GRIPS Discussion Paper 20-09

Multivariate Stochastic Volatility with Co-Heteroscedasticity

By

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

September 2020

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

This version: September 2020

(First version: October 2018)

[1]The authors thank seminar participants at the 12^{th} RCEA Bayesian Workshop, the 4^{th} Hitotsubashi Summer Institute and CEMFI Econometrics Workshop for helpful comments and suggestions. Roberto León-González acknowledges financial support from the GRIPS Policy Research Center under the grant "Multivariate Stochastic Volatility with Partial Homoscedasticity", from the Nomura Foundation (BE-004) and from JSPS (category C, 19K01588). Rodney Strachan acknowledges financial support from the GRIPS Policy Research Center for a research visit to GRIPS. Rodney Strachan is a Senior Fellow of the Rimini Centre for Economic Analysis (RCEA) and a Research Associate of the Centre for Applied Macroeconomic Analysis. Roberto León-González is a Senior Fellow of the RCEA. All errors are, of course, our own.

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; a property known as 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 volatility estimation. By incorporating testable co-heteroscedasticity restrictions, 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 an empirical application to a large Vector Autoregression (VAR), in which we find strong evidence for co-heteroscedasticity and that the new method compares favorably to previous ones in terms of forecasting from horizon 3 onward. A Monte Carlo experiment illustrates that the new method estimates well the characteristics of approximate factor models with heteroscedastic errors.

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.

Multivariate GARCH (MGARCH) models often suffer from the problem that the number of parameters to be estimated grows too quickly with the dimension. Another problem that often occurs is related to ensuring that the conditional variance matrix remains positive definite without imposing too strong restrictions on the parameters (e.g. Bauwens et al. (2006)). In the MGARCH context, conditionally heteroscedastic latent factor models (e.g. Diebold and Nerlove (1989), King et al. (1994)) provide a simple way to ensure positive definiteness, and help to reduce the number of parameters. However, the conditional variance matrix is no longer only a function of observables, which makes inference more complicated as the latent factor has to be marginalized in the likelihood (e.g. Gourieroux (1997), section 6.3).

Recent advances in computation have made possible the estimation of Multivariate SV (MSV) models, providing parsimonious yet flexible models in which the conditional variance is always positive definite. 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 has been found in MSV models (e.g. Clark and Ravazzolo (2015)).

The MSV literature is not free from problems. One common problem is that the order of the variables in a vector affects the estimates of its covariance matrix, implying that the methods are not invariant to ordering (e.g. Primiceri (2005)). Another problem is that, with the exception of factor SV models (e.g. Chib et al. (2006), Kastner (2019)), computational difficulties make most methods only applicable in practice to relatively small covariance matrices. The computational problems are related to the dimension of the latent variables and the typically high correlation between parameters and latent variables.

Most of the MSV literature uses a log-normal distribution for the volatility, and often result in methods that are not invariant to ordering, with several exceptions such as Yu and Meyer (2006) or Kastner (2019), among others. Furthermore, 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, Philipov and Glickman (2006), Casarin and Sartore (2007), Asai and McAleer (2009), Fox and West (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. However, the methods do not scale up to higher dimensions, unless very strong restrictions are imposed on the time variation of the process (e.g. Uhlig (1997), Triantafyllopoulos (2012)).

The approach in this paper uses a stationary inverse Wishart process together with testable co-heteroscedasticity (Engle and Kozicki (1993)) restrictions that reduce the dimension of the latent variables, permitting the analysis of larger covariance matrices and an interesting decomposition of the errors into homoscedastic and heteroscedastic parts. The model can be interpreted as an approximate factor model (Chamberlain and Rothschild (1983)) where the factors have a time varying covariance matrix. This allows for a more general time variation than hitherto proposed strict factor models, which assume independence among the heteroscedastic factors and independent idiosyncratic errors.

In order to surmount the computational difficulties, 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 additionally 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.

Our approach also allows us to obtain new variance decompositions as well as new insights into the characteristics of the structural shocks and their impacts on the variables of interest. Co-heteroscedasticity might also be relevant to find portfolio allocations with thinner tails and constant variance returns (e.g. Diebold and Nerlove, (1989), Engle and Susmel (1993) and Hernández-Trillo (1997)).

The structure of the paper is as follows. Section 2 describes the model and the identification strategy. Section 3 deals with decompositions of impulse responses and the variance. Section 4 develops the particle filter to estimate the likelihood while in Section 5 the particle Gibbs algorithm and tools

for model comparison are explained. Section 6 contains an application to US macro data, a comparison in terms of predictive performance with previous methods, and a Monte Carlo experiment to illustrate the properties of the new method.

2 Model and Identification Strategy

We consider the following model of stochastic volatility:

$$y_t = \beta x_t + e_t, \quad \text{with} \quad e_t|x_t \sim N(0, \Sigma_t), \quad t = 1, ..., T$$

where $\beta : r \times k_x$ is a matrix of unknown parameters, $x_t : k_x \times 1$ is a vector of predetermined variables, and $e_t : r \times 1$ is the vector of unobserved errors. We assume that $G = E(\Sigma_t)$ exists and is finite, and use its singular value decomposition (SVD, e.g. Magnus and Neudecker (1999, p. 18)), defined as:

$$G = E(\Sigma_t) = USU' = U_1 S_1 U_1' + U_2 S_2 U_2' \tag{1}$$

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:

$$\begin{array}{rcl} U & = & \left(\begin{array}{cc} U_1 & U_2 \end{array} \right) \quad S = \left(\begin{array}{cc} S_1 & 0 \\ 0 & S_2 \end{array} \right) \\ U_1 & : & r \times r_1, \quad U_2 : r \times r_2, \; S_1 : r_1 \times r_1, \;\; S_2 : r_2 \times r_2 \end{array} \tag{2}$$

We assume that $var(e_t|x_t) = \Sigma_t$ can be written as:

$$\Sigma_t = U_1 S_1^{1/2} \Upsilon_t^{-1} S_1^{1/2} U_1' + U_2 S_2 U_2' \tag{3}$$

where Υ_t is a Wishart Autoregressive process of order 1 (WAR(1), Gourieroux et al. (2009)), normalized such that $E(\Upsilon_t^{-1}) = I_{r_1}$. This implies that vector e_t can be decomposed into r_1 heteroscedastic errors ($u_{1t} : r_1 \times 1$, $var(u_{1t}) = \Upsilon_t^{-1}$) and r_2 homoscedastic errors ($u_{2t} : r_2 \times 1$, $var(u_{2t}) = I_{r_2}$, $cov(u_{1t}, u_{2t}) = 0$), with $r = r_1 + r_2$:

$$e_t = A_1 u_{1t} + A_2 u_{2t} = \left(\begin{array}{cc} A_1 & A_2 \end{array} \right) \left(\begin{array}{c} u_{1t} \\ u_{2t} \end{array} \right) = A u_t \tag{4}$$

and $A_1 = U_1 S_1^{1/2}$, $A_2 = U_2 S_2^{1/2}$. The WAR(1) can be described by first defining $K_t = Z_t' Z_t$, where Z_t

is a $n \times r_1$ matrix distributed as a Gaussian AR(1) process:

$$Z_t = Z_{t-1}\rho + \varepsilon_t \qquad vec(\varepsilon_t) \sim N(0, I_{r_1} \otimes I_n) \tag{5}$$

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 $N(0, (I_{r_1} - \rho^2)^{-1} \otimes I_n)$. The parameter n represents the degrees of freedom in the WAR(1) process and it will be estimated[2]. Because $E(K_t^{-1}) = (n - r_1 - 1)^{-1}(I - \rho^2)$, we normalize K_t^{-1} as $\Upsilon_t^{-1} = (n - r_1 - 1)(I - \rho^2)^{-1/2} K_t^{-1} (I - \rho^2)^{-1/2}$, so that $E(\Upsilon_t^{-1}) = I_{r_1}$. We assume that $n > r_1 + 1$, such that $E(\Sigma_t)$ is finite[3].

We follow a Bayesian approach and put a prior directly on the model parameters (G, ρ, n, β). Section 5 will describe a Particle Gibbs algorithm to sample from the posterior of model parameters and latent variables given the observed data Y, defined as $Y = (y_1, x_1, ..., y_T, x_T)$. The latent variables are related to the volatility matrices K_t, and will be defined in Section 5[4].

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 rk_x to $r_1 k_x$.

The properties of the WAR(1) 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 $E(K_t)$:

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

The correlations between different elements of Σ_t^{-1} conditional on Σ_{t-1}^{-1} are controlled by the parameters in A_1 and can be calculated using the properties of the non-central Wishart.

Assumption (3) is natural if it is assumed that e_t has an approximate factor structure with only r_1 heteroscedastic factors and r is large relative to r_1, because a factor structure implies that the first

[2]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 $n \geq 2r_1$.

[3]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.

[4]Similarly to Jacquier et al. (1994), who estimate a univariate stochastic volatility model, we expand the parameter space by treating the latent variables as parameters, which leads to a more tractable form for the likelihood function. The particle Gibbs allows us to obtain a draw of all volatilities jointly conditional on the model parameters (G, ρ, n, β), and a draw of model parameters conditional on the volatilities.

r_1 eigenvalues of G grow without bound as r gets larger, whereas the other eigenvalues are bounded (provided that each of the common factors affect a large number of variables, and hence the factors are 'pervasive', see for example Chamberlain and Rothschild (1983) or Bai and Ng (2002)). Hence we can interpret u_{1t} as the heteroscedastic factors and A_1 as the factor loadings. Note that only the space spanned by A_1 ($sp(A_1)$) is identified, and we solve the indeterminacy by choosing $A_1 = U_1 S_1^{1/2}$, which restricts $A_1' A_1$ to be a diagonal matrix. In Section 6.3 we provide a Monte Carlo experiment that illustrates that our model performs well as a factor model for moderate values of r.

Another practical reason for assumption (3) is that it is well-known that stochastic volatility is empirically important, and therefore it seems reasonable to introduce Υ_t^{-1} into the model in the way that it will have the greatest impact, which is by interacting Υ_t^{-1} with the largest singular values of G.

Note that our model 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. It can be said then that the elements of the vector y_t share heteroscedastic shocks, a property known in the literature as common heteroscedasticity (Engle and Kozicki (1993)), and which for convenience we refer to as co-heteroscedasticity. However, it should be noted that there might be co-heteroscedasticity patterns that our model fails to capture, especially in small VARs of 2 up to 10 variables.

Note that our model differs from traditional factor models of unobserved volatility (e.g. King et al. (1994), Yu and Meyer (2006), Chib et al. (2006), Kastner (2019)) in two important aspects. Firstly, we allow for time-varying correlations among the heteroscedastic factors. This implies that the time-varying process has dimension $r_1(r_1 + 1)/2$, which approaches the number of free elements of Σ_t as r_1 approaches r. Secondly, as in Chamberlain and Rothschild (1983), we do not impose the strict factor structure, which would imply the existence of independent idiosyncratic errors, but instead allow correlation among the homoscedastic components. Although in previous factor models the correlation between two elements in e_t changes with time, the restrictions on the correlation among the latent factors imply that the time-variation of Σ_t is more restricted that in our setup (Kastner (2019, p. 100)). On the other hand, some previous models (e.g. Chib et al. (2006)) allow for features that our model does not have, such as jumps in the mean or student-t errors, but these could also be incorporated in our setup.

