on 1043 daily returns of 5 exchange rates (AUD, CAD, EUR, JPY, SWF) with respect to the US dollar, beginning in January 2007 and ending in December 2010 (same data as in Chan et al. (2018)). Table 4 shows model selection criteria for models under the restriction $n > 2r_1$, and all indicate that the best model is $r_1 = 5$. Models under the restriction $n = r_1 + 2$ also suggested $r_1 = 5$, but had worse predictive likelihoods for observations $t = 501 - 1043$. Therefore we can conclude that these 5 exchange rates are not co-heteroscedastic. Figure 6 shows the posterior volatility estimates and the squared Bayesian residuals, where we can see that the exchange rates have idiosyncratic spikes, reinforcing the impression that they do not share heteroscedastic shocks. Figure 7 shows the trace plots and autocorrelation functions for $r_1 = 4$, indicating that autocorrelations vanish within 10 lags for most parameters (40 lags for ρ), and the mixing and convergence of the algorithm is good, taking about 2.8 hours to obtain 15000 iterations with 60 particles. The evaluation of the loglikelihood, using the average of 100 independent replications of the particle filter, each one with $Tr_1 = 4172$ particles, took 5.4 minutes, with a numerical standard error of 0.7.

However, the autocorrelations for the model $r_1 = 5$ were much more persistent, with the most persistent correlations being that of the (1,1) elements of ρ and $\overline{K} = (K_1 + ... + K_T)/T$, which become smaller than 10% only after 310 and 375 lags, respectively. Even in this case the MCSE values of ρ and $\overline{K}$ were only about 8% of the posterior standard deviation, using 15000 iterations with 50 particles, repeating the sampling of $L_{1:T}$ 30 times within each iteration, and using about 14 hours of computation time. Appendix II presents more details of how the number of particles affect the effective sample size of K_t. The evaluation of the loglikelihood for $r_1 = 5$ took about 11.3 minutes, with a numerical standard error of 0.7.

24

r_1	$l(Y\|\theta)$	$l(Y)$	BIC	Pred Lik	Pred BIC
0	-5239.7	-5398.9	-5239.7	-2919.4	-2904.0
1	-5058.4	-5197.9	-5065.4	-2894.7	-2875.9
2	-4990.9	-5114.8	-5001.3	-2885.8	-2850.4
3	-4943.5	-5017.7	-4957.4	-2872.4	-2851.5
4	-4909.3	-4936.6	-4926.7	-2868.8	-2853.6
5	-4659.8	-4682.1	-4680.6	-2755.4	-2711.5

Table 4: Model selection criteria for each value of r_1, with n estimated subject to $n > 2r_1$ (exchange rates data). Pred Lik is the predictive likelihood for observations 501 to 1043. Column labels as defined in Table 2.
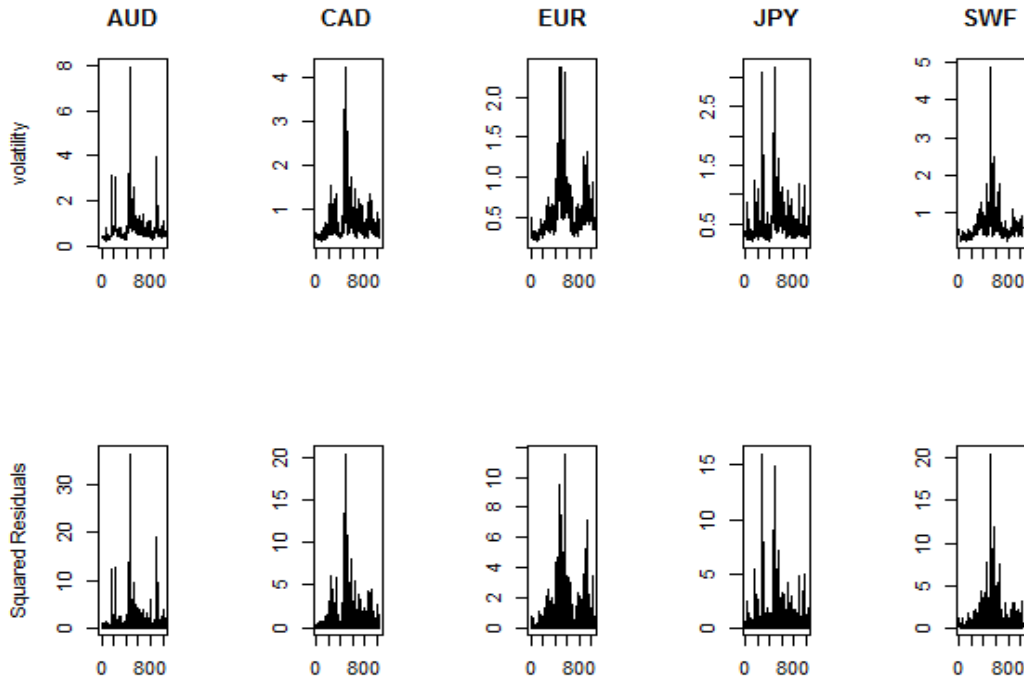

Figure 6: Posterior volatility estimates and squared Bayesian residuals for the returns of five exchange rates: AUD, CAD, EUR, JPY, SWF.
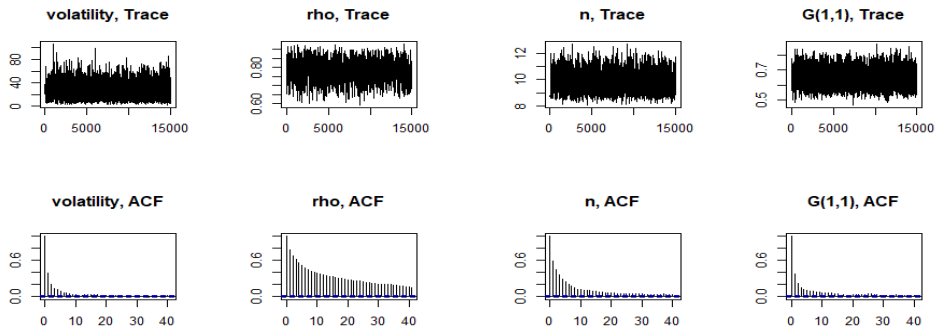

Figure 7: Trace plot and autocorrelations for the (4,4) element of K_{521}, the 4th diagonal element of ρ, n and the (1,1) element of G (exchange rate data) ($r_1 = 4$).

7 Concluding remarks

In this paper we have developed methods for Bayesian inference in an inverted Wishart process for Stochastic Volatility. Unlike most of the previous literature, the model is invariant to the ordering of the variables and allows for fat tails and the non-existence of higher moments. We provide a novel algorithm for posterior simulation which uses a pseudo posterior to define a more efficient particle filter, and to obtain a reparameterization that we find improves computational efficiency.

