GRIPS Discussion Paper 24-2

Approximate Factor Models with a Common Multiplicative Factor for Stochastic Volatility

By

Roberto Leon-Gonzalez
Blessings Majoni

April 2024

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

Approximate Factor Models with a Common Multiplicative Factor for Stochastic Volatility*

Roberto Leon-Gonzalez, Blessings Majoni

National Graduate Institute for Policy Studies

Japan

April, 2024

*We thank seminar participants at the 2023 Örebro Financial Econometrics Workshop, 2023 China Forum on Bayesian Econometrics, Strathclyde University Economics Seminars and Department of Quantitative Methods in Economics and Business (ULPGC) for helpful comments and suggestions. We also thank Niko P. Hauzenberger, Gary Koop, Yasuhiro Omori, Takashi Takenouchi, Takashi Tsuchiya, Francisco Vazquez-Polo, Mike West and Ping Wu for helpful comments and suggestions. We gratefully acknowledge financial support from JSPS (category C, 19K01588) and from GRIPS Policy Research Center (grant G241RP208). Roberto Leon-Gonzalez is a Senior Fellow of the RCEA. All errors are of course our own.

Abstract

Common factor stochastic volatility (CSV) models capture the commonality that is often observed in volatility patterns. However, they assume that all the time variation in volatility is driven by a single multiplicative factor. This paper has two contributions. Firstly we develop a novel CSV model in which the volatility follows an inverse gamma process (CSV-IG), which implies fat Student's t tails for the observed data. We obtain an analytic expression for the likelihood of this CSV model, which facilitates the numerical calculation of the marginal and predictive likelihood for model comparison. We also show that it is possible to simulate exactly from the posterior distribution of the volatilities using mixtures of gammas. Secondly, we generalize this CSV-IG model by parsimoniously substituting conditionally homoscedastic shocks with heteroscedastic factors which interact multiplicatively with the common factor in an approximate factor model (CSV-IG-AF). In empirical applications we compare these models to other multivariate stochastic volatility models, including different types of CSV models and exact factor stochastic volatility (FSV) models. The models are estimated using daily exchange rate returns of 8 currencies. A second application estimates the models using 20 macroeconomic variables for each of four countries: US, UK, Japan and Brazil. The comparison method is based on the predictive likelihood. In the application to exchange rate data we find strong evidence of CSV and that the best model is the IG-CSV-AF. In the Macro application we find that 1) the CSV-IG model performs better than all other CSV models, 2) the CSV-IG-AF is the best model for the US, 3) the CSV-IG is the best model for Brazil and 4) exact factor SV models are the best for UK and JP.

1 Introduction

Since the seminal work of Sims (1980), Vector Autoregressions (VAR) models have been a workhorse for informing macroeconomic policy making. They are used, for example, to estimate the impact of fiscal and monetary policies in the economy. The work of Engle (1982) demonstrated that it is very important to explicitly model the time varying variance of macroeconomic or financial variables, proposing Autoregressive Conditional Heteroscedasticity (ARCH) models, and later the literature proposed Stochastic Volatility (SV) models as an improvement (e.g. Shephard (1994), Kim et al. (1998)).

In recent years VAR models with stochastic volatility are used extensively in economics (e.g. Clark and Mertens (2023)). A simple approach is the Common Stochastic Volatility (CSV) model (e.g. Pajor (2006), Yu and Meyer (2006)) which assumes that the var-cov matrix of the error term e_t can be written as $var(e_t) = \sigma_t \Sigma$, where Σ is a constant unrestricted positive definite symmetric matrix, and σ_t is a univariate Log-Normal Autoregressive (LNAR) process. As argued by Carriero et al. (2016), the simplicity of the CSV model is an advantage because it permits the estimation of large VAR models, which provide a better understanding of the relationships among variables, and often better forecasting. Furthermore, Carriero et al. (2016) found that the empirical performance of the CSV model was not far from that of more flexible SV models using US macroeconomic data.