Asai and McAleer (2015) propose a factor model for realized volatility, the fMSV-WAR, which is related to ours. One difference is that the fMSV-WAR specifies a WAR(1) process for Υ_t^{-1}, implying that all the moments of Σ_t exist. Our model specifies a WAR(1) process for Υ_t, implying that the number of finite moments of Σ_t depends on n. Another difference is that the availability of realized volatility data permits the fMSV-WAR to identify and estimate more parameters, and to do without

the assumption of a large r. Asai and McAleer (2015) provide generalizations of the fMSV-WAR model to incorporate asymmetric and heterogeneous-time effects, which could perhaps be incorporated into our framework in future research.

3 Decompositions, Identification of Structural Shocks and Impulse Responses

Equation (4) implies that e_t has a heteroscedastic component $e_t^{het} = A_1 u_{1t}$ and a homoscedastic component $e_t^{hom} = A_2 u_{2t}$. By defining H_1 as $H_1 = U_1 U_1'$, the heteroscedastic component of e_t can be obtained as $e_t^{het} = H_1 e_t = A_1 u_{1t}$, 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:

$$\begin{aligned} y_t &= y_t^{het} + y_t^{hom} = \mu + e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + ... \\ y_t^{het} &= H_1\mu + H_1 e_t + \Psi_1 H_1 e_{t-1} + \Psi_2 H_1 e_{t-2} + ... + \\ y_t^{hom} &= H_2\mu + H_2 e_t + \Psi_1 H_2 e_{t-1} + \Psi_2 H_2 e_{t-2} + ... \end{aligned} \tag{6}$$

where $H_2 = I_r - H_1 = U_2 U_2'$. The matrices H_1 and H_2 are also useful to isolate the heteroscedastic component Σ_t^{het} of the volatility matrix as $\Sigma_t^{het} = H_1 \Sigma_t$, and the homoscedastic component $\Sigma_t^{hom} = H_2 \Sigma_t$, such that $\Sigma_t = \Sigma_t^{het} + \Sigma_t^{hom}$.

In the VAR literature it is common to identify the structural shocks ξ_t by finding a matrix P_t such that $\Sigma_t = P_t P_t'$ and defining $\xi_t = P_t^{-1} e_t$. For example P_t could be a triangular matrix based on the assumption that some variables do not react contemporaneously to the impulse (e.g. Banbura et al. (2010)). In this case the impulse response function can be calculated as the difference of two forecasts (e.g. Koop et al. (1996)):

$$\begin{aligned} IR_t(s;i,j) &= E(y_{i,t+s}|\xi_{jt} = 1, y_{1:(t-1)}) - E(y_{i,t+s}|\xi_{jt} = 0, y_{1:(t-1)}) \\ IR_t^{het}(s;i,j) &= E(y_{i,t+s}^{het}|\xi_{jt} = 1, y_{1:(t-1)}) - E(y_{i,t+s}^{het}|\xi_{jt} = 0, y_{1:(t-1)}) \\ IR_t^{hom}(s;i,j) &= E(y_{i,t+s}^{hom}|\xi_{jt} = 1, y_{1:(t-1)}) - E(y_{i,t+s}^{hom}|\xi_{jt} = 0, y_{1:(t-1)}) \end{aligned}$$

where $IR_t(s;i,j)$, $IR_t^{het}(s;i,j)$, $IR_t^{hom}(s;i,j)$ correspond to the (i,j) elements of $\Psi_s P_t$, $\Psi_s H_1 P_t$ and $\Psi_s H_2 P_t$, respectively, with $\Psi_0 = I_r$, $y_{1:(t-1)} = (y_1, ..., y_{t-1})$, $y_t = (y_{1t}, ..., y_{rt})$ and $\xi_t = (\xi_{1t}, ..., \xi_{rt})$. Note that $IR_t(s;i,j) = IR_t^{het}(s;i,j) + IR_t^{hom}(s;i,j)$ and that $IR^{het}(s;i,j)$ can be interpreted as the

impact of ξ_{jt} on y_t^{het} at horizon s[5].

Note that y_t^{het} is a sum of only heteroscedastic shocks, and y_t^{hom} is a sum of only homoscedastic shocks. Equation (6) can be used to obtain variance decompositions, calculating the proportion of $var(y_{t+h}|y_t), var(y_{t+h}^{het}|y_t)$ or $var(y_{t+h}^{hom}|y_t)$ caused by the heteroscedastic shocks u_{1t} or by ξ_t, as we illustrate in the empirical section. As the impulse response functions, these variance decompositions change with t.

By decomposing y_t into y_t^{het} and y_t^{hom} using equation (6), 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.

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

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. (1993)), such an 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[6].

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]In factor models it is common to identify the structural shocks ξ_t as a rotation of the factors: $\xi_t = C_t \Upsilon_t^{1/2} u_{1t}$, where C_t is a $r_1 \times r_1$ orthogonal matrix such that $C_t' C_t = I$. The matrix C_t can be defined to facilitate the economic interpretation of the impulse responses (e.g. Uhlig (2005) or Gonzalo and Ng (2001)). We leave this as an avenue for future research.

[6]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 $\pi(K_{1:T}|Y,\theta)$[7].

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:

$$L(Y|\Sigma_{1:T},\beta) = \left(\prod_{t=1}^{T}|\Sigma_t|^{-1/2}\right)\exp\left(-\frac{1}{2}\delta\sum_{t=1}^{T}tr\left(\Sigma_t^{-1}e_te_t^{'}\right)\right)\widetilde{L}(Y|\Sigma,\beta)$$

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

$$\widetilde{L}(Y|\Sigma_{1:T},\beta) = (2\pi)^{-Tr/2}\exp\left(-\frac{1}{2}(1-\delta)\sum_{t=1}^{T}tr\left(\Sigma_t^{-1}e_te_t^{'}\right)\right), \text{ where } e_t = y_t - \beta x_t \tag{7}$$

where δ is a scalar between 0 and 1 that can be tuned to improve the efficiency of the particle filter. Setting $\delta = 1$ gives the bootstrap particle filter, and after some experimentation we set $\delta = 0.8$ in the empirical analysis of Section 6.1. 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:

$$\begin{aligned} Z_t &= Z_{t-1}\rho V_t + \varepsilon_t \qquad & vec(\varepsilon_t) \sim N(0, V_t \otimes I_n), \text{ for } t > 1 \\ Z_1 &= \varepsilon_1 \qquad & vec(\varepsilon_1) \sim N(0, V_1 \otimes I_n) \end{aligned} \tag{8}$$

where V_t is given by the following recursion:

$$\begin{aligned} V_T &= (I + (1-\delta)B_1'e_Te_T'B_1)^{-1} \\ V_t &= (I + (1-\delta)B_1'e_te_t'B_1 + \rho(I - V_{t+1})\rho)^{-1} \qquad t > 1 \\ V_1 &= (I + (1-\delta)B_1'e_1e_1'B_1 - \rho V_2\rho)^{-1} \end{aligned} \tag{9}$$

where $B = (B_1, B_2) = ((\widetilde{A}_1, A_2)^{-1})'$, with $\widetilde{A}_1 = \sqrt{n - r_1 - 1}A_1(I - \rho^2)^{-1/2}$ and $B_1 : r \times r_1$. Appendix I shows that the true likelihood, after integrating out the latent K, can be compactly written as:

[7]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).

$$L(Y|\beta,\theta) = \widetilde{c}_L E_{\widetilde{\pi}}\left(\prod_{t=1}^{T}\left[|K_t|^{1/2}\exp\left(-\frac{\delta}{2}tr(K_t B_1' e_t e_t' B_1)\right)\right]\right) \quad (10)$$

$$\widetilde{c}_L = \left|I-\rho^2\right|^{n/2}\left|\widetilde{A}_1\widetilde{A}_1' + A_2 A_2'\right|^{-T/2}\exp\left(-\frac{1}{2}\sum_{t=1}^{T} tr\left(B_2 B_2' e_t e_t'\right)\right)\left(\prod_{t=1}^{T}|V_t|^{n/2}\right) \quad (11)$$

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 (10) 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 1 *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:

$$C_1 := \frac{1}{N}\sum_{m=1}^{N}\frac{|K_1^m|^{1/2}}{\exp\left(\frac{\delta}{2}tr\left(K_1^m B_1' e_1 e_1' B_1\right)\right)}, \quad W_1^k := \frac{1}{NC_1}\frac{\left|K_1^k\right|^{1/2}}{\exp\left(\frac{\delta}{2}tr\left(K_1^k B_1' e_1 e_1' B_1\right)\right)}$$

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

$$C_t := \frac{1}{N}\sum_{m=1}^{N}\frac{|K_t^m|^{1/2}}{\exp\left(\frac{\delta}{2}tr\left(K_t^m B_1' e_t e_t' B_1\right)\right)}, \quad W_t^k := \frac{1}{NC_t}\frac{\left|K_t^k\right|^{1/2}}{\exp\left(\frac{\delta}{2}tr\left(K_t^k B_1' e_t e_t' B_1\right)\right)}$$

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 (5) 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 (5), we get the following:

$$\begin{aligned} K_t &\sim W(n, (I - \rho^2)^{-1}), \text{ for } t = 1 \\ L_{1t} &= (K_{t-1})^{1/2} \rho + \varepsilon_{1t}, \quad \varepsilon_{1t} : r_1 \times r_1, \quad vec(\varepsilon_{1t}) \sim N(0, I_{r_1} \otimes I_{r_1}), \text{ for } t > 1 \\ K_{2t} &\sim W(n - r_1, I_{r_1}), \text{ for } t > 1 \end{aligned} \tag{12}$$

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 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[8]. 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 and $\rho = 0.8I_{r_1}$ (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 2 *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:

$$w_1^k := \frac{\left|K_1^k\right|^{1/2}}{\exp\left(\frac{\delta}{2} tr\left(K_1^k B_1' e_1 e_1' B_1\right)\right)},\ W_1^k := w_1^k / (\sum_{m=1}^{N} w_1^m),\ k = 1, ..., N$$

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 $vec(L_{1t}^k) \sim N(\left(K_{t-1}^{A_{t-1}^k}\right)^{1/2} \rho V_t, I_{r_1} \otimes V_t)$ 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

[8]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.