The modelling framework allows us to determine whether a set of variables share heteroscedastic shocks and are therefore co-heteroscedastic. Furthermore, the framework allows us to obtain new variance decompositions as well as new insights on the characteristics of the structural shock and its impact on the variables of interest. We find strong evidence of co-heteroscedasticity in an application to a large VAR of 20 macroeconomic variables, and provided a detailed analysis of how monetary policy and the stock market impact the heteroscedasticity of macroeconomic variables. We also illustrated the methodology using data on exchange rates, but found no evidence of co-heteroscedasticity.

Future research could look into the implications of co-heteroscedasticity for finding portfolio allocations with smaller risk and for decision making. Another possible venue is to find ways to reduce the number of free parameters in G through, for example, the use of Wishart graphical models (e.g. Dawid and Lauritzen (1993), Wang and West (2009)).

References

Andrieu, C., N.D. Freitas, A. Doucet and M.I. Jordan "An Introduction to MCMC for Machine Learning" *Machine Learning*, 50, 5–43.

Andrieu, C., A. Doucet, and R. Holenstein, (2010), "Particle Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, 269–342.

Anderson T.W. and Girshick, M.A. (1944) "Some Extensions of the Wishart Distribution," *The Annals of Mathematical Statistics*, 15, 345-357.

Asai, M. and M. McAleer (2009) "The Structure of Dynamic Correlations in Multivariate Stochastic Volatility Models," *Journal of Econometrics*, 150, 182-192.

Banbura, B., D. Giannone and L. Reichlin, (2010) "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25, 71-92.

Bauwens, L., M. Lubrano, and J.F. Richard (1999) *Bayesian Inference in Dynamic Econometric Models*. Oxford: Oxford University Press.

Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

Carriero, A., Clark, T.E. and Marcellino, M. (2015) "Large vector autoregressions with asymmetric priors and time varying volatilities," *Working Paper No. 759, School of Economics and Finance, Queen Mary University of London.*

Casarin, R. and D. Sartore, (2007), "Matrix-state particle filters for Wishart stochastic volatility processes," in *Proceedings SIS, 2007 Intermediate Conference, Risk and Prediction*, 399-409, CLEUP Padova.

Chan, J. (2013) " Moving average stochastic volatility models with application to inflation forecast," *Journal of Econometrics*, 176, 162-172

Chan, J. (2018) "Large Bayesian VARs: A Flexible Kronecker Error Covariance Structure," *Journal of Business & Economic Statistics* DOI: 10.1080/07350015.2018.1451336

Chan, J., R. Leon-Gonzalez, and R. Strachan (2018) "Invariant Inference and Efficient Computation in the Static Factor Model," *Journal of the American Statistical Association,* 113, 819-828.

Chib, S., I. Jeliazkov (2001) "Marginal Likelihood From the Metropolis–Hastings Output," *Journal of the American Statistical Association,* 96, 270-281.

Chiu, C.W., H. Mumtaz, and G. Pintér (2017) "Forecasting with VAR models: Fat tails and stochastic volatility," *International Journal of Forecasting*, 33, 1124-1143

Clark, T.E. and F. Ravazzolo (2015), "Macroeconomic Forecasting Performance Under Alternative Specifications of Time-Varying Volatility," *Journal of Applied Econometrics*, 30, 551–575.

Dawid, A.P. and S.L. Lauritzen (1993) "Hyper-Markov laws in the statistical analysis of decomposable graphical models," *The Annals of Statistics*, 3, 1272-1317.

Durbin, J. and S.J. Koopman (2001) *Time Series Analysis by State Space Methods,* Oxford University Press.

Engle, R. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation" *Econometrica*, 50, 987-1007.

Flegal, J.M., M. Haran and G.L. Jones, (2008). Markov Chain Monte Carlo: Can We Trust the Third Significant Figure? *Statistical Science* 23, 250-260.

Fox, E.B. and M. West (2011) "Autoregressive Models for Variance Matrices: Stationary Inverse Wishart Processes" arXiv:1107.5239.

Gleser, L.J. (1976) "A Canonical Representation for the Noncentral Wishart Distribution Useful for Simulation," *Journal of the American Statistical Association*, 71, 690-695.

Gordon, N., D. Salmond and A.F.M. Smith (1993). "Novel approach to nonlinear and non-Gaussian Bayesian state estimation," *Proc. Inst. Elect. Eng., F* 140, 107–113.

Gourieroux, C., J. Jasiak and R. Sufana (2009) "The Wishart Autoregressive process of multivariate stochastic volatility," *Journal of Econometrics*, 150, 167 - 181.

Guarniero, P., A.M. Johansen and A. Lee (2017) "The iterated auxiliary particle filter," *Journal of the American Statistical Association*, 112, 1636-1647.

Gupta, A.K. and D.K. Nagar (2000) *Matrix Variate Distributions,* Chapman & Hall/CRC.

Hamilton, J.D. (1994) *Time Series Analysis*, Princenton University Press, Princenton.

Hall, B. (2012) LaplacesDemon: Complete Environment for Bayesian Inference. R package version 12.10.01. URL http:http://cran.r-project.org/web/packages/LaplacesDemon/

Jones G., M. Haran, B. Caffo, R. Neath (2006). "Fixed-Width Output Analysis for Markov chain Monte Carlo," *Journal of the American Statistical Association*, 100, 1537–1547

Karapanagiotidis, P. (2012) "Improving Bayesian VAR Density Forecasts through Autoregressive Wishart Stochastic Volatility", *MPRA Paper* No. 38885.

Kim, S., N. Shephard, and S. Chib (1998) "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65, 361-393.

Koop, G. (2003), *Bayesian Econometrics*, Wiley.

Koop G., R. León-González and R. Strachan (2011), "Bayesian Inference in the Time-Varying Cointegration Model," *Journal of Econometrics*, 165, 210-220.

León-González, R. (2018), ""Efficient Bayesian Inference in Generalized Inverse Gamma Processes for Stochastic Volatility," forthcoming in *Econometric Reviews.*

Magnus, R. J. and Neudecker, H. (1999) *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley.

Muirhead, R.J. (2005) *Aspects of Multivariate Statistical Theory*, Wiley.

Papaspiliopoulos, O., G.O. Roberts and M.A. Sköld (2007) "General Framework for the Parametrization of Hierarchical Models" *Statistical Science*, 22, 59-73

Philipov, A. and M. E. Glickman (2006) "Multivariate stochastic volatility via Wishart processes," *Journal of Business and Economic Statistics* 24, 313–328.

Pitt, M.K. and N. Shephard (1999) "Analytic convergence rates and parameterization issues for the Gibbs sampler applied to state space models," *Journal of Time Series Analysis*, 20, 63-85.

Sims, C. and Zha, T., (1998) "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review* 39, 949-68.

Triantafyllopoulos, K. (2010) "Multi-variate stochastic volatility modelling using Wishart autoregressive processes" *Journal of Time Series Analysis*, 33, 48-60.