Subsequent literature improved the CSV model by adding serial correlation and fat tails in the errors (Chan (2020), Hartwig (2022)) and some studies found that variants of the CSV model outperform more flexible SV models (e.g. Poon (2018), Hou et al. (2023), Götz and Hauzenberger (2021)). For example, Hou et al. (2023) compared variants of the CSV model to among others exact SV factor models (henceforth FSV, e.g. Chib et al. (2006), Kastner (2019)) and Cholesky SV models (henceforth BVAR-SV, e.g. Cogley and Sargent (2005)) with and without fat tails using Australian macroeconomic data and found that the best model according to joint predictive likelihoods was a CSV with Student's t errors (CSV-t) in a large VAR of 20 variables, and the same model extended with Moving Average (MA) serial correlation (CSV-MA-t) for a small VAR of 3 variables.

Another advantage of the CSV structure is that the impact of σ_t can be interpreted as the effect of uncertainty on the economy (e.g. Mumtaz (2016), Mumtaz (2018)). The literature has also emphasized the importance of allowing for fat tails in the distribution of the errors even when using general forms of SV (e.g. Cross and Poon (2016), Chiu et al. (2017)). Furthermore, a recent literature has stressed the importance of using order-invariant approaches to multivariate SV (e.g. Chan et al. (2018), Chan et al. (2023), Wu and Koop (2022), Arias et al. (2023)), especially in large VARs. For example Chan et al. (2018) proposed an order-invariant approximate SV factor model which relaxes the assumption of independence among idiosyncratic factors that is made in FSV models. As noted by Chamberlain and Rothschild (1983) even a small departure from this assumption will imply that exact factor models will require a large number of factors to adequately capture the correlation structure.

This paper has three main contributions. Firstly we propose a novel CSV model, denoted as CSV-IG, in which the distribution of SV is a gamma autoregressive process (León-González (2019), Sundararajan and Barreto-Souza (2023), Leon-Gonzalez and Majoni (2023)), which implies a Student's t distribution for the observed dependent variables. In contrast to previous CSV models, we are able to derive an analytic expression of the integrated likelihood, and to sample exactly from the posterior distribution of the volatilities, obtaining a simpler numerical method for the calculation of the marginal likelihood, and permitting Maximum Likelihood estimation (MLE).

Secondly, we generalize the CSV structure by increasing parsimoniously the number of heteroscedastic factors while keeping the assumption of a multiplicative factor that impacts all volatilities simultaneously. We therefore allow for some of the components of the normalized vector $\tilde{e}_t = (1/\sqrt{\sigma_t})e_t$ to be heteroscedastic, while keeping the unconditional variance of e_t unrestricted. We use an approximate factor structure for $\tilde{e}_t$ and therefore denote this model as CSV-IG-AF.

Thirdly, we carry out an extensive empirical exercise to evaluate the new models and a large number of competing models, including variants of the CSV model, in addition to more general SV structures such as FSV and BVAR-SV models. In one application we use data on 8 daily exchange rates from the main trading partners of Zimbabwe and in another one we use quarterly data on 20 macroeconomic variables from each of 4 countries: US, UK, Japan and Brazil.

Section 2 describes the CSV-IG model, providing the analytic expression of the likelihood, the posterior distribution of the volatilities and the calculation of the marginal likelihood. Section 3 describes the CSV-IG-AF model, Section 4 presents the empirical exercise and Section 5 concludes.

2 Inverse Gamma CSV Model

Section 2.1 describes the model, the likelihood and the joint posterior density of the volatilities. Section 2.2 provides a method to calculate the marginal likelihood.

2.1 Model, Likelihood and Posterior Density of Volatilities

The model can be described as follows:

$$Y_t = \Pi x_t + e_t, \qquad e_t|\sigma_t \sim N\left(0, \sigma_t \Sigma\right) \tag{2.1}$$

where Y_t is a $r \times 1$ vector of observed dependent variables, Π is a $r \times k_x$ matrix of coefficients, x_t is a $k_x \times 1$ vector of observed regressors, and e_t is a $r \times 1$ vector of errors which is independent of x_t and i.i.d. Define the time varying stochastic process k_t as $k_t = (\sigma_t)^{-1}$, and assume that $k_t = z_t' z_t$, where z_t is an $n \times 1$ vector. The vector z_t has the following Gaussian AR(1) representation:

$$z_t = z_{t-1}\rho + \epsilon_t \qquad vec(\epsilon_t) \sim N(0, \theta^2 I_n) \tag{2.2}$$