(c) compute and normalize the weights

$$w_t^k := \frac{\left|K_t^k\right|^{1/2}}{\exp\left(\frac{\delta}{2} tr\left(K_t^k B_1' e_t e_t' B_1\right)\right)}, \quad W_t^k := w_t^k / (\sum_{m=1}^{N} w_t^m),\ k = 1, ..., N$$

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

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

(a) compute the updated weights

$$\begin{aligned} \widetilde{w}_t^k &= w_t^k f(K_{t+1}^{b_{t+1}} | K_t^k), ..., \widetilde{W}_t^k := \widetilde{w}_t^k / (\sum_{m=1}^{N} \widetilde{w}_t^m),\ k = 1, ..., N, \text{ where} \\ \mu_t^k &= (K_t^k)^{1/2} \rho V_{t+1}, \quad f(K_{t+1}^{b_{t+1}} | K_t^k) = \exp(-\frac{1}{2} tr(V_{t+1}^{-1} (L_{1(t+1)}^{b_{t+1}} - \mu_t^k)' (L_{1(t+1)}^{b_{t+1}} - \mu_t^k)) \end{aligned}$$

(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 (5) 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 (12) we need to define first $(\widetilde{L}_{1t}, \widetilde{K}_{2t})$ as follows:

$$\begin{aligned} \widetilde{L}_{1t} &= (K_{t+1}^{-1/2})' L_{1,t+1}' K_t^{1/2}, \quad \text{for } t = 1, ..., (T-1) \\ \widetilde{K}_{2t} &= K_t - \widetilde{L}_{1t}' \widetilde{L}_{1t} \end{aligned} \tag{13}$$

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

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

Then the pseudo-posterior in (8) can be written in reverse order by first adapting the recursion in (9) as follows:

$$\begin{aligned}
\widetilde{V}_1 &= (I + (1-\delta)B_1' e_1 e_1' B_1)^{-1} \\
\widetilde{V}_t &= (I + (1-\delta)B_1' e_t e_t' B_1 + \rho(I - \widetilde{V}_{t-1})\rho)^{-1} \qquad t < T \\
\widetilde{V}_T &= (I + (1-\delta)B_1' e_T e_T' B_1 - \rho\widetilde{V}_{T-1}\rho)^{-1}
\end{aligned} \tag{15}$$

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

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

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 (13):

$$\begin{aligned}
L_{1t} &= (K_{t-1}^{-1/2})'\widetilde{L}_{1,t-1}' K_t^{1/2}, \quad \text{for } t = 2, ..., T \\
K_{2t} &= K_t - L_{1t}' L_{1t}
\end{aligned} \tag{17}$$

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[9] $L^*_{1:T} = f^{-1}_{\theta^*}(\epsilon)$. 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. 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[10]. In our context, we can apply this strategy by rewriting the pseudo-posterior in (8) using the decomposition of the non-central Wishart:

$$\begin{aligned} K_t &\sim W(n, V_t), \text{ for } t = 1 \\ L_{1t} &= (K_{t-1})^{1/2}\rho V_t + \varepsilon_{1t}, \quad \varepsilon_{1t} : r_1 \times r_1, \quad vec(\varepsilon_{1t}) \sim N(0, V_t \otimes I_{r_1}), \text{ for } t > 1 \\ K_{2t} &\sim W(n - r_1, V_t), \text{ for } t > 1 \end{aligned} \tag{18}$$

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 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).

[9] 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).

[10] 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 independent 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.

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, and its conditional posterior is normal. Although the conditional posterior of β is a normal, when rk is large it could be time consuming to calculate the var-cov matrix of β, because it requires inverting matrices of the order $rk \times rk$. To avoid this problem, we follow the large VAR literature and choose the prior var-cov matrix for β to have a Kronecker structure, $\underline{V}_\beta = G \otimes \underline{V}_0$, such that $(\underline{V}_\beta)^{-1} = G^{-1} \otimes (\underline{V}_0)^{-1}$, where $\underline{V}_0$ controls the degree of shrinkage. In the homoscedastic VAR, this implies that the posterior var-cov matrix $(\overline{V}_\beta)$ also has a Kronecker structure, and hence inverting the posterior variance only requires to invert matrices of the order $r \times r$ and $k \times k$. In heteroscedastic VARs such as ours, $\overline{V}_\beta$ does not necessarily have a Kronecker structure. However, in our case the Kronecker structure in the prior implies that $\overline{V}_\beta$ and $(\overline{V}_\beta)^{1/2}$ can be calculated by inverting matrices of order $r_1 k \times r_1 k$ and $k \times k$. In our empirical application with $r = 20$ and $k = 81$, we conclude that the best value of r_1 is 5, and therefore we only need to invert matrices of order 405 (as opposed to 1620). The following proposition, whose proof is in appendix IV, summarizes this result.

Proposition 3 *Assuming that the conditional prior of $vec(\beta')|G$ is a normal with mean $\underline{\mu}_\beta$ and covariance matrix $\underline{V}_\beta$ given by:*

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

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:

$$\begin{aligned}
\overline{V}_\beta &= (\widetilde{A}_1 \otimes I_k)\left(\sum_{t=1}^{T}(K_t \otimes x_t x_t^{'}) + \widetilde{I}_{r_1}^{-1} \otimes \underline{V}_0^{-1}\right)^{-1}(\widetilde{A}_1' \otimes I_k) + \\
&\quad \left(A_2 A_2' \otimes \left(\sum_{t=1}^{T}(x_t x_t^{'}) + \underline{V}_0^{-1}\right)^{-1}\right) \qquad (19)\\
\overline{\mu}_\beta &= \overline{V}_\beta\left(vec(\sum_{t=1}^{T} x_t y_t' \Sigma_t^{-1}) + \left(G^{-1} \otimes \underline{V}_0^{-1}\right)\underline{\mu}_\beta\right) \qquad (20)\\
with\ \widetilde{I}_{r_1}^{-1} &= (n - r_1 - 1)(I_{r_1} - \rho^2)^{-1},\ \widetilde{A}_1 = \sqrt{n - r_1 - 1}A_1(I - \rho^2)^{-1/2}
\end{aligned}$$

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

standard normal variates, and $\left(\overline{V}_\beta\right)^{1/2}$ *can be calculated as:*

$$\begin{aligned}
\left(\overline{V}_\beta\right)^{1/2} &= (D_1, D_2)', \quad D_1 : rk \times r_1 k, \quad D_2 : rk \times r_2 k \qquad (21)\\
D_1 &= (\widetilde{A}_1 \otimes I_k)\left(\left(\sum_{t=1}^{T}(K_t \otimes x_t x_t^{'}) + \widetilde{I}_{r_1}^{-1} \otimes \underline{V}_0^{-1}\right)^{-1/2}\right)'\\
D_2 &= A_2 \otimes \left(\left(\sum_{t=1}^{T}(x_t x_t^{'}) + \underline{V}_0^{-1}\right)^{-1/2}\right)'
\end{aligned}$$

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 i^{th} iteration are generated as follows:

- If i is even:
  - Generate a value $L_{1:T}^*$ for $L_{1:T}$, using the natural order Algorithm 2 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 (13).
  - 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 (17).
- 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 (28) 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 3 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. The prior for each of the diagonal elements of ρ is a beta distribution, so we approximate the posterior ordinate of ρ using independent beta calibrated with the posterior mean and variance of ρ delivered by the MCMC algorithm. Similarly, the prior for $\widetilde{n} = \log(n - 2r_1)$ is a normal, so we use also a normal to approximate the posterior of $\widetilde{n}$. We also use the BIC and marginal likelihood to compare the model that assumes $n \geq 2r_1$ with models that assume $r_1 + 2 \leq n < 2r_1$. The Monte Carlo experiment in Section 6.3 suggests that the marginal likelihood calculated in this manner gives good estimates of r_1, and is slightly more accurate than the BIC criterion.

6 Empirical Application

6.1 Macroeconomic Data

We use 20 macroeconomic quarterly variables (defined in Table 1) with 4 lags in the VAR, for the period 1959Q3-2013Q4 and specify a normal-inverse-Wishart prior with shrinkage for (β, G)[11]. Therefore y_t is a 20×1 vector, and x_t is a 81×1 vector containing 4 lags of y_t and the intercept. Table 2 shows model selection criteria when the restriction $n \geq 2r_1$ is imposed, indicating that the best value for r_1 is 5 according to both the marginal likelihood and BIC. Models that impose the restriction[12] $n = r_1 + 2$, not shown here for brevity, also estimate a value of r_1 equal to 5. Although the model with $n = r_1 + 2$ gives a higher value of the log-likelihood and marginal likelihood, the model with $n \geq 2r_1$ performs better in terms of the predictive likelihoods calculated in Section 6.2, so this is the model we choose. Estimating the preferred model took 4.8 hours in a Intel Xeon CPU (E5-2690) with 2.9 GHz, using 10 cores, and obtaining 12000 iterations with no thinning, a burn-in of 20 iterations and 320 particles[13]. The evaluation of the loglikelihood, using the average of 100 independent replications of the particle filter, each one with $2Tr_1 = 2140$ particles, took 4.8 minutes, with a numerical standard error of 0.09. For the preferred model, Figure 1 shows the trace plot and autocorrelations for the fifth 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.

[11] This is the same dataset used in Chan (2020), and the shrinkage prior has $k_1 = 0.04, k_2 = 100$, with (k_1,k_2) defined in Chan (2020). The degrees of freedom parameter for the inverse-Wishart prior of G is $r + 3$. The prior for each diagonal element of ρ is a beta distribution $B(95, 5)$ and the prior for $\widetilde{n} = \log(n - 2r_1)$ is a Normal: $N(\ln(23.5 - 2r_1), 1.5)$.

[12] Note that when we estimate n with the posterior simulator, we require $n \geq 2r_1$. To check whether this restriction is appropriate, we also estimate the model with the restriction $n = r_1 + 2$, which is the minimum integer value of n that ensures that $E(\Sigma_t)$ exists. For simplicity we do not check other values of n.

[13] 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.

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. (2019), 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 2[14].

For the purpose of illustration, variance decompositions and impulse responses are calculated at the period 1978 Q2. Table 3 shows the proportion of the variance $var(y_{t+h}|y_t)$ 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 (99%, 17%), whereas they become (88%, 30%) 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 $var(y_{t+h}^{het}|y_t)$ 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[15] 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. A variance decomposition shows that the structural monetary policy shock is mostly homoscedastic, with only 1.3% of its conditional variance caused by the heteroscedastic shocks u_{1t} at 1978Q2.

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}.

[14]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 (2019).

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

Variable	Transformation	Speed	Label
Real gross domestic product	400 Δ log	Slow	y_{1t}
Consumer price index	400 Δ log	Slow	y_{2t}
Effective Federal funds rate	no transformation		y_{3t}
M2 money stock	400 Δ log	Fast	y_{4t}
Personal income	400 Δ log	Slow	y_{5t}
Real personal consumption expenditure	400 Δ log	Slow	y_{6t}
Industrial production index	400 Δ log	Slow	y_{7t}
Civilian unemployment rate	no transformation	Slow	y_{8t}
Housing starts	log	Slow	y_{9t}
Producer price index	400 Δ log	Slow	$y_{10,t}$
Personal consumption expenditures: chain-type price index	400 Δ log	Slow	$y_{11,t}$
Average hourly earnings: manufacturing	400 Δ log	Slow	$y_{12,t}$
M1 money stock	400 Δ log	Fast	$y_{13,t}$
10-Year Treasury constant maturity rate	no transformation	Fast	$y_{14,t}$
Real gross private domestic investment	400 Δ log	Slow	$y_{15,t}$
All employees: total nonfarm	400 Δ log	Slow	$y_{16,t}$
ISM manufacturing: PMI composite index	no transformation	Slow	$y_{17,t}$
ISM manufacturing: new orders index	no transformation	Slow	$y_{18,t}$
Business sector: real output per hour of all Persons	400 Δ log	Slow	$y_{19,t}$
Real stock prices (S&P 500 index divided by PCE index)	100 Δ log	Fast	$y_{20,t}$

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
0	-5824.0	-6529.5	-8727.1	-6529.5
1	-5916.5	-6542.4	-8716.6	-6547.7
2	-5859.6	-6486.5	-8674.1	-6494.5
3	-5822.6	-6469.5	-8653.2	-6480.3
4	-5776.7	-6458.4	-8644.3	-6471.8
5	-5763.0	-6455.5	-8638.1	-6471.6
6	-5742.3	-6455.8	-8648.2	-6474.6
7	-5735.6	-6464.2	-8651.7	-6485.7
8	-5727.1	-6466.4	-8653.4	-6490.5

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. Numerical standard errors (NSE) for the log likelihood values were estimated using 100 independent replications of the particle filter, with $N = 2Tr_1$ particles. In all cases, the NSE values were smaller than 0.42.