Uhlig, H. "Bayesian Vector Autoregressions with Stochastic Volatility," *Econometrica*, 65, 59-73.

Wang, H. and M. West (2009) "Bayesian analysis of matrix normal graphical models," *Biometrika*, 96, 821-834.

Whiteley, N. (2010) "Discussion of Particle Markov chain Monte Carlo methods" *Journal of the Royal Statistical Society* B, 72, 306–307.

Zha, T. (1999) "Block Recursion and Structural Vector Autoregressions," *Journal of Econometrics*, 90, 291-316.

Appendix I

This appendix is for section 4, and derives the pseudo posterior and the compact expression for the likelihood in (12). The prior $\pi(Z_{1:T}|\theta)$ is:

$$(2\pi)^{-Tnr/2} \left|I - \rho^2\right|^{n/2} \exp\left(-\frac{1}{2}tr\left(\sum_{t=2}^{T}\left(Z_t - Z_{t-1}\rho\right)'\left(Z_t - Z_{t-1}\rho\right) + \left(Z_1'Z_1\right)\left(I - \rho^2\right)\right)\right) \tag{21}$$

$$= (2\pi)^{-Tnr/2} \left|I - \rho^2\right|^{n/2} \exp\left(-\frac{1}{2}tr\left(\sum_{t=1}^{T-1}\left(\rho'Z_t'Z_t\rho\right) + \sum_{t=2}^{T}\left(Z_t'Z_t\right) - 2\sum_{t=2}^{T}\left(Z_t'Z_{t-1}\rho\right) + \left(Z_1'Z_1\right)\left(I - \rho^2\right)\right)\right)$$

Given the definition of Σ_t in (3), the inverse Σ_t^{-1} can be written as:

$$\Sigma_t^{-1} = \left(A^{-1}\right)'\begin{pmatrix} K_t & 0 \\ 0 & I \end{pmatrix} A^{-1} = B_1 K_t B_1' + B_2 B_2', \qquad \text{for } B = \left(A^{-1}\right)' = \begin{pmatrix} B_1 & B_2 \end{pmatrix} \tag{22}$$

where $K_t = Z_t'Z_t$. Thus, when we multiply (21) times the pseudo likelihood in (9) we obtain:

$$\pi(Z_{1:T}|\theta)\widetilde{L}(Y|\Sigma_{1:T},\beta) = \tag{23}$$

$$= \exp\left(-\frac{1}{2}tr\left(\sum_{t=1}^{T-1}\left(\rho'Z_t'Z_t\rho\right) + \sum_{t=2}^{T}\left(Z_t'Z_t\right)\left(I + B_1'e_te_t'B_1\right) - 2\sum_{t=2}^{T}\left(Z_t'Z_{t-1}\rho\right)\right)\right) \times$$

$$\exp\left(-\frac{1}{2}tr\left(\left(Z_1'Z_1 + B_1'e_1e_1'B_1\right)\left(I - \rho^2\right)\right)\right)(2\pi)^{-Tnr/2}\left|I - \rho^2\right|^{n/2}c_{\widetilde{B}}$$

where:

$$c_{\widetilde{B}} = \exp\left(-\frac{1}{2}\sum_{t=1}^{T}tr\left(B_2B_2'e_te_t'\right)\right)$$

The terms that depend on Z_T in (23) are $\exp(-0.5tr((Z_T'Z_T)\left(I + B_1'e_Te_T'B_1\right) - 2\left(Z_T'Z_{T-1}\rho\right)))$ and so $Z_T|Z_{1:T-1}$ is a normal with mean equal to $Z_{T-1}\rho V_T$ and variance $V_T = (I + B_1'e_Te_T'B_1)^{-1}$. Integrating out Z_T from (23) we obtain:

$$\int \pi(Z_{1:T}|\theta)\widetilde{L}(Y|\Sigma_{1:T},\beta)dZ_T$$

$$= (2\pi)^{nr/2}\left|V_T\right|^{n/2}\exp(\frac{1}{2}tr(\rho'Z_{T-1}'Z_{T-1}\rho V_T)) \times$$

$$\exp\left(-\frac{1}{2}tr\left(\sum_{t=1}^{T-1}\left(\rho'Z_t'Z_t\rho\right) + \sum_{t=2}^{T-1}\left(Z_t'Z_t\right)\left(I + B_1'e_te_t'B_1\right) - 2\sum_{t=2}^{T-1}\left(Z_t'Z_{t-1}\rho\right)\right)\right) \times$$

$$\exp\left(-\frac{1}{2}tr\left(\left(Z_1'Z_1 + B_1'e_1e_1'B_1\right)\left(I - \rho^2\right)\right)\right)(2\pi)^{-Tnr/2}\left|I - \rho^2\right|^{n/2}c_{\widetilde{B}}$$

The terms that depend on Z_{T-1} are $\exp(-0.5tr((Z_{T-1}'Z_{T-1})\left(I + B_1'e_{T-1}e_{T-1}'B_1 + \rho\rho'\right) - 2\left(Z_{T-1}'Z_{T-2}\rho\right) - (Z_{T-1}'Z_{T-1}\rho V_T\rho')))$ and so $Z_{T-1}|Z_{1:T-2}$ is a normal with mean equal to $Z_{T-2}\rho V_{T-1}$ and variance $V_{T-1} = (I + B_1'e_{T-1}e_{T-1}'B_1 + \rho(I - V_T)\rho')^{-1}$. This process can be repeated recursively to find the densities described in (10) and (11) and also the following integrating constant:

$$\int \pi(Z_{1:T}|\theta)\widetilde{L}(Y|\Sigma_{1:T},\beta) = \left|I - \rho^2\right|^{n/2}\widetilde{c}_B\prod_{t=1}^{T}\left|V_t\right|^{n/2} = \widetilde{c}$$

which allows us to write the pseudo posterior as:

$$\widetilde{\pi}(Z_{1:T}|Y,\theta) = \frac{\pi(Z_{1:T}|\theta)\widetilde{L}(Y|\Sigma_{1:T},\beta)}{\widetilde{c}}$$

Then we can write the likelihood as:

$$
\begin{aligned}
L(Y|\beta, \theta) &= \int \left(\prod_{t=1}^{T} |\Sigma_t|^{-1/2} \right) \pi(Z_{1:T}|\theta) \widetilde{L}(Y|\Sigma_{1:T}, \beta) dZ_{1:T} \\
&= \int \left(\prod_{t=1}^{T} |\Sigma_t^{-1}|^{1/2} \right) \widetilde{c}\widetilde{\pi}(Z_{1:T}|Y, \theta) dZ_{1:T} \\
&= \int \left(\prod_{t=1}^{T} |K_t|^{1/2} \right) \left(|A_1 A_1' + A_2 A_2'|^{-T/2} \right) \widetilde{c}\widetilde{\pi}(Z_{1:T}|Y, \theta) dZ_{1:T} = \widetilde{c}_L E_{\widetilde{\pi}} \left(\prod_{t=1}^{T} |K_t|^{1/2} \right)
\end{aligned}
$$

which is equal to (12).