The scalar parameter ρ controls the persistence of the volatility and n determines the degrees of freedom of the marginal distribution of σ_t, which is inverse gamma. This representation of z_t implies that the conditional distribution of $k_t|k_{t-1}$ is a non central chi squared. The non central chi-squared distribution is well defined for non integer values of n, therefore we will treat the unknown parameter n as continuous. Given the properties of a gamma, the conditional mean of the inverse time varying volatility k_t is a weighted average of the unconditional mean of k_t and its previous value k_{t-1}:

$$E(k_t|k_{t-1}) = \rho^2 k_{t-1} + (1-\rho^2)E(k_t)$$

To ensure stationarity we assume that $|\rho| < 1$. We also assume that the initial distribution of k_1 is the same as the stationary distribution of k_t:

$$k_1 \backsim Gamma\left(\frac{n}{2}, \frac{2\theta^2}{1-\rho^2}\right) \tag{2.3}$$

As a normalization we fix θ^2 such that the unconditional mean of σ_t is equal to one, which implies $\theta^2 = \frac{1-\rho^2}{n-2}$. For this purpose we impose the restriction that $n > 2$, so that e_t has a finite stationary variance. With this normalization there are only two volatility parameters to estimate: ρ^2 and n.

Thus, this model has the same framework as in the CSV literature that follows the seminal paper of Carriero et al. (2016) in that only σ_t varies with time. In this model however, σ_t is inverse gamma (IG) whereas the CSV literature has σ_t following a log normal distribution. The inverse gamma specification implies a Student's t distribution for Y_t thus enabling us to model heavy tailed distributions. Furthermore, the IG specification allows us to integrate out the volatilities, obtaining an analytic expression of the likelihood and exact sampling from the joint posterior distribution of the volatilities, as the following two propositions that are proved in the Appendix show.

Proposition 2.1. *Define $\varepsilon_t^2 = e_t'\Sigma^{-1}e_t$, with $e_t = Y_t - \Pi x_t$, then the likelihood function of the IG-CSV model described in equations (2.1)-(2.3) is as follows:*

$$L(Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}2^{\frac{r}{2}}\frac{\Gamma\left(\frac{n+r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left|\varepsilon_1^2 + V_1^{-1}\right|^{-\frac{n+r}{2}}V_1^{-\frac{n}{2}}$$

$$L(Y_2|Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\frac{(\varepsilon_2^2+1)^{-\frac{n+r}{2}}}{(1-\delta_2)^{-\frac{n+r}{2}}}\hat{C}_2$$

$$L(Y_3|Y_2,Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_3}\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\frac{\Gamma\left(\frac{n+r+2h_2}{2}\right)}{(\varepsilon_3^2+1)^{\frac{n+r}{2}}}(2S_3)^{\frac{n+r+2h_2}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\hat{C}_3$$

and for any $t \geq 3$:

$$L(Y_t|Y_{1:t-1}) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_t}\sum_{h_{t-1}=0}^{\infty}\widetilde{C}_{t-1,h_{t-1}}\frac{\Gamma\left(\frac{n+r+2h_{t-1}}{2}\right)}{(\varepsilon_t^2+1)^{\frac{n+r}{2}}}(2S_t)^{\frac{n+r+2h_{t-1}}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\left(\frac{n+r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\hat{C}_t$$

where:

$$
\begin{aligned}
&V_1 = (1-\rho^2)^{-1} \\
&\widetilde{V}_2^{-1} = V_1^{-1} + \varepsilon_1^2 \\
&\delta_2 = \rho^2(\widetilde{V}_2^{-1} + \rho^2)^{-1} \\
&Z_2 = (\varepsilon_t^2 + 1)^{-1}\delta_2 \\
&\widetilde{C}_{2,h_2} = \frac{[(n+r)/2]_{h_2}}{[n/2]_{h_2}} \left(\frac{1}{2}\rho^2(\widetilde{V}_2^{-1} + \rho^2)^{-1}\right)^{h_2} \frac{1}{h_2!} \\
&\widetilde{C}_{3,h_3} = \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} \Gamma\left(\frac{n+r+2h_2}{2}\right) \frac{[(n+r)/2 + h_2]_{h_3}}{[n/2]_{h_3}} \left(\frac{1}{2}\rho^2 S_3\right)^{h_3} \frac{1}{h_3!} (2S_3)^{\frac{n+r+2h_2}{2}} \\
&c_3 = {}_2F_1\left(\frac{n+r}{2}, \frac{n+r}{2}; \frac{n}{2}; \delta_3\right) \Gamma\left(\frac{n+r}{2}\right) (1-\rho^2 S_3)^{-\frac{n+r}{2}} (2S_3)^{\frac{n+r}{2}} \\
&\hat{C}_t = {}_2F_1\left(\frac{n+r+2h_{t-1}}{2}, \frac{n+r}{2}; \frac{n}{2}; Z_t\right) \ \textit{for } t \geq 2 \textit{ and where } h_1 = 0
\end{aligned}
$$

for $T \geq t \geq 3$

$$
\begin{aligned}
&S_t = (\varepsilon_{t-1}^2 + 1 + \rho^2)^{-1} \\
&\widetilde{V}_t^{-1} = \varepsilon_{t-1}^2 + 1 \\
&Z_t = (\varepsilon_t^2 + 1)^{-1} S_t \rho^2 \\
&\delta_t = \left((1-\rho^2 S_t)^{-1} S_t \rho^2 (\widetilde{V}_{t-1}^{-1} + \rho^2)^{-1}\right)
\end{aligned}
$$

and for $T+1 \geq t \geq 4$*:*

$$
c_t = \sum_{h_{t-1}=0}^{\infty} \widetilde{C}_{t-1,h_{t-1}} (1-\rho^2 S_t)^{-\frac{n+r+2h_{t-1}}{2}} \Gamma\left(\frac{n+r+2h_{t-1}}{2}\right) (2S_t)^{\frac{n+r+2h_{t-1}}{2}}
$$

$$
\widetilde{C}_{t-1,h_{t-1}} = \sum_{h_{t-2}=0}^{\infty} \widetilde{C}_{t-2,h_{t-2}} \Gamma\left(\frac{n+r+2h_{t-2}}{2}\right) \frac{[(n+r)/2 + h_{t-2}]_{h_{t-1}}}{[n/2]_{h_{t-1}}} \left(\frac{1}{2}\rho^2 S_{t-1}\right)^{h_{t-1}} \frac{(2S_{t-1})^{\frac{n+r+2h_{t-2}}{2}}}{h_{t-1}!}
$$

and $S_{T+1} = (1+\varepsilon_T^2)^{-1}$

$[x]_h$ denotes the rising factorial and ${}_2F_1$ a Gauss hypergeometric function (e.g. Muirhead (2005, p. 20)). These hypergeometric functions can be transformed to accelerate their convergence in a number of ways. Abramowitz et al. (1988, p. 559) defines several transformations such as the Euler transformation where:

$$ {}_2F_1(a,b;c;z) = (1-z)^{c-a-b}{}_2F_1(c-a,c-b;c;z) $$

or a linear combination approach:

$$ \begin{aligned} {}_2F_1(a,b;c;z) = & \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}{}_2F_1(a,b;a+b-c+1;1-z) \\ & + (1-z)^{c-a-b}\frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}{}_2F_1(c-a,c-b;c-a-b+1;1-z) \\ & \text{for } (|arg(1-z)| < \pi) \end{aligned} $$

Applying the Euler transformation to $\hat{C}_t$ gives:

$$ \hat{C}_t = (1-Z_t)^{-\frac{n+2r+2h_{t-1}}{2}}{}_2F_1\bigg(-\frac{r+2h_{t-1}}{2}, -\frac{r}{2}; \frac{n}{2}; Z_t\bigg) \text{ for t} \geq 2 \text{ and where } h_1 = 0 $$

However, in our coding we used the Euler acceleration only for $\hat{C}_2$ and c_3. Instead, we accelerated the calculations by implementing parallel computing in the code. This is possible because many of the coefficients in the series are the same for every t, therefore they only need to be computed once, which can be done in parallel. We also calculate all the $\hat{C}_t$ in parallel.