	Mon Policy Shock (ξ_t)						Hetero shock (u_{1t})					
	1	2	3	4	5	∞	1	2	3	4	5	∞
GDP	0	0	1	2	2	3	86	82	76	73	71	61
GDP^{het}	0	0	0	0	0	0	100	100	100	100	100	100
GDP^{hom}	0	1	6	7	7	9	0	0	0	0	0	0
INF	0	3	3	3	2	4	55	64	65	65	68	44
INF^{het}	0	0	0	0	0	0	100	100	100	100	100	100
INF^{hom}	0	7	9	8	8	7	0	0	0	0	0	0
FFR	72	51	42	34	31	11	17	38	48	56	59	30
FFR^{het}	0	0	0	0	0	0	100	100	100	100	100	100
FFR^{hom}	83	81	79	77	75	15	0	0	0	0	0	0
BOND	10	11	12	13	14	4	42	48	50	51	52	27
BOND^{het}	0	0	0	0	0	0	100	100	100	100	100	100
BOND^{hom}	16	20	22	25	27	6	0	0	0	0	0	0
SP	0	0	0	0	0	1	99	97	96	96	95	88
SP^{het}	0	0	0	0	0	0	100	100	100	100	100	100
SP^{hom}	0	1	3	3	4	8	0	0	0	0	0	0

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

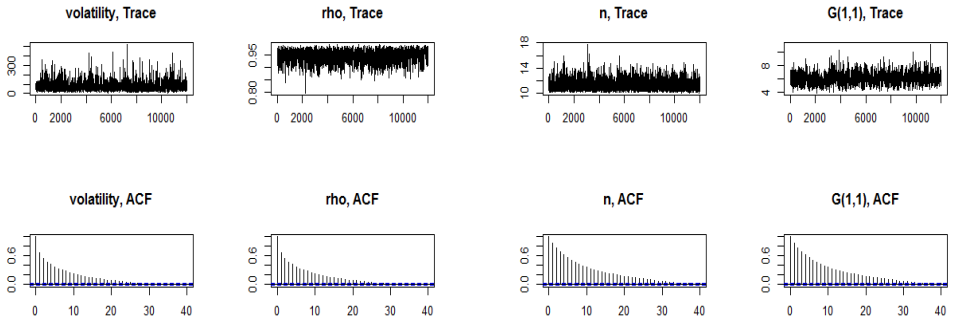

Figure 1: Trace plot and autocorrelations for the (5,5) element of K_{107}, the 4^{th} diagonal element of ρ, for n and the (1,1) element of G (Macro data).

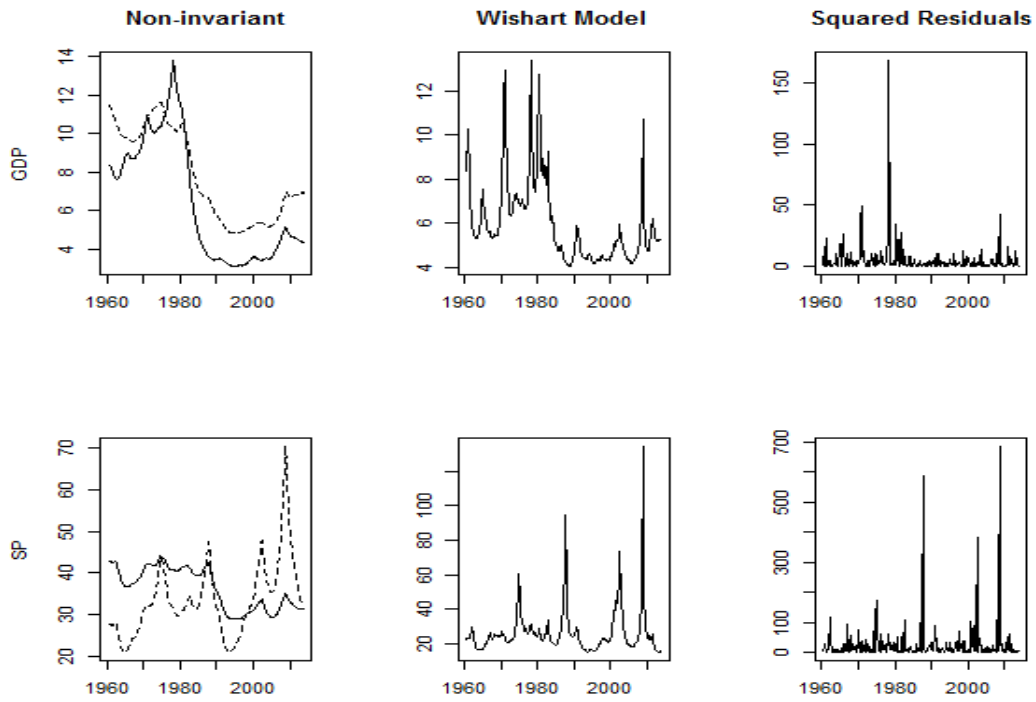

Figure 2: Estimated reduced form volatilities and squared Bayesian residuals for the real GDP growth rate (GDP) and the real stock returns (SP). The left panel corresponds to the method of Carriero et al. (2019) under two different orderings: the solid line is for ordering 1 and the dotted line is for ordering 2. The central panel corresponds to our method with $r_1 = 5$, and the right panel are the squared of the Bayesian residuals.

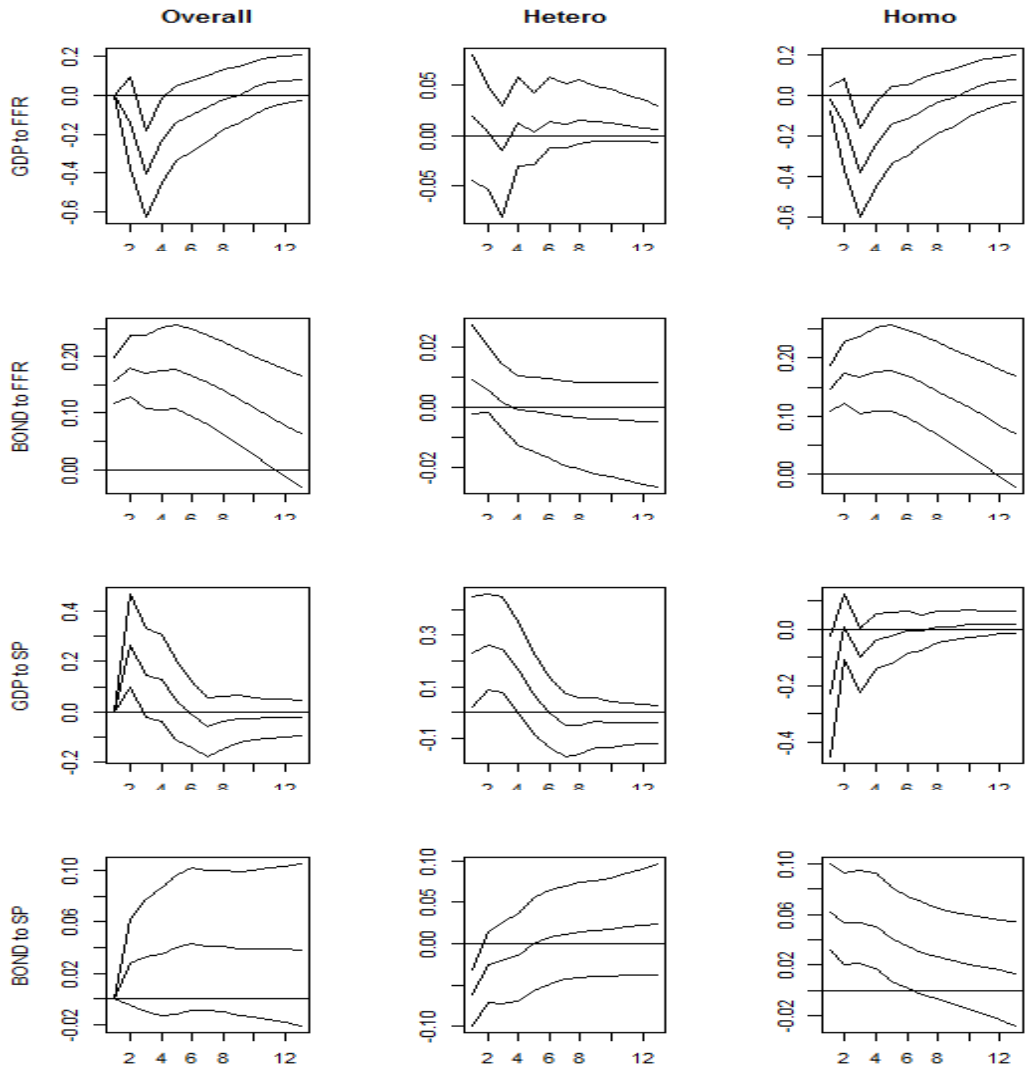

Figure 3: Impulse Responses of GDP growth (GDP) and the 10 year bond yield (BOND) to a monetary policy shock (FFR) and to a stock return shock (SP). The left panel plots the overall response, the central panel plots the reponse of the heteroscedastic component (y^{het}), and the right panel the homoscedastic component (y^{hom}). Posterior median and 90% credible interval. Horizons 1 up to 12.

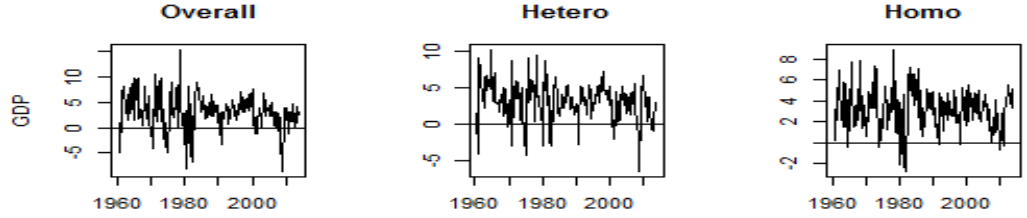

Figure 4: Decomposition of GDP growth into its heteroscedastic and homoscedastic components.

6.2 Comparison to Other Approaches

We compare our model with $r_1 = 5$ and $r_1 = 6$ (denoted as WAR-5 and WAR-6 respectively) to other approaches in terms of out-of-sample forecasting. The evaluation period consists of 116 observations, from 1985Q1 to 2013Q4. We calculate the average log predictive score for the joint vector $y = (y_1, ..., y_{20})$ for several horizons h ($\pi_{1:20,h}$), as well as the Root Mean Squared Forecasting Error (RMSFE,) and average log predictive score for each variable y_{it} for horizon 1 (denoted as E_i and π_i, respectively, for $i = 1, ..., 20$)[16].

We consider the factor stochastic volatility model of Chib et al. (2006), with both normal and student-t idiosyncratic errors. We use 5 and 6 factors because this gives a better value for $\pi_{1:20,h}$ than a smaller amount of factors. The models with normal errors are denoted as fSV-5 and fSV-6 (for 5 and 6 factors, respectively), and with student-t errors as fSVt-5 and fSVt-6. These models have a stationary process for stochastic volatility, and we fix the prior mean for the coefficient of the lag volatility to 0.95 (as we do in the Wishart model). The prior for $vec(\beta')$ has the same shrinkage structure.

We also consider models with non-stationary stochastic volatility, such as the previously mentioned model of Carriero et al. (2019), denoted as SV, and also the common drifting Volatility model of Carriero et al. (2016), denoted as CSV and CSV-t for normal and student-t errors[17], respectively.