Appendix II

This appendix is for Section 5.2. It derives the reverse order representation in (15) - (16) and outlines the algorithm to obtain a draw of $\widetilde{L}_{1:T}$. It also provides a simulation to illustrate the computational gains of alternating the order.

To prove the connection between (L_{1t}, K_{2t}) and $(\widetilde{L}_{1t}, \widetilde{K}_{2t})$ in (15), we obtain (L_{1t}, K_{2t}) by first defining the semi-orthogonal matrix $\overrightarrow{Z}_t = Z_t(K_t)^{-1/2}$, with $K_t = Z_t' Z_t$ and its orthogonal complement $\overrightarrow{Z}_{t\perp} : n \times (n - r_1)$ such that:

$$
\overrightarrow{Z}_t' \overrightarrow{Z}_t = I_{r_1}, \quad \overrightarrow{Z}_{t\perp}' \overrightarrow{Z}_{t\perp} = I_{n-r_1}, \quad \overrightarrow{Z}_t' \overrightarrow{Z}_{t\perp} = 0
$$

And transform from Z_t to (L_{1t}, L_{2t}), where $L_{1t} = \overrightarrow{Z}_{t-1}' Z_t$, $L_{2t} = \overrightarrow{Z}_{t-1,\perp}' Z_t$, such that $L_{1t} : r_1 \times r_1$ is independent of $L_{2t} : (n - r_1) \times r_1$. Then the process $Z_t = Z_{t-1}\rho + \varepsilon_t$ implies that:

$$
L_{1t} = \overrightarrow{Z}_{t-1}' Z_t = \underbrace{\left(\overrightarrow{Z}_{t-1} \right)' Z_{t-1}\rho}_{K_{t-1}^{1/2}} + \underbrace{\left(\overrightarrow{Z}_{t-1} \right)' \varepsilon_t}_{\varepsilon_{1t}} \tag{24}
$$

$$
L_{1t} = K_{t-1}^{1/2}\rho + \varepsilon_{1t}, \quad \varepsilon_{1t} : r_1 \times r_1, \quad vec(\varepsilon_{1t}) \sim N(0, I_{r_1} \otimes I_{r_1}), \, t > 1
$$

and that:

$$
L_{2t} = \overrightarrow{Z}_{t-1,\perp}' Z_t = \underbrace{\left(\overrightarrow{Z}_{t-1,\perp} \right)' Z_{t-1}\rho}_{0} + \left(\overrightarrow{Z}_{t-1,\perp} \right)' \varepsilon_t = \underbrace{\left(\overrightarrow{Z}_{t-1,\perp} \right)' \varepsilon_t}_{\varepsilon_{2t}}
$$

$$
L_{2t} = \varepsilon_{2t}, \quad vec(L_{2t}) \sim N(0, I_{r_1} \otimes I_{n-r_1})
$$

$$
K_{2t} = L_{2t}' L_{2t} \sim W_r(n - r_1, I_{r_1})
$$

Note that the process $Z_t = Z_{t-1}\rho + \varepsilon_t$ with Z_1 following the stationary distribution can be equiv-

alently written as $Z_{t-1} = Z_t\rho + \varepsilon_t$ with Z_T following the stationary distribution[16]. Therefore we can apply the same decomposition to the reverse process as follows:

$$\widetilde{L}_{1,t-1} = \overrightarrow{Z}'_t Z_{t-1} = \underbrace{\left(\overrightarrow{Z}_t\right)' Z_t\rho}_{K_t^{1/2}} + \underbrace{\left(\overrightarrow{Z}_t\right)' \varepsilon_t}_{\varepsilon_{1t}} \tag{25}$$

$$\widetilde{L}_{2,t-1} = \overrightarrow{Z}'_{t\perp} Z_{t-1} = \underbrace{\left(\overrightarrow{Z}_{t\perp}\right)' Z_t\rho}_{0} + \left(\overrightarrow{Z}_{t\perp}\right)' \varepsilon_t = \underbrace{\left(\overrightarrow{Z}_{t-1,\perp}\right)' \varepsilon_t}_{\varepsilon_{2t}}$$

$$\widetilde{K}_{2t} = \widetilde{L}'_{2t}\widetilde{L}_{2t}$$

Note that $L_{1t} = \overrightarrow{Z}'_{t-1} Z_t = (K_{t-1}^{-1/2})' Z'_{t-1} Z_t$, while $\widetilde{L}_{1,t-1} = \overrightarrow{Z}'_t Z_{t-1} = (K_t^{-1/2})' Z'_t Z_{t-1}$, and therefore we can conclude that $Z'_t Z_{t-1} = (K_t^{1/2})' \widetilde{L}_{1,t-1} = ((K_{t-1}^{1/2})' L_{1t})'$, from where we can arrive at (15) and the inverse transformation (19).

The algorithm to generate a value for $\widetilde{L}_{1:T}$ in reverse order is analogous to the natural order algorithm in Section 5.1. Let $\widetilde{L}_t = (\widetilde{L}_{1t}, \widetilde{K}_{2t})$ for $t < T$, $\widetilde{L}_T = K_T$, and $\widetilde{L}_{1:T} = (\widetilde{L}_1, \widetilde{L}_2, ..., \widetilde{L}_T)$. Let the N particles be denoted as $\widetilde{L}_t^k = (\widetilde{L}_{1t}^k, \widetilde{K}_{2t}^k)$ for $t < T$ and $\widetilde{L}_T^k = (K_T^k)$. Define $K_t^k = (\widetilde{L}_{1t}^k)' \widetilde{L}_{1t}^k + \widetilde{K}_{2t}^k$ for $t < T$. Given the definition of $\widetilde{V}_t$ in (17), the value of $\widetilde{L}_{1:T}$ at iteration i, denoted as $\widetilde{L}_{1:T}(i) = (\widetilde{L}_1(i), ..., \widetilde{L}_T(i))$, can be generated given the previous value $\widetilde{L}_{1:T}(i-1)$ as follows:

Algorithm 4 *Reverse order cPFBS.*

Step 1: Fix the last particle equal to $\widetilde{L}_{1:T}(i-1)$, that is, $\widetilde{L}_{1:T}^N = \widetilde{L}_{1:T}(i-1)$

 Step 2: at time $t = T$,

 (a) sample $K_T^k \sim W_{r_1}(n, \widetilde{V}_T)$ for $k = 1, ..., N-1$, and

 (b) compute and normalize the weights:

$$w_T^k := \left|K_T^k\right|^{0.5}, \ W_T^k := w_T^k / (\sum_{m=1}^{N} w_T^m), \ k = 1, ..., N$$

 Step 3: at times $t = T-1, ..., 1$,

 (a) sample the indices A_{t+1}^k, for $k = 1, ..., N-1$, from a multinomial distribution on $(1, ..., N)$ with probabilities $\mathbf{W}_{t+1} = (W_{t+1}^1, ..., W_{t+1}^N)$

---

[16]This can easily be proved from $Z_t = Z_{t-1}\rho + \varepsilon_t$ by writing $(Z_1, ..., Z_T)$ as a vector $Z = (vec(Z_1)', vec(Z_2)', ..., vec(Z_T)')'$, and noting that the joint distribution of Z is normal with 0 mean and covariance V :

$$V = \begin{pmatrix} (I-\rho^2)^{-1} \otimes I_n & (I-\rho^2)^{-1}\rho \otimes I_n & \cdots & (I-\rho^2)^{-1}\rho^{T-1} \otimes I_n \\ \rho'(I-\rho^2)^{-1} \otimes I_n & (I-\rho^2)^{-1} \otimes I_n & & (I-\rho^2)^{-1}\rho^{T-2} \otimes I_n \\ \\ \rho^{(T-1)'}(I-\rho^2)^{-1} \otimes I_n & \rho^{(T-2)'}(I-\rho^2)^{-1} \otimes I_n & & (I-\rho^2)^{-1} \otimes I_n \end{pmatrix}$$

and also the distribution of $\widetilde{Z} = (vec(Z_T)', vec(Z_{T-1})', ..., vec(Z_1)')'$ is normal with the same 0 mean and covariance matrix V, and so we can also write $Z_t = Z_{t+1}\rho + \varepsilon_t$ provided Z_T comes from the stationary distribution.

(b) sample $vec(\widetilde{L}_{1t}^k) \sim N\left(\left(K_{t+1}^{A_{t+1}^k}\right)^{1/2}\rho\widetilde{V}_t, I_{r_1}\otimes\widetilde{V}_t\right)$ and $\widetilde{K}_{2t}^k \sim W_{r_1}(n-r_1,\widetilde{V}_t)$, calculate $K_t^k = (\widetilde{L}_{1t}^k)'\widetilde{L}_{1t}^k + \widetilde{K}_{2t}^k$ for $k = 1, ..., N-1$ and

(c) compute and normalize the weights

$$w_t^k := \left|K_t^k\right|^{0.5}, \quad W_t^k := w_t^k/(\sum_{m=1}^{N} w_t^m), \ k = 1, ..., N$$

Step 4: at time $t = 1$, sample b_1 from a multinomial distribution on $(1, ..., N)$ with probabilities $\mathbf{W}_1$, and set $\widetilde{L}_1(i) = \widetilde{L}_1^{b_1}$.

Step 5: at times $t = 2, ..., T$

(a) compute the updated weights

$$\widetilde{w}_t^k = w_t^k f(K_{t-1}^{b_{t-1}}|K_t^k), ..., \widetilde{W}_t^k := \widetilde{w}_t^k/(\sum_{m=1}^{N} \widetilde{w}_t^m), \ k = 1, ..., N, \ where$$

$$\mu_t^k = (K_t^k)^{1/2}\rho\widetilde{V}_{t-1}, \quad f(K_{t-1}^{b_{t-1}}|K_t^k) = \exp(-\frac{1}{2}tr(\widetilde{V}_{t-1}^{-1}(\widetilde{L}_{1(t-1)}^{b_{t-1}} - \mu_t^k)'(\widetilde{L}_{1(t-1)}^{b_{t-1}} - \mu_t^k))$$

(b) sample b_t from a multinomial distribution on $(1, ..., N)$ with probabilities $\widetilde{\mathbf{W}}_t = (\widetilde{W}_t^1, ..., \widetilde{W}_t^N)$, and set $\widetilde{L}_t(i) = \widetilde{L}_t^{b_t}$.

In order to see the impact of alternating the order, we fix (G, β) equal to their posterior means and (n, ρ) equal to their posterior medians and calculate the Effective Sample Size (ESS) of the (1,1) element of K_t for 10000 iterations using the Macro data with $r_1 = 7$ (Table 5) and the Exchange Rate Data with r_1 equal to 4 and 5 (Table 6). The ESS values are not adjusted by computing time because the differences in computing time among algorithms are negligible. We can see that the ESS values of K_T are always bigger (smaller) than the ESS values of K_1 in the natural order (reverse order) algorithm, respectively. The alternating order algorithm has always better ESS than the natural order for K_1, and better ESS than the reverse order for K_T. The gains imply that ESS is increased at least by a factor of 3 (with Exchange rate data, 95 particles and $r_1 = 4$) and at most by a factor of 260 (with Exchange rates data, 95 particles and $r_1 = 5$). However, the alternating order algorithm has lower ESS than the natural order for K_T, and lower ESS than the reverse order for K_1, but the factor of increase of the single order algorithms over the alternating algorithm is only between 1.4 and 2.9. Looking at the ESS of $K_{T/2}$ (i.e. at the middle of the sample), the alternating order algorithm has always better ESS when using the Macro data (Table 5), with the factor of improvement ranging between 1.01 (with 110 particles) and 1.55 (with 25 particles). With the exchange rates data the alternating order algorithm is always better than the natural order to sample $K_{T/2}$, but the reverse order algorithm is even better in 5 out of 8 cases. To get a measure of which algorithm is overall better to sample $K_{1:T}$, we calculate the

N	order	K_1	$K_{T/2}$	K_T	$\overline{K}$	Gain (%)	
	-1	7853	107	86	91		
25	0	3764	124	3938	103	14	31
	1	88	80	7587	79		
	-1	8661	174	332	231		
50	0	4640	272	4581	371	61	32
	1	321	258	8762	281		
	-1	9222	579	558	586		
80	0	4929	654	4971	801	37	33
	1	709	475	9289	601		
	-1	9843	880	578	1052		
110	0	5325	988	5357	1539	46	49
	1	1226	974	9427	1030		

Table 5: Effective sample size of K with 10000 iterations and $r_1 = 7$ using the Macro data. The column with label 'order' takes values (-1,0,1) for reverse, alternating and natural order, respectively. N is the number of particles and the columns (K_1, $K_{T/2}$, K_T,$\bar{K}$) give the Effective Sample Size (ESS) with 10000 iterations of sampling the (1,1) element of (K_1, $K_{T/2}$, K_T,$\overline{K}$), where $\bar{K} = (K_1 + ... + K_T)/T$. The two columns under 'Gain (%)' give the percentage increase in the ESS of the (1,1) element of $\overline{K}$ when using alternating order with respect to reverse order (left) and natural order (right). The parameters (G, β) were fixed equal to their posterior means and (n, ρ) equal to their posterior medians.

ESS of $\overline{K} = (K_1 + ... + K_T)/T$. We find that the alternating order algorithm is the best in all but one case, increasing the ESS by at most 140% and at least by 1%. The only exception is with the exchange rate data with $r_1 = 5$ and 95 particles, where the ESS of the reverse order algorithm is 42% better. Overall we can conclude that the ordering in sampling $K_{1:T}$ affects the efficiency of the algorithm, and that for our datasets alternating the ordering brings almost always computational gains.