Table 4 shows that the WAR-5 and WAR-6 models have the best values of $\pi_{1:20,h}$ among all models for horizons 3, 4 and 8. The relative advantage of the WAR models increases with the horizon. The second best model for horizon 3 is the fSV6-t, and for horizons 4 and 8 the second best model is the CSV-t. For horizon one, the fSV models and the SV model are better than the WAR models. For the second horizon, only the fSVt-6 model is better than the WAR models. All models are better than the homoscedastic VAR model.

In terms of the RMSFE and π_i for each of the variables, the results in Table 4 indicate that there is no model that is better for all the variables. In terms of the RMSFE, the WAR-5 model is equal or better than the non-WAR models for 9 of the variables, the CSV is equal or better than the non CSV models for 7 of the variables, the CSV-t is also equal or better (than the non-CSV models) for 7 of the variables, the SV is equal or better than the other models for 6 of the variables, whereas the fSV-6 or fSVt-6 models are equal or better for only 3 variables. Doing similar comparisons in terms of π_i, the SV is the best for 7 variables, followed by the WAR and CSV-t models with 4 variables.

[16] See for example Chan (2020) for a more detailed definition of π_i and E_i

[17] Pajor (2006) and Yu and Meyer (2006) also used the common volatility assumption, but assuming stationarity.

	VAR	WAR-5	WAR-6	fSV-5	fSV-6	fSVt-5	fSVt-6	SV	CSV	CSVt
$\pi_{1:20,1}$	-36.33	-35.43	-35.53	-35.39	-35.26	-35.22	-35.08	-35.15	-35.61	-35.56
$\pi_{1:20,2}$	-39.19	-38.24	-38.30	-38.42	-38.4	-38.35	-38.12	-38.37	-38.49	-38.37
$\pi_{1:20,3}$	-40.76	-39.88	-39.80	-40.22	-40.19	-40.08	-39.94	-40.13	-40.22	-40.01
$\pi_{1:20,4}$	-41.97	-41.17	-41.14	-41.65	-41.70	-41.44	-41.36	-41.50	-41.43	-41.22
$\pi_{1:20,8}$	-44.58	-43.87	-43.84	-44.43	-44.28	-44.10	-44.19	-44.35	-44.18	-44.04
E_1	2.18	2.20	2.17	2.41	2.46	2.42	2.45	2.08	2.14	2.14
π_1	-2.25	-2.26	-2.25	-2.32	-2.35	-2.32	-2.33	-2.17	-2.21	-2.20
E_2	2.11	2.06	2.06	1.98	1.97	1.96	1.96	2.05	1.99	1.99
π_2	-2.27	-2.02	-2.03	-1.93	-1.92	-1.91	-1.91	-2.00	-1.99	-1.97
E_3	0.66	0.64	0.63	0.50	0.50	0.49	0.51	0.51	0.65	0.65
π_3	-1.05	-1.08	-1.07	-0.61	-0.64	-0.60	-0.63	-0.62	-0.94	-0.95
E_4	2.57	2.60	2.63	2.79	2.84	2.75	2.76	2.82	2.71	2.69
π_4	-2.37	-2.36	-2.37	-2.42	-2.44	-2.40	-2.40	-2.46	-2.38	-2.37
E_5	2.87	2.85	2.84	3.00	2.97	3.00	2.98	2.79	2.82	2.80
π_5	-2.46	-2.44	-2.45	-2.47	-2.45	-2.46	-2.45	-2.38	-2.43	-2.43
E_6	1.97	1.96	1.94	2.01	2.00	2.02	2.00	1.92	1.98	1.95
π_6	-2.11	-2.12	-2.11	-2.14	-2.12	-2.15	-2.14	-2.08	-2.06	-2.06
E_7	3.39	3.39	3.39	4.30	4.32	4.33	4.33	3.48	3.62	3.57
π_7	-2.69	-2.69	-2.69	-2.82	-2.82	-2.82	-2.82	-2.63	-2.65	-2.64
E_8	0.18	0.18	0.18	0.22	0.22	0.22	0.22	0.21	0.18	0.19
π_8	0.21	0.18	0.19	0.10	0.10	0.09	0.09	0.14	0.14	0.14
E_9	0.08	0.08	0.08	0.10	0.10	0.10	0.11	0.10	0.08	0.08
π_9	0.99	0.95	0.95	0.89	0.89	0.89	0.88	0.92	0.76	0.76
E_{10}	5.50	5.34	5.33	5.28	5.29	5.23	5.25	5.42	5.24	5.22
π_{10}	-3.26	-3.03	-3.04	-2.96	-2.95	-2.94	-2.94	-3.03	-3.07	-3.04
E_{11}	1.52	1.49	1.48	1.45	1.44	1.44	1.43	1.51	1.43	1.43
π_{11}	-1.89	-1.70	-1.70	-1.68	-1.67	-1.65	-1.65	-1.74	-1.67	-1.65
E_{12}	1.77	1.76	1.77	1.81	1.80	1.81	1.81	1.76	1.84	1.83
π_{12}	-2.03	-2.04	-2.04	-2.06	-2.06	-2.06	-2.06	-2.02	-2.06	-2.05
E_{13}	4.71	4.63	4.66	5.37	5.46	5.38	5.42	5.69	4.75	4.74
π_{13}	-3.13	-2.94	-2.96	-3.12	-3.11	-3.07	-3.08	-3.13	-3.10	-3.08
E_{14}	0.46	0.46	0.46	0.47	0.48	0.46	0.48	0.46	0.46	0.46
π_{14}	-0.64	-0.65	-0.66	-0.65	-0.67	-0.65	-0.67	-0.61	-0.65	-0.64
E_{15}	9.56	9.55	9.53	12.05	12.07	12.07	12.08	10.04	9.61	9.61
π_{15}	-3.73	-3.74	-3.73	-3.90	-3.90	-3.89	-3.90	-3.71	-3.69	-3.69
E_{16}	0.79	0.81	0.80	0.91	0.91	0.92	0.92	0.78	0.84	0.84
π_{16}	-1.27	-1.30	-1.29	-1.32	-1.32	-1.33	-1.33	-1.22	-1.28	-1.27
E_{17}	2.83	2.77	2.79	3.26	3.27	3.27	3.27	2.82	2.77	2.77
π_{17}	-2.47	-2.46	-2.47	-2.62	-2.63	-2.63	-2.62	-2.46	-2.43	-2.44
E_{18}	4.45	4.38	4.42	5.26	5.33	5.26	5.30	4.52	4.50	4.50
π_{18}	-2.91	-2.87	-2.88	-3.09	-3.10	-3.09	-3.08	-2.92	-2.91	-2.91
E_{19}	2.58	2.59	2.57	2.55	2.57	2.55	2.55	2.49	2.48	2.48
π_{19}	-2.39	-2.41	-2.40	-2.40	-2.43	-2.40	-2.42	-2.36	-2.38	-2.38
E_{20}	6.41	6.38	6.38	6.49	6.49	6.45	6.46	6.50	6.34	6.30
π_{20}	-3.28	-3.19	-3.20	-3.26	-3.26	-3.21	-3.21	-3.27	-3.22	-3.21

Table 4: Comparison with Other Approaches. $\pi_{1:20,h}$ is the average log predictive score for horizon h for the whole vector $(y_{1t}, ..., y_{20,t})$. π_i is the average log predictive score for the first horizon for y_i. E_i is the RMSE when predicting y_{it} in the first horizon.

6.3 Monte Carlo Experiment

The data generating process is given by the following heteroscedastic approximate factor model:

$$\begin{aligned} y_t &= \beta y_{t-1} + \Gamma u_{1t} + \sqrt{\theta}\widetilde{u}_{2t} \\ u_{1t} &: r_1 \times 1,\ var(u_{1t}) = \Upsilon_t^{-1},\ E(\Upsilon_t^{-1}) = I_{r_1},\ \Upsilon_t \sim WAR(1) \\ \widetilde{u}_{2t} &: r \times 1, \quad var(\widetilde{u}_{2t}) = \Delta = (\delta_{ij}),\ \delta_{ij} = 0.5^{|i-j|},\ \theta : 1 \times 1 \\ G &= var(\Gamma u_{1t} + \sqrt{\theta}\widetilde{u}_{2t}) = \Gamma\Gamma' + \theta\Delta \end{aligned} \tag{22}$$

with $r_1 = 2$ and the parameters of the WAR(1) process fixed as $n = 8$ and $\rho = 0.95 I_{r_1}$. The matrix β is 0 in the data generating process, but estimated for each dataset. For each dataset generated, each of the elements of the matrix of factor loadings Γ is generated using iid $N(0,1)$ variates. Note that the elements of $\widetilde{u}_{2t}$ are correlated with each other and that its dimension is $r \times 1$, which is bigger than the dimension of u_{2t} in our model. Recalling the singular value decomposition of G in Section 2 ($G = U_1 S_1 U_1' + U_2 S_2 U_2'$), model (22) becomes equivalent to our model when $\Gamma\Gamma' = U_1 S_1 U_1'$. This will be the case as r gets larger relative to r_1, or θ decreases (Chamberlain and Rothschild (1983)). The degree of misspecification of our model with respect to (22) can be captured with two measures: d_1^2 and d_2. d_1^2 is the squared distance between the spaces $sp(\Gamma)$ and $sp(U_1)$: $d_1^2 = d^2(sp(\Gamma)), sp(U_1))$, as defined by Larsson and Villani (2001) and normalized to be between 0 and 100. d_2 is the ratio of the absolute value of the (1,1) element of the matrix $(U_1 S_1 U_1' - \Gamma\Gamma')$ over the (1,1) element of G. It measures the asymptotic (i.e. for large T) mean absolute error in estimating the proportion of the variance that changes with time.

The prior is the same as in the empirical application to US data. We consider the cases of $r = 10, 20, 30, 40$ and $T = 250, 500$. As a baseline we assume $\theta = r_1 = 2$, which implies that (unconditionally on Γ) the common (heteroscedastic) components have the same variances as the idiosyncratic (homoscedastic) components (as in the simulations of Bai and Ng (2002)), but we also illustrate the case of $\theta = 0.1$ for $r = 10$. Table 5 gives the average values (calculated with 50000 replications) of d_1^2 and d_2, which indicates that even for $r = 10$ the degree of misspecification is relatively small, with $E(d_1^2) = 4.4$ and $E(d_2) = 0.11$ for ($r = 10$, $\theta = 2$), and quickly decreases with increasing r, such that for example $E(d_1^2) = 0.78$ and $E(d_2) = 0.05$ for ($r = 20$, $\theta = 2$). Decreasing θ from 2 to 0.1 has a similar effect on (d_1^2, d_2) as increasing r from 10 to 40 (with $E(d_1^2)$ about 0.175 and E(d_2) about 0.015).

For each data configuration, we simulate 100 datasets, and apply our methodology to each of them. Table 5 shows the average estimate of r_1 when it is estimated using the BIC ($\hat{r}_1^{BIC}$), the Marginal Likelihood ($\hat{r}_1^{ML}$) described in Section 5.6, by Bayesian model averaging with BIC weights ($\hat{r}_1^{BMA,BIC}$)

and with weights calculated using the Marginal Likelihood ($\hat{r}_1^{BMA,ML}$). Using the Marginal Likelihood gives better results than the BIC. The best performance is provided by $\hat{r}_1^{BMA,ML}$, which gives an average estimate (1.77) close to the true value (2) even in the most difficult case of ($r = 10$, $\theta = 2$, $T = 250$). As expected, the performance improves as r or T increases or θ decreases. Table 5 also shows the proportion of times that each model was chosen under each criterion ($\hat{\pi}_{r_1}^{BIC}$, $\hat{\pi}_{r_1}^{ML}$). As a comparison, the criteria IC_{p1} and PC_{p1} of Bai and Ng (2002) only perform well in the case of $r = 40$.