Appendix III

This appendix is for Section 5.3 and it gives details on the reparameterization function $\epsilon = f_\theta(L_{1:T})$, its inverse $f_\theta^{-1}(\epsilon)$, the density $\pi^t(\theta|\epsilon, \beta)$, and a simulation to illustrate the better performance of the reparameterization.

Defining $\epsilon = (a_{2:T}, p_{1:T}, c_{1:T})$, the following algorithm can be used to calculate $\epsilon = f_\theta(L_{1:T})$.

Algorithm 5 *To obtain $\epsilon = f_\theta(L_{1:T})$.*

Step 1: at time $t = 1$,

(a) Define $\widehat{K}_1 = (V_1^{-1/2})' K_1 (V_1^{-1/2})$, and Δ_1 as an upper triangular matrix such that $\widehat{K}_1 = \Delta_1' \Delta_1$, with c_1 being the off-diagonal elements of Δ_1.

(b) Let $d_1 = (d_{11}, d_{21}, ..., d_{r_1 1})$ be the diagonal elements of Δ_1. Let $p_1 = (p_{11}, p_{21}, ..., p_{r_1 1})$, with $p_{i1} = F_{\chi^2(f_{i1}^n)}(d_{i1}^2)$, where $F_{\chi^2(f_{i1}^n)}(d_{i1}^2)$ is the distribution function of a chi-squared with $f_{i1}^n = n + i + 1$ degrees of freedom.

Step 2: at times $t = 2, ..., T$,

r_1	N	order	K_1	$K_{T/2}$	K_T	$\overline{K}$	Gain (%)	
		-1	4755	1277	436	127		
	25	0	1651	1353	3078	240	88	48
		1	92	1251	5608	161		
		-1	6482	2425	886	337		
	50	0	3428	2414	3713	345	2	37
4		1	228	2272	6983	251		
		-1	7141	4358	1599	516		
	95	0	4208	4243	5228	627	21	23
		1	340	3925	7303	509		
		-1	7170	5364	1574	645		
	125	0	3971	4775	5953	660	2	1
		1	404	3227	7745	655		
		-1	6627	372	93	33		
	25	0	3348	405	3875	37	12	188
		1	29	277	7180	13		
		-1	8463	1044	177	64		
	50	0	4100	958	4562	88	37	140
5		1	47	569	8202	36		
		-1	8875	2281	580	146		
	95	0	4461	2357	4900	85	-42	25
		1	17	1515	9256	68		
		-1	8675	2925	794	134		
	125	0	4709	2356	5201	171	28	27
		1	44	1811	9655	135		

Table 6: Effective sample size of K with 10000 iterations and $r_1 = 4$ and 5, using the Exchange Rate data. Column labels have the same meaning as in Table 5. The parameters (G, β) were fixed equal to their posterior means and (n, ρ) equal to their posterios medians.

(a) Calculate a_t as:

$$a_t = (L_{1t} - (K_{t-1})^{1/2}\rho V_t)V_t^{-1/2}, \quad \text{for } t > 1$$

(b) Calculate $\widehat{K}_{2t} = (V_t^{-1/2})'K_{2t}(V_t^{-1/2})$, and Δ_t as an upper triangular matrix such that $\widehat{K}_{2t} = \Delta_t'\Delta_t$, with c_t being the off-diagonal elements of Δ_t.

(c) Let $d_t = (d_{1t}, d_{2t}, ..., d_{r_1t})$ be the diagonal elements of Δ_t. Let $p_t = (p_{1t}, p_{2t}, ..., p_{r_1t})$, with $p_{it} = F_{\chi^2(f_{it}^n)}(d_{it}^2)$, where $F_{\chi^2(f_{it}^n)}(d_{it}^2)$ is the distribution function of a chi-squared with $f_{it}^n = n - r_1 + i + 1$ degrees of freedom.

The inverse transformation is denoted as $L_{1:T}^{\theta^*} = f_{\theta^*}^{-1}(\epsilon)$, for $\theta^* = (G^*, \rho^*, n^*)$, where we write $L_{1:T}^{\theta^*}$ instead of $L_{1:T}$ to make clear that for a fixed value of ϵ, $L_{1:T}$ changes when θ changes. For this reason we also use the notation $(K_t^\theta, K_{2t}^\theta, L_{1t}^\theta, V_t^\theta)$ for $(K_t, K_{2t}, L_{1t}, V_t)$ below. The following algorithm describes how to obtain $L_{1:T}^{\theta^*} = f_{\theta^*}^{-1}(\epsilon)$.

Algorithm 6 To obtain $L_{1:T}^{\theta^*} = f_{\theta^*}^{-1}(\epsilon)$

Step 1: Calculate $(V_1^{\theta^*}, V_2^{\theta^*}, ..., V_T^{\theta^*})$ using the recursion in (11) and the value of $\theta^* = (G^*, \rho^*, n^*)$.

Step 2: at time $t = 1$

(a) Obtain the vector $d_t^{n^*} = (d_{1t}^{n^*}, d_{2t}^{n^*}, ..., d_{r_1t}^{n^*})$, by calculating $d_{it}^{n^*} = \left(F_{\chi^2(f_{it}^{n^*})}^{-1}(p_{it}) \right)^{1/2}$, where $F_{\chi^2(f_{it}^{n^*})}^{-1}$ is the inverse of the distribution function of a chi-squared with $f_{it}^{n^*} = n^* + i + 1$ degrees of freedom.

(b) Construct $\Delta_t^{\theta^*}$ as an upper triangular matrix with off diagonal elements equal to c_t and diagonal equal to $d_{it}^{n^*}$.

(c) Calculate $K_1^{\theta^*} = \left((V_t^{\theta^*})^{1/2}\right)' \left(\Delta_t^{\theta^*}\right)' \Delta_t^{\theta^*}(V_t^{\theta^*})^{1/2}$

Step 3: at times $t = 2, ..., T$

(a) Calculate

$$L_{1t}^{\theta^*} = (K_{t-1}^{\theta^*})^{1/2}\rho^*V_t^{\theta^*} + a_t(V_t^{\theta^*})^{1/2}$$

(b) Obtain the vector $d_t^{n^*} = (d_{1t}^{n^*}, d_{2t}^{n^*}, ..., d_{r_1t}^{n^*})$, by calculating $d_{it}^{n^*} = \left(F_{\chi^2(f_{it}^{n^*})}^{-1}(p_{it}) \right)^{1/2}$, where $F_{\chi^2(f_{it}^{n^*})}^{-1}$ is the inverse of the distribution function of a chi-squared with $f_{it}^{n^*} = n^* - r_1 + i + 1$ degrees of freedom.

(c) Construct $\Delta_t^{\theta^*}$ as an upper triangular matrix with off diagonal elements equal to c_t and with diagonal equal to $d_t^{n^*}$

(d) Calculate $K_{2t}^{\theta^*} = \left((V_t^{\theta^*})^{1/2}\right)' \left(\Delta_t^{\theta^*}\right)' \Delta_t^{\theta^*}(V_t^{\theta^*})^{1/2}$.

The following proposition gives the conditional posterior density of θ given (ϵ, β).

Proposition 7 *The conditional posterior density of θ given (ϵ, β), denoted by $\pi^t(\theta|\epsilon, \beta)$, is such that:*

$$\pi^t(\theta|\epsilon, \beta) \propto \pi(L^\theta_{1:T}|\theta) L(Y|K^\theta_{1:T}, \beta) J_\varepsilon \pi(\theta) \pi(\beta|\theta) \tag{26}$$

where

$$
\begin{aligned}
L(Y|K^\theta_{1:T}, \beta) &= (2\pi)^{-Tr/2} \left(\prod_{t=1}^{T} |\Sigma_t^{-1}|^{1/2} \right) \exp\left(-\frac{1}{2} \sum_{t=1}^{T} tr\left(\Sigma_t^{-1} e_t e_t' \right) \right), \\
\text{with } e_t &= y_t - \beta x_t, \quad \Sigma_t^{-1} = B_1 K_t^\theta B_1' + B_2 B_2' \\
\pi(L^\theta_{1:T}|\theta) &= \left| I_{r_1} - \rho^2 \right|^{n/2} \exp\left(-\frac{1}{2} tr\left(\left(I_{r_1} - \rho^2 \right) K_1^\theta + M \right) \right) |K_1^\theta|^{(n-r_1-1)/2} \times \\
&\quad \prod_{t=2}^{T} |K_{2t}|^{(n-r_1-(r_1+1))/2} \frac{2^{-\frac{((T-1)(n-r_1)r_1)}{2}}}{(\Gamma_{r_1}((n-r_1)/2))^{T-1}} \frac{2^{-\frac{nr_1}{2}}}{(\Gamma_{r_1}(n/2))} (2\pi)^{-\frac{(T-1)r_1^2}{2}} \\
M &= \sum_{t=2}^{T} K_t^\theta + \sum_{t=2}^{T} \rho' K_{t-1}^\theta \rho - 2\rho' \sum_{t=2}^{T} ((K_{t-1}^\theta)^{1/2})' L_{1t}^\theta \\
J_\varepsilon &= (V_1^\theta)^{(r_1+1)/2} \left(\prod_{t=2}^{T} (V_t^\theta)^{(r_1+1)/2 + r_1/2} \right) \left(\prod_{t=2}^{T} \prod_{i=1}^{r_1} \left(\frac{(d_{it}^n)^{(r_1-i)}}{F'_{\chi^2(f_{it}^n)}((d_{it}^n)^2)} \right) \right)
\end{aligned}
$$

where for $t > 1$, $d_t^n = (d_{1t}^n, ..., d_{r_1 t}^n)$ is the diagonal of the Cholesky decomposition of $\widehat{K}_{2t} = (V_t^{-1/2})' K_{2t} (V_t^{-1/2})$, whereas for $t = 1$, $d_t^n = (d_{1t}^n, ..., d_{r_1 t}^n)$ is the diagonal of the Cholesky decomposition of $\widehat{K}_1 = (V_1^{-1/2})' K_1 (V_1^{-1/2})$. $F'_{\chi^2(f_{it}^n)}(d_{it}^2)$ is the density function of a χ^2 distribution evaluated at d_{it}^2 with degrees of freedom $f_{it}^n = n - r_1 + i + 1$ for $t > 1$, and $f_{it}^n = n + i + 1$ for $t = 1$. $\pi(\theta)$ is the prior of θ, and $\pi(\beta|\theta)$ is the conditional prior of β given θ.

Proof. $L(Y|K^\theta_{1:T}, \beta)$ is the likelihood given $K^\theta_{1:T}$, $\pi(L^\theta_{1:T}|\theta)$ is the prior of $L^\theta_{1:T}$, and J_ε is the Jacobian of the transformation from $L_{1:T}$ to ε. $\pi(L^\theta_{1:T}|\theta)$ can be calculated as:

$$\pi(L^\theta_{1:T}|\theta) = \pi(K_1^\theta|\theta) \prod_{t=2}^{T} \pi(L^\theta_{1t}|\theta, K^\theta_{t-1}) \pi(K^\theta_{2t}|\theta)$$

and then use the expressions for the multivariate normal and Wishart densities with parameters specified as in (14). To calculate the Jacobian J_ε note that the Jacobian from K_{2t} to $\widehat{K}_{2t}$ (or from K_1 to $\widehat{K}_1$) is $(V_t^\theta)^{(r_1+1)/2}$ and the Jacobian from $\widehat{K}_{2t}$ to its Cholesky decomposition is $\prod_{i=1}^{r_1} \left((d_{it}^n)^{(r_1-i+1)} \right)$. The Jacobian from d_{it}^n to p_{it} is $\left((d_{it}^n) F'_{\chi^2(f_{it}^n)}((d_{it}^n)^2) \right)^{-1}$, and the Jacobian from L_{1t} to a_t is $(V_t^\theta)^{r_1/2}$. ∎

In our estimations we use a normal inverse Wishart prior for G, and a normal prior for $\tilde{n} = \log(n - 2r_1)$ and a beta distribution for ρ. However, in order to perform the Metropolis step we target the

conditional posterior of $(G, \tilde{n}, \tilde{\rho})$, where $\tilde{\rho} = \ln(-\ln(1 - \rho^2))$. The prior for $\pi(\tilde{\rho})$ can be obtained as:

$$\pi(\tilde{\rho}) = \pi(\rho) J_\rho$$
$$J_\rho = \frac{1 - \rho^2}{2\rho}(-\ln(1 - \rho^2))$$

where J_ρ is the Jacobian. As a proposal density we use an inverse Wishart for G centered at $((1 - \tau_G)G(i - 1) + \tau_G \hat{G})$, where $G(i - 1)$ is the value of G in the previous iteration and $\hat{G}$ is a preliminary estimate of G. For $\varrho = (\tilde{n}, \tilde{\rho})$ we use a normal proposal density centered at $((1 - \tau_\rho)\varrho(i - 1) + \tau_\rho \hat{\varrho})$, where $\hat{\varrho}$ is a preliminary estimate and $\varrho(i - 1)$ is the value of ϱ in the previous iteration.