We also report the bias, mean absolute error (MAE), and coverage of 90%, 95%, and 99% credible intervals[18] for the following six parameters. g_{11} is the (1,1) element of G. ϱ_{12} is the correlation defined as $\varrho_{12} = g_{12}/\sqrt{g_{11}g_{22}}$. p_{11} is the ratio of the (1,1) element of $U_1 S_1 U_1'$ over g_{11} (i.e. proportion of the variance of the first variable that is time varying). $\sigma_{11,t}$ is the (1,1) element of Σ_t, with t being the integer nearest to $T/3$. $\varrho_{12,t}$ is the correlation defined as $\varrho_{12,t} = \sigma_{12,t}/\sqrt{\sigma_{11,t}\sigma_{22,t}}$. $\overline{\sigma}_{11}$ is the average of $\sigma_{11,t}$: $\overline{\sigma}_{11} = (1/T)\sum_{t=1}^{T}\sigma_{11,t}$. The bias and MAE are calculated as a percentage of the true value in the cases of non-bounded parameters, namely g_{11}, $\sigma_{11,t}$ and $\overline{\sigma}_{11}$.

In addition to the performance of the BMA estimates (calculated with ML weights), we report the performance of estimates calculated assuming that $r_1 = 0$. Table 6 for $T = 250$ and Table 7 for $T = 500$ show that the actual coverage is close to the nominal one for g_{11}, $\sigma_{11,t}$ and $\overline{\sigma}_{11}$ in all cases, except only for g_{11} when $T = 250$, $r = 40$, in which case it is slightly undersized. The actual coverages for the correlations ϱ_{12}, $\varrho_{12,t}$ are also close to the nominal ones in almost all cases, but are slightly undersized in the case of $r = 10$, $T = 250$ for ϱ_{12} and $r = 10$, $T = 250$, 500 for $\varrho_{12,t}$. The credible intervals of p_{11} are undersized in all cases, but the bias and MAE are relatively small. Using any of the measures of coverage, bias or MAE, the performance of the BMA estimates is much better than for the estimates based on $r_1 = 0$, which illustrates the importance of accounting for heteroscedasticity

[18]This is the percentage of times that the posterior credible interval, calculated with the posterior quantiles, contains the true value.

	$\theta = 2$								$\theta = 0.1$	
	$T = 250$				$T = 500$				T=250	T=500
r	10	20	30	40	10	20	30	40	10	10
$\hat{r}_1^{ML}$	1.63	1.93	2.07	2.01	1.79	2.05	2.04	2.01	2.01	2.00
$\hat{r}_1^{BIC}$	0.93	1.71	1.86	1.92	1.50	1.97	1.95	1.98	1.66	2.00
$\hat{r}_1^{BMA,ML}$	1.77	1.95	2.07	2.00	1.85	2.06	2.04	2.01	1.98	2.00
$\hat{r}_1^{BMA,BIC}$	0.93	1.69	1.83	1.90	1.49	1.97	1.96	1.99	1.65	2.00
$\hat{\pi}_0^{ML}$	0.11	0.01	0.00	0.00	0.05	0.00	0.00	0.00	0.01	0.00
$\hat{\pi}_1^{ML}$	0.25	0.12	0.02	0.01	0.18	0.05	0.02	0.01	0.05	0.00
$\hat{\pi}_2^{ML}$	0.58	0.80	0.92	0.97	0.71	0.93	0.93	0.97	0.87	1.00
$\hat{\pi}_3^{ML}$	0.04	0.07	0.03	0.02	0.05	0.02	0.04	0.02	0.05	0.00
$\hat{\pi}_4^{ML}$	0.00	0.00	0.03	0.00	0.01	0.00	0.01	0.00	0.00	0.00
$\hat{\pi}_0^{BIC}$	0.39	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.13	0.00
$\hat{\pi}_1^{BIC}$	0.29	0.31	0.21	0.17	0.19	0.05	0.05	0.04	0.09	0.00
$\hat{\pi}_2^{BIC}$	0.32	0.67	0.73	0.75	0.64	0.93	0.95	0.94	0.77	1.00
$\hat{\pi}_3^{BIC}$	0.00	0.02	0.05	0.07	0.01	0.02	0.00	0.02	0.01	0.00
$\hat{\pi}_4^{BIC}$	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
PC_{p1}	8.00	7.87	7.69	7.18	8.00	8.00	7.74	7.08	8.00	8.00
IC_{p1}	8.00	7.49	4.84	2.45	8.00	7.91	4.95	2.04	8.00	8.00
$E(d_1^2)$	4.40	0.78	0.31	0.17	4.40	0.78	0.31	0.17	0.18	0.18
$E(d_2)$	0.11	0.05	0.03	0.02	0.11	0.05	0.03	0.02	0.01	0.01

Table 5: Average estimated values for r_1 using 100 simulations ($\hat{r}_1^{ML}$, $\hat{r}_1^{BIC}$, $\hat{r}_1^{BMA,ML}$, $\hat{r}_1^{BMA,BIC}$, PC_{p1},IC_{p1}), and proportion of times that each model was chosen ($\hat{\pi}_{r_1}^{ML}$, $\hat{\pi}_{r_1}^{BIC}$). $E(d_1^2)$ and $E(d_2)$ are measures of model misspecification, and are calculated with 50000 replications.

		$\theta = 2$								$\theta = 0.1$	
		$r = 10$		$r = 20$		$r = 30$		$r = 40$		$r = 10$	
		$r_1 = 0$	BMA	$r_1 = 0$	BMA	$r_1 = 0$	BMA	$r_1 = 0$	BMA	$r_1 = 0$	BMA
g_{11}	c_{99}	90	98	82	98	79	99	75	96	80	100
	c_{95}	76	92	66	92	74	92	65	83	65	96
	c_{90}	70	88	54	83	68	84	50	70	58	87
	Bias (%)	-6.0	-2.0	-5.3	4.0	-10.6	6.1	-12.4	14.4	-5.8	-0.8
	MAE (%)	10.9	10.0	14.3	13.5	12.5	11.6	14.5	18.7	13.7	14.1
ϱ_{12}	c_{99}	82	93	94	100	95	98	94	97	86	99
	c_{95}	77	81	87	96	87	94	88	94	72	96
	c_{90}	73	79	76	94	81	89	81	86	64	91
	Bias	0.003	0.002	0.009	0.009	0.003	-0.005	0.000	-0.013	0.010	0.012
	MAE	0.064	0.065	0.048	0.048	0.053	0.057	0.056	0.062	0.058	0.053
p_{11}	c_{99}	0	94	0	84	0	90	0	89	0	89
	c_{95}	0	83	0	64	0	80	0	72	0	77
	c_{90}	0	74	0	55	0	71	0	62	0	63
	Bias	-0.43	0.06	-0.73	0.07	-0.40	0.05	-0.39	0.08	-0.89	-0.01
	MAE	0.43	0.13	0.73	0.07	0.40	0.07	0.39	0.09	0.89	0.03
$\sigma_{11,t}$	c_{99}	64	98	43	100	70	99	64	98	37	99
	c_{95}	50	94	37	97	56	96	47	94	29	93
	c_{90}	45	90	34	91	48	92	41	85	24	85
	Bias (%)	1.97	-2.85	12.01	3.14	-4.73	1.57	-4.38	1.91	33.60	12.07
	MAE (%)	19.90	15.82	32.12	22.05	18.07	14.82	21.96	17.05	51.33	33.97
$\varrho_{12,t}$	c_{99}	65	95	51	99	72	98	71	99	43	100
	c_{95}	49	79	43	95	61	94	60	97	30	97
	c_{90}	44	74	39	92	55	91	54	92	24	85
	Bias	-0.04	-0.03	0.03	0.02	-0.02	0.00	-0.01	0.00	0.00	-0.01
	MAE	0.13	0.10	0.12	0.10	0.10	0.07	0.11	0.09	0.14	0.11
$\bar{\sigma}_{11}$	c_{99}	97	99	89	99	88	99	77	100	92	97
	c_{95}	90	93	78	94	77	95	69	96	85	90
	c_{90}	80	90	69	89	69	89	54	89	80	87
	Bias (%)	-5.95	-2.95	-7.01	-0.62	-10.55	1.30	-12.94	3.47	-5.93	-2.78
	MAE (%)	8.56	7.49	10.46	8.16	11.22	7.65	13.40	9.19	9.75	8.48

Table 6: Performance of estimates with $T = 250$. c_α is the coverage of an α% posterior credible interval, measured as the number of times that the interval contained the true value over 100 simulations. The columns labeled as BMA refer to the performance of BMA estimates, and those labeled with $r_1 = 0$ refer to the estimates from the model that assumes $r_1 = 0$

		$\theta = 2$								$\theta = 0.1$	
		$r = 10$		$r = 20$		$r = 30$		$r = 40$		$r = 10$	
		$r_1 = 0$	BMA	$r_1 = 0$	BMA	$r_1 = 0$	BMA	$r_1 = 0$	BMA	$r_1 = 0$	BMA
g_{11}	c_{99}	96	98	90	96	88	99	86	100	75	100
	c_{95}	83	97	81	92	77	94	71	96	63	97
	c_{90}	77	90	71	88	72	87	61	91	57	94
	Bias (%)	-2.27	-0.66	-3.15	0.84	-5.00	2.46	-7.45	2.09	1.14	1.15
	MAE (%)	6.86	6.63	7.39	6.32	7.76	7.30	8.61	6.52	10.76	9.53
ϱ_{12}	c_{99}	95	96	94	98	97	100	97	99	82	99
	c_{95}	92	93	82	92	86	95	87	93	72	98
	c_{90}	82	90	75	87	77	89	76	84	64	95
	Bias	0.00	-0.01	0.00	0.00	0.01	0.00	-0.01	-0.01	-0.01	0.00
	MAE	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
p_{11}	c_{99}	0	76	0	80	0	87	0	98	0	72
	c_{95}	0	59	0	63	0	75	0	89	0	59
	c_{90}	0	53	0	50	0	64	0	84	0	51
	Bias	-0.38	0.07	-0.41	0.05	-0.42	0.04	-0.34	0.03	-0.87	0.01
	MAE	0.38	0.13	0.41	0.07	0.42	0.05	0.34	0.04	0.87	0.02
$\sigma_{11,t}$	c_{99}	58	95	50	98	43	98	66	99	75	100
	c_{95}	47	92	40	89	38	92	56	96	63	97
	c_{90}	38	88	35	83	33	86	47	90	57	94
	Bias (%)	2.3	-1.2	6.9	1.9	2.6	-1.9	-5.5	-1.5	1.1	1.1
	MAE (%)	17.8	13.6	23.7	16.5	21.8	15.2	15.8	11.0	10.8	9.5
$\varrho_{12,t}$	c_{99}	57	92	62	99	61	98	63	99	82	99
	c_{95}	41	84	48	92	54	95	53	93	72	98
	c_{90}	35	83	39	88	43	92	46	88	64	95
	Bias	-0.01	-0.01	-0.02	-0.01	-0.01	0.00	0.01	0.02	-0.01	0.00
	MAE	0.11	0.09	0.11	0.09	0.09	0.07	0.10	0.08	0.04	0.04
$\bar{\sigma}_{11}$	c_{99}	97	98	99	99	93	99	92	98	94	98
	c_{95}	90	93	89	97	84	95	73	95	85	93
	c_{90}	84	88	83	93	76	87	59	87	78	89
	Bias (%)	-3.1	-1.7	-4.3	-0.9	-5.2	0.7	-7.3	0.5	-1.2	-0.4
	MAE (%)	5.6	5.2	6.3	4.9	6.7	5.6	8.4	5.7	7.0	6.1

Table 7: Performance of estimates with $T = 500$. c_α is the coverage of an α% posterior credible interval, measured as the number of times that the interval contained the true value over 100 simulations. The columns labeled as BMA refer to the performance of BMA estimates, and those labeled with $r_1 = 0$ refer to the estimates from the model that assumes $r_1 = 0$.