Figure 8 shows the trace plot and autocorrelations when no reparameterization is used, using the macro data of Section 6.1. We can see that the autocorrelations are much more persistent than those in Figure 1, particularly for ρ and n. For example, the lag 40 autocorrelation of n (ρ) is 0.87 (0.69) with no reparameterization, but equal to 0.01 (0.025) with the reparameterization, respectively. The effective sample sizes (ESS) of 10000 after burn-in iterations for (K, ρ, n, G) are (439, 81, 24, 107), without reparameterization and equal to (960, 878, 877, 522) with the reparameterization. However, the computation time with the latter is about 2.3 higher. Therefore, taking into account computation time, the algorithm with the reparameterization is about 16 times more efficient to sample n, 4.7 times more efficient to sample ρ, 2.3 times more efficient to sample G and roughly equally efficient to sample K.
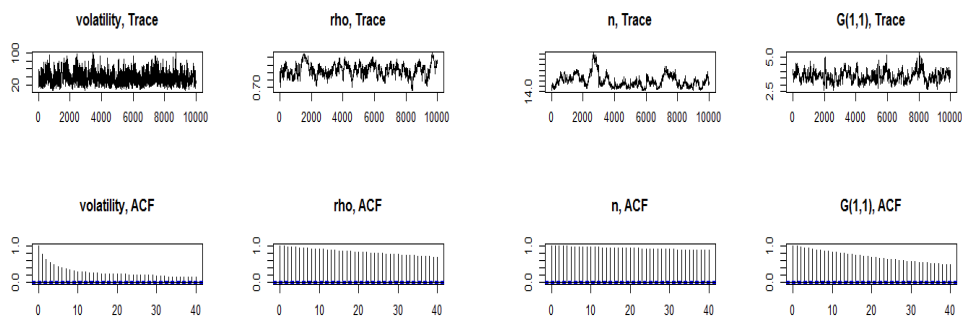

Figure 8: Trace plot and autocorrelations when no reparameterization is used in the model with $r_1 = 7$ (Macro data). Trace plots and autocorrelations are for the (7,7) element of $K_{T/2}$, the 4th diagonal element of ρ, for n and the (1,1) element of G.

Appendix IV

This appendix gives the details of the conditional prior and posterior of β, discussed in Section 5.4.

Proposition 8 *Assuming that the conditional prior of $vec(\beta')|G$ is a normal with mean $\underline{\mu}_\beta$ and co-variance matrix $\underline{V}_\beta$ given by:*

$$\underline{V}_\beta = (G \otimes \underline{V}_0) \, , \, for \, \underline{V}_0 : k \times k$$

the conditional posterior $vec(\beta')|G, K_{1:T}$ is also normal with mean $\overline{\mu}_\beta$ and covariance matrix $\overline{V}_\beta$ given by:

$$\overline{V}_\beta = (A_1 \otimes I_k) \left(\sum_{t=1}^{T} (K_t \otimes x_t x_t') + \widetilde{I}_{r_1}^{-1} \otimes \underline{V}_0^{-1} \right)^{-1} (A_1' \otimes I_k) + \tag{27}$$

$$\left(A_2 A_2' \otimes \left(\sum_{t=1}^{T} (x_t x_t') + \underline{V}_0^{-1} \right)^{-1} \right)$$

$$\overline{\mu}_\beta = \overline{V}_\beta \left(vec(\sum_{t=1}^{T} x_t y_t' \Sigma_t^{-1}) + \left(G^{-1} \otimes \underline{V}_0^{-1} \right) \underline{\mu}_\beta \right) \tag{28}$$

$$with \quad \widetilde{I}_{r_1}^{-1} = (n - r_1 - 1)(I_{r_1} - \rho^2)^{-1}$$

A draw of $vec(\beta')|G, K_{1:T}$ can be obtained as $\left((\overline{V}_\beta)^{1/2} \right)' \eta + \overline{\mu}_\beta$ where η is a $rk \times 1$ vector of independent standard normal variates, and $(\overline{V}_\beta)^{1/2}$ can be calculated as:

$$(\overline{V}_\beta)^{1/2} = (D_1, D_2)', \quad D_1 : rk \times r_1 k, \quad D_2 : rk \times r_2 k \tag{29}$$

$$D_1 = (A_1 \otimes I_k) \left(\left(\sum_{t=1}^{T} (K_t \otimes x_t x_t') + \widetilde{I}_{r_1}^{-1} \otimes \underline{V}_0^{-1} \right)^{-1/2} \right)'$$

$$D_2 = A_2 \otimes \left(\left(\sum_{t=1}^{T} (x_t x_t') + \underline{V}_0^{-1} \right)^{-1/2} \right)'$$

Proof. Standard calculations, similar to those in a multivariate regression model, show that $\overline{V}_\beta$ is given by

$$\overline{V}_\beta = \left(\sum_{t=1}^{T} \left(\Sigma_t^{-1} \otimes x_t x_t' \right) + G^{-1} \otimes \underline{V}_0^{-1} \right)^{-1} \tag{30}$$

and that $\overline{\mu}_\beta$ is given by (28). However, expression (30) requires the inversion of a $rk \times rk$ matrix. To obtain (27) first note that (3) and (22) imply that

$$G^{-1} = B \begin{pmatrix} \widetilde{I}_{r_1}^{-1} & 0 \\ 0 & I_{r_2} \end{pmatrix} B', \quad \Sigma_t^{-1} = B \begin{pmatrix} K_t & 0 \\ 0 & I_{r_2} \end{pmatrix} B'$$

Hence we can write (30) as:

$$\overline{V}_\beta = \left((B \otimes I_k) \left(\sum_{t=1}^{T} \left(\left(\begin{array}{cc} K_t & 0 \\ 0 & I_{r_2} \end{array} \right) \otimes x_t x_t' \right) + \left(\begin{array}{cc} \widetilde{I}_{r_1}^{-1} & 0 \\ 0 & I_{r_2} \end{array} \right) \otimes \underline{V}_0^{-1} \right) (B' \otimes I_k) \right)^{-1}$$

$$\overline{V}_\beta = (A \otimes I_k) \left(\left(\begin{array}{cc} \sum_{t=1}^{T} K_t \otimes x_t x_t' + \widetilde{I}_{r_1}^{-1} \otimes \underline{V}_0^{-1} & 0 \\ 0 & I_{r_2} \otimes \left(\sum_{t=1}^{T} x_t x_t' + \underline{V}_0^{-1} \right) \end{array} \right) \right)^{-1} (A' \otimes I_k)$$

$$\overline{V}_\beta = (A_1 \otimes I_k) \left(\sum_{t=1}^{T} (K_t \otimes x_t x_t') + \widetilde{I}_{r_1}^{-1} \otimes \underline{V}_0^{-1} \right)^{-1} (A_1' \otimes I_k) + A_2 A_2' \otimes \left(\sum_{t=1}^{T} x_t x_t' + \underline{V}_0^{-1} \right)^{-1} \quad (31)$$

which is equal to (27), and where we have used that $A \otimes I_k = (A_1, A_2) \otimes I_k = (A_1 \otimes I_k, A_2 \otimes I_k)$. To derive (29) simply note that $\left((\overline{V}_\beta)^{1/2} \right)' (\overline{V}_\beta)^{1/2} = D_1 D_1' + D_2 D_2' = \overline{V}_\beta$, as we wanted to show. ∎