7 Concluding remarks

In this paper we have developed methods for Bayesian inference in an inverted Wishart process for Stochastic Volatility. 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 an 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 shocks and their impacts on the variables of interest. We find strong evidence of co-heteroscedasticity in an application to a large VAR of 20 macroeconomic variables, and found that our model performs better than previous methods in terms of forecasting performance from horizon 3 onward. A Monte Carlo experiment indicates that our model recovers well the characteristics of approximate factor models with heteroscedastic factors.

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 (2003) "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.

Asai, M. and M. McAleer (2015) "Forecasting co-volatilities via factor models with asymmetry and long memory in realized covariance," *Journal of Econometrics*, 189, 251-262.

Bai, J., and S. Ng, (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

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.

Bauwens, L., S. Laurent AND J. V. K. Rombouts (2006) "Multivariate GARCH Models: A Survey," *Journal of Applied Econometrics,* 21, 79-109.

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

Carriero, A., Clark, T.E. and Marcellino, M. (2016) "Common Drifting Volatility in Large Bayesian VARs," *Journal of Business Economic Statistics*, 34, 375-390.

Carriero, A., Clark, T.E. and Marcellino, M. (2019) "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212, 137-154.

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.

Chamberlain, G., and M. Rothschild (1983) "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1305–1324.

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

Chan, J. (2020) "Large Bayesian VARs: A Flexible Kronecker Error Covariance Structure," *Journal of Business & Economic Statistics* , 38, 68-79.

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

Chib, S., F. Nardari, and N. Shephard (2006) "Analysis of high dimensional multivariate stochastic volatility models," *Journal of Econometrics*, 134, 2, 341-371.

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.

Diebold, F. and M. Nerlove, (1989). "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model".*Journal of Applied Econometrics*, 4, 1, 1-21.

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.

Engle, R. and S. Kozicki (1993) "Testing for Common Features," *Journal of Business & Economic Statistics*, 11, 4, 369-380

Engle, R.F. and R. Susmel (1993), "Common Volatility in International Equity Markets," *Journal of Business & Economic Statistics*, 11, 2, 167-176.

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.

Gonzalo, J. and S. Ng (2001) "A systematic framework for analyzing the dynamic e!ects of permanent and transitory shocks" *Journal of Economic Dynamics Control* 25, 1527-1546
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. (1997) *ARCH Models and Financial Applications.* Springer-Verlag: New York.

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.

Hernández-Trillo, F. (1997) "Re-exploring financial diversification: The Mexican case," In The Global Structure of Financial Markets: An Overview, edited by Dilip K. Ghosh and Edgar Ortiz, Routledge.

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

Kastner, G. (2019) "Sparse Bayesian Time-Varying Covariance Estimation in Many Dimensions," *Journal of Econometrics*, 210, 98-115.

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.

King, M., E. Sentana and S. Wadhwani (1994) "Volatility and Links between National Stock Markets," *Econometrica*, 62, 4 901-933.

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

Koop, G., Pesaran, M. H. and S. M. Potter (1996), "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 74, 119–147.

Larsson, R. and Villani, M. (2001), "A distance measure between cointegration spaces," *Economic Letters*,70, 21-27

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.

Pajor, A. (2006), â€œBayesian Analysis of the Conditional Correlation Between Stock Index Returns with Multivariate Stochastic Volatility Models,â€ *Acta Physica Polonica B*, 37, 3093â€“3103.

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.

Triantafyllopoulos, K. (2012) "Multi-variate stochastic volatility modelling using Wishart autoregressive processes" *Journal of Time Series Analysis*, 33, 48-60.

Uhlig, H. (1997) "Bayesian Vector Autoregressions with Stochastic Volatility," *Econometrica*, 65, 59-73.

Uhlig, H. (2005). "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure." *Journal of Monetary Economics*, 52, 381–419

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.

Yu, J. and R. Meyer (2006) "Multivariate Stochastic Volatility Models: Bayesian Estimation and Model Comparison," *Econometric Reviews* 25, 361-384.

Appendix I

This appendix is for section 4, and derives the pseudo posterior and the compact expression for the likelihood in (10). The prior $\pi(Z_{1:T}|\theta)$ is:

$$(2\pi)^{-Tnr/2}\left|I-\rho^2\right|^{n/2}\exp\left(-\frac{1}{2}tr\left(\sum_{t=2}^{T}\left(Z_t-Z_{t-1}\rho\right)'\left(Z_t-Z_{t-1}\rho\right)+\left(Z_1'Z_1\right)\left(I-\rho^2\right)\right)\right) \tag{23}$$

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

Given the definition of Σ_t in (3), and defining $\widetilde{A}=(\widetilde{A}_1,A_2)$, the inverse Σ_t^{-1} can be written as:

$$\Sigma_t^{-1}=\left(\widetilde{A}^{-1}\right)'\begin{pmatrix}K_t & 0\\ 0 & I\end{pmatrix}\widetilde{A}^{-1}=B_1K_tB_1'+B_2B_2', \qquad \text{for } B=\left(\widetilde{A}^{-1}\right)'=\begin{pmatrix}B_1 & B_2\end{pmatrix} \tag{24}$$

where $K_t=Z_t'Z_t$. Thus, when we multiply (23) times the pseudo likelihood in (7) we obtain:

$$\begin{aligned}&\pi(Z_{1:T}|\theta)\widetilde{L}(Y|\Sigma_{1:T},\beta)= \qquad (25)\\ =\ &\exp\left(-\frac{1}{2}tr\left(\sum_{t=1}^{T-1}\left(\rho' Z_t'Z_t\rho\right)+\sum_{t=2}^{T}\left(Z_t'Z_t\right)\left(I+(1-\delta)B_1'e_te_t'B_1\right)-2\sum_{t=2}^{T}\left(Z_t'Z_{t-1}\rho\right)\right)\right)\times\\ &\exp\left(-\frac{1}{2}tr\left(\left(\left(I-\rho^2\right)+(1-\delta)B_1'e_1e_1'B_1\right)Z_1'Z_1\right)\right)(2\pi)^{-Tnr/2}\left|I-\rho^2\right|^{n/2}c_{\widetilde{B}}\end{aligned}$$

where:

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

The terms that depend on Z_T in (25) are $\exp(-0.5tr((Z_T'Z_T)(I+(1-\delta)B_1'e_Te_T'B_1)-2(Z_T'Z_{T-1}\rho)))$ and so $vec(Z_T)|Z_{1:T-1}$ is a normal with mean equal to $vec(Z_{T-1}\rho V_T)$ and variance $V_T\otimes I$, with $V_T=(I+(1-\delta)B_1'e_Te_T'B_1)^{-1}$. Integrating out Z_T from (25) we obtain:

$$
\begin{aligned}
& \int \pi(Z_{1:T}|\theta)\widetilde{L}(Y|\Sigma_{1:T},\beta)dZ_T \\
= \;& (2\pi)^{nr/2}\left|V_T\right|^{n/2}\exp(\frac{1}{2}tr(\rho' Z'_{T-1}Z_{T-1}\rho V_T))\times \\
& \exp\left(-\frac{1}{2}tr\left(\sum_{t=1}^{T-1}(\rho' Z'_t Z_t\rho)+\sum_{t=2}^{T-1}(Z'_tZ_t)\left(I+(1-\delta)B'_1e_te'_tB_1\right)-2\sum_{t=2}^{T-1}(Z'_tZ_{t-1}\rho)\right)\right)\times \\
& \exp\left(-\frac{1}{2}tr\left(\left((I-\rho^2)+(1-\delta)B'_1e_1e'_1B_1\right)Z'_1Z_1\right)\right)(2\pi)^{-Tnr/2}\left|I-\rho^2\right|^{n/2}c_{\widetilde{B}}
\end{aligned}
$$

The terms that depend on Z_{T-1} are $\exp(-0.5tr(\left(Z'_{T-1}Z_{T-1}\right)(I+(1-\delta)B'_1e_{T-1}e'_{T-1}B_1+\rho\rho')-2\left(Z'_{T-1}Z_{T-2}\rho\right)-(Z'_{T-1}Z_{T-1}\rho V_T\rho')))$ and so $vec(Z_{T-1})|Z_{1:T-2}$ is a normal with mean equal to $vec(Z_{T-2}\rho V_{T-1})$ and variance $V_{T-1}\otimes I$, with $V_{T-1}=(I+(1-\delta)B'_1e_{T-1}e'_{T-1}B_1+\rho(I-V_T)\rho')^{-1}$. This process can be repeated recursively to find the densities described in (8) and (9) 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:

$$
\begin{aligned}
L(Y|\beta,\theta) &= \int\left(\prod_{t=1}^{T}\left|\Sigma_t\right|^{-1/2}\right)\exp(-\frac{\delta}{2}\sum_{t=1}^{T}tr(\Sigma_t^{-1}e_te'_t))\pi(Z_{1:T}|\theta)\widetilde{L}(Y|\Sigma_{1:T},\beta)dZ_{1:T} \\
&= \int\left(\prod_{t=1}^{T}\left|\Sigma_t^{-1}\right|^{1/2}\exp(-\frac{\delta}{2}tr(B_1K_tB'_1e_te'_t))\right)\exp(-\frac{\delta}{2}\sum_{t=1}^{T}tr(B_2B'_2e_te'_t))\widetilde{c}\widetilde{\pi}(Z_{1:T}|Y,\theta)dZ_{1:T} \\
&= \int\left(\prod_{t=1}^{T}\frac{\left|K_t\right|^{1/2}}{\exp(\frac{\delta}{2}tr(K_tB'_1e_te'_tB_1))}\right)\left(\left|\widetilde{A}_1\widetilde{A}'_1+A_2A'_2\right|^{-T/2}\right)\frac{\widetilde{c}\widetilde{\pi}(Z_{1:T}|Y,\theta)}{\exp(\frac{\delta}{2}\sum_{t=1}^{T}tr(B_2B'_2e_te'_t))}dZ_{1:T} \\
&= \widetilde{c}_LE_{\widetilde{\pi}}\left(\prod_{t=1}^{T}\left[\left|K_t\right|^{1/2}\exp(-\frac{\delta}{2}tr(K_tB'_1e_te'_tB_1))\right]\right)
\end{aligned}
$$

which is equal to (10).

Appendix II

This appendix is for Section 5.2. It derives the reverse order representation in (13) - (14) 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 (13), 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:

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

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:

$$\begin{aligned} L_{1t} &= \overrightarrow{Z}_{t-1}'Z_t = \underbrace{\left(\overrightarrow{Z}_{t-1}\right)'Z_{t-1}}_{K_{t-1}^{1/2}}\rho + \underbrace{\left(\overrightarrow{Z}_{t-1}\right)'\varepsilon_t}_{\varepsilon_{1t}} \qquad (26) \\ L_{1t} &= K_{t-1}^{1/2}\rho + \varepsilon_{1t}, \quad \varepsilon_{1t} : r_1 \times r_1, \quad vec(\varepsilon_{1t}) \sim N(0, I_{r_1} \otimes I_{r_1}),\ t > 1 \end{aligned}$$

and that:

$$\begin{aligned} L_{2t} &= \overrightarrow{Z}_{t-1,\perp}'Z_t = \underbrace{\left(\overrightarrow{Z}_{t-1,\perp}\right)'Z_{t-1}\rho}_{0} + \left(\overrightarrow{Z}_{t-1,\perp}\right)'\varepsilon_t = \underbrace{\left(\overrightarrow{Z}_{t-1,\perp}\right)'\varepsilon_t}_{\varepsilon_{2t}} \\ L_{2t} &= \varepsilon_{2t}, \qquad vec(L_{2t}) \sim N(0, I_{r_1} \otimes I_{n-r_1}) \\ K_{2t} &= L_{2t}'L_{2t} \sim W_r(n - r_1, I_{r_1}) \end{aligned}$$

Note that the process $Z_t = Z_{t-1}\rho + \varepsilon_t$ with Z_1 following the stationary distribution can be equivalently written as $Z_{t-1} = Z_t\rho + \varepsilon_t$ with Z_T following the stationary distribution[19]. Therefore we can apply the same decomposition to the reverse process as follows:

$$\begin{aligned} \widetilde{L}_{1,t-1} &= \overrightarrow{Z}_t'Z_{t-1} = \underbrace{\left(\overrightarrow{Z}_t\right)'Z_t\rho}_{K_t^{1/2}} + \underbrace{\left(\overrightarrow{Z}_t\right)'\varepsilon_t}_{\varepsilon_{1t}} \qquad (27) \\ \widetilde{L}_{2,t-1} &= \overrightarrow{Z}_{t\perp}'Z_{t-1} = \underbrace{\left(\overrightarrow{Z}_{t\perp}\right)'Z_t\rho}_{0} + \left(\overrightarrow{Z}_{t\perp}\right)'\varepsilon_t = \underbrace{\left(\overrightarrow{Z}_{t-1,\perp}\right)'\varepsilon_t}_{\varepsilon_{2t}} \\ \widetilde{K}_{2t} &= \widetilde{L}_{2t}'\widetilde{L}_{2t} \end{aligned}$$

[19]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 :

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

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.

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 (13) and the inverse transformation (17).

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 (15), 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:

$$w_T^k := \frac{\left|K_T^k\right|^{0.5}}{\exp\left(\frac{\delta}{2}tr(K_T^k B_1' e_T e_T' B_1)\right)},\ W_T^k := w_T^k/(\sum_{m=1}^{N} w_T^m),\ k = 1, ..., N$$

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* $\mathbf{W}_{t+1} = (W_{t+1}^1, ..., W_{t+1}^N)$

(b) sample $vec(\widetilde{L}_{1t}^k) \sim N(\left(K_{t+1}^{A_{t+1}^k}\right)^{1/2}\rho\widetilde{V}_t, I_{r_1} \otimes \widetilde{V}_t)$ *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

$$w_t^k := \frac{\left|K_t^k\right|^{0.5}}{\exp\left(\frac{\delta}{2}tr(K_t^k B_1' e_t e_t' B_1)\right)}, \quad W_t^k := w_t^k/(\sum_{m=1}^{N} w_t^m),\ k = 1, ..., N$$

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

$$\begin{aligned}
\widetilde{w}_t^k &= w_t^k f(K_{t-1}^{b_{t-1}}|K_t^k), ..., \ \widetilde{W}_t^k := \widetilde{w}_t^k/(\sum_{m=1}^{N} \widetilde{w}_t^m),\ k = 1, ..., N, \text{ where} \\
\mu_t^k &= (K_t^k)^{1/2}\rho\widetilde{V}_{t-1}, \quad f(K_{t-1}^{b_{t-1}}|K_t^k) = \exp(-\frac{1}{2}tr(\widetilde{V}_{t-1}^{-1}(\widetilde{L}_{1(t-1)}^{b_{t-1}} - \mu_t^k)'(\widetilde{L}_{1(t-1)}^{b_{t-1}} - \mu_t^k))
\end{aligned}$$

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

Table 8: 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, n equal to its posterior median and $\rho = 0.8I_{r_1}$.

(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, n equal to the posterior median and $\rho = 0.8I_{r_1}$ 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 8). 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 6.9 and at most by a factor of 45.7. 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 1.8 and 2.1. Looking at the ESS of $K_{T/2}$ (i.e. at the middle of the sample), the alternating order algorithm has always better ESS (Table 8), with the factor of improvement ranging between 1.01 (with 110 particles) and 1.55 (with 25 particles). To get a measure of which algorithm is overall better to sample $K_{1:T}$, we calculate the ESS of $\overline{K} = (K_1 + ... + K_T)/T$. We find that the alternating order algorithm is the best in all cases, increasing the ESS by at most 61% and at least by 14%. Overall we can conclude that the ordering in sampling $K_{1:T}$ affects the efficiency of the algorithm, and that for our dataset alternating the ordering brings 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 $t = 2, ..., T$,

(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 $d_t = (d_{1t}, d_{2t}, ..., d_{r_1 t})$ *be the diagonal elements of* Δ_t. *Let* $p_t = (p_{1t}, p_{2t}, ..., p_{r_1 t})$, *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 (9) and the value of* $\theta^* = (G^*, \rho^*, n^*)$.

Step 2: at time $t = 1$

(a) Obtain the vector $d_t^{n^*} = (d_{1t}^{n^*}, d_{2t}^{n^*}, ..., d_{r_1 t}^{n^*})$, *by calculating* $d_{it}^{n^*} = \left(F^{-1}_{\chi^2(f_{it}^{n^*})}(p_{it})\right)^{1/2}$, *where* $F^{-1}_{\chi^2(f_{it}^{n^*})}$ *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 $d_t^{n^*} = (d_{1t}^{n^*}, d_{2t}^{n^*}, ..., d_{r_1 t}^{n^*})$, *by calculating* $d_{it}^{n^*} = \left(F^{-1}_{\chi^2(f_{it}^{n^*})}(p_{it})\right)^{1/2}$, *where* $F^{-1}_{\chi^2(f_{it}^{n^*})}$ *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:*

$$\pi^t(\theta|\epsilon, \beta) \propto \pi(L_{1:T}^{\theta}|\theta) L(Y|K_{1:T}^{\theta}, \beta) J_\varepsilon \pi(\theta) \pi(\beta|\theta) \tag{28}$$

where

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

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

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

and then use the expressions for the multivariate normal and Wishart densities with parameters specified as in (12). 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 $\left(V_t^\theta\right)^{(r_1+1)/2}$ 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 $\left(V_t^\theta\right)^{r_1/2}$. ■

In our estimations we use a normal inverse Wishart prior for G, and a normal prior for $\widetilde{n} = \log(n - 2r_1)$ and a beta distribution for ρ. However, in order to perform the Metropolis step we target the conditional posterior of $(G, \widetilde{n}, \widetilde{\rho})$, where $\widetilde{\rho} = \ln(-\ln(1-\rho^2))$. The prior for $\pi(\widetilde{\rho})$ can be obtained as:

$$\begin{aligned}\pi(\widetilde{\rho}) &= \pi(\rho)J_\rho \\ J_\rho &= \frac{1-\rho^2}{2\rho}(-\ln(1-\rho^2))\end{aligned}$$

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

Figure 5 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.

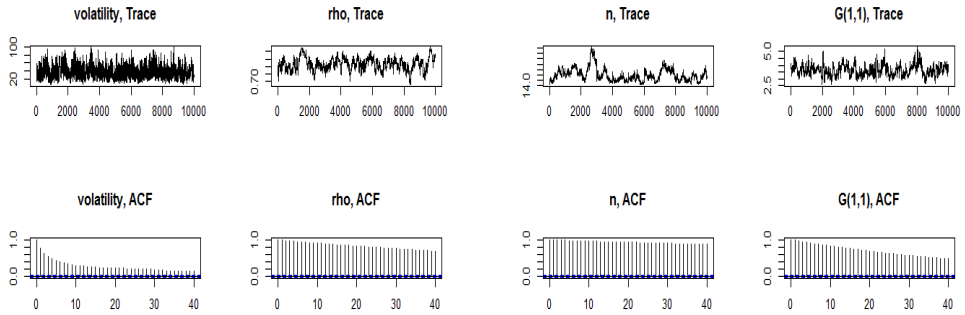

Figure 5: Trace plot and autocorrelations when no reparameterization is used in the model with $r_1 = 7$ (Macro data). Trace plots and autocorrelations are for the (7,7) element of $K_{T/2}$, the 4^{th} diagonal element of ρ, for n and the (1,1) element of G.

Appendix IV

This appendix gives the proof of Proposition 3 presented in Section 5.4, which gives the conditional posterior of β.

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

$$\overline{V}_\beta = \left(\sum_{t=1}^{T} \left(\Sigma_t^{-1} \otimes x_t x_t' \right) + G^{-1} \otimes \underline{V}_0^{-1} \right)^{-1} \tag{29}$$

and that $\overline{\mu}_\beta$ is given by (20). However, expression (29) requires the inversion of a $rk \times rk$ matrix. To obtain (19) first note that (3) and (24) 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 (29) as:

$$\begin{aligned}
\overline{V}_\beta &= \left((B \otimes I_k) \left(\sum_{t=1}^{T} \left(\begin{pmatrix} K_t & 0 \\ 0 & I_{r_2} \end{pmatrix} \otimes x_t x_t' \right) + \begin{pmatrix} \widetilde{I}_{r_1}^{-1} & 0 \\ 0 & I_{r_2} \end{pmatrix} \otimes \underline{V}_0^{-1} \right) (B' \otimes I_k) \right)^{-1} \\
\overline{V}_\beta &= (\widetilde{A} \otimes I_k) \left(\begin{pmatrix} \sum_{t=1}^{T} K_t \otimes x_t x_t' + \widetilde{I}_{r_1}^{-1} \otimes \underline{V}_0^{-1} & 0 \\ 0 & I_{r_2} \otimes \left(\sum_{t=1}^{T} x_t x_t' + \underline{V}_0^{-1} \right) \end{pmatrix} \right)^{-1} (\widetilde{A}' \otimes I_k) \\
\overline{V}_\beta &= (\widetilde{A}_1 \otimes I_k) \left(\sum_{t=1}^{T} (K_t \otimes x_t x_t') + \widetilde{I}_{r_1}^{-1} \otimes \underline{V}_0^{-1} \right)^{-1} (\widetilde{A}_1' \otimes I_k) + A_2 A_2' \otimes \left(\sum_{t=1}^{T} x_t x_t' + \underline{V}_0^{-1} \right)^{-1}
\end{aligned} \tag{30}$$

which is equal to (19), and where we have used that $\widetilde{A} \otimes I_k = (\widetilde{A}_1, A_2) \otimes I_k = (\widetilde{A}_1 \otimes I_k, A_2 \otimes I_k)$. To derive (21) simply note that $\left(\left(\overline{V}_\beta \right)^{1/2} \right)' \left(\overline{V}_\beta \right)^{1/2} = D_1 D_1' + D_2 D_2' = \overline{V}_\beta$, as we wanted to show. ■