The following proposition shows that the posterior of $k_t|k_{1:(t-1)}$ is a mixture of gammas and therefore it is possible to simulate exactly from the volatilities.

Proposition 2.2. *The joint posterior distribution $\pi(k_{1:T}|Y_{1:T})$ can be obtained from the following conditional densities each of which is a mixture of gammas:*

$$ \pi(k_t|k_{(t+1):T}, Y_{1:T}) \propto |k_t|^{\frac{n+r-2}{2}} \exp\bigg(-\frac{1}{2}S_{t+1}^{-1}k_t\bigg)\sum_{h=0}^{\infty}\big(C_{t,h}|k_t|^h\big),\ t = 1, ..., T $$

where

$$ \begin{aligned} C_{1,h} &= \frac{1}{h!}\frac{1}{[n/2]_h}\bigg(\frac{1}{4}\rho^2 k_2\bigg)^h \\ S_2 &= (\varepsilon_1^2 + 1)^{-1} \\ S_{T+1} &= (\varepsilon_T^2 + 1)^{-1} \end{aligned} $$

for $3 \leq t \leq T$

$$ S_t = (\varepsilon_{t-1}^2 + 1 + \rho^2)^{-1} $$

and for $2 \leq t < T$*:*

$$C_{t,h} = \sum_{h_t=0}^{h} \widetilde{C}_{t,h-h_t} \frac{1}{[n/2]_{h_t}} \left(\frac{1}{4}\rho^2\right)^{h_t} \frac{k_{t+1}^{h_t}}{h_t!}$$

while for $t = T$*,* $C_{t,h} = \widetilde{C}_{t,h}$*, and where* $\widetilde{C}_{t,h}$ *has been defined in Proposition 2.1.*

We tested the code that implements the sampling from the posterior densities of $k_{1:T}$ described in Proposition 2.2 using the test proposed by Geweke (2004) with the U.S. Macro data described in Section 4. The result gives evidence that our procedure is sampling from the true posterior densities (Section 6.3 in the Appendix).

A Gibbs sampling type of posterior simulator for this model can be implemented with the following 4 steps: 1) generate Π conditional on $(k_{1:T}, \Sigma)$ from a Multivariate Normal, 2) generate Σ conditional on $(k_{1:T}, \Pi)$ from an Inverted Wishart, 3) generate (ρ^2, n) using a Metropolis-step that targets the likelihood given in Proposition 2.1, 4) generate the inverse volatilities $k_{1:T}$ as mixtures of gammas according to Proposition 2.2.

2.2 Calculation of the Marginal Likelihood

The numerical calculation of the marginal likelihood for our model is simpler because we have an exact analytic expression for the integrated likelihood, requiring only two estimations of the model and providing reliable calculations even for the large VARs models considered in Section 4. The marginal likelihood for the lognormal volatility CSV models has been calculated in the literature by adapting the method of Chib and Jeliazkov (2001), which requires more runs of the MCMC algorithm (e.g. Chan (2020)), or by combining conditional Monte Carlo with modified importance sampling (Chan (2023)), which requires finding an importance density for the vector of volatilities.

We use an importance sampling approach in which we compare our CSV-IG model (denoted as M_2) with a fictitious model M_1 for which the marginal likelihood is available in analytic form. M_1 is the same as M_2 except that $var(e_t) = \hat{\sigma}_t \Sigma$, where $\hat{\sigma}_t$ is fixed and equal to the posterior mean of σ_t under model M_2. Therefore M_2 has two parameters more than M_1: the volatility parameters ρ^2 and n. To compare these two models by importance sampling we need to expand the posterior of M_1 with a distribution $\hat{f}(n, \rho^2)$, which is an approximation of the conditional posterior $\pi(n, \rho^2 | Y, M_2)$. Because the priors of n and ρ^2 are lognormal and beta, respectively, we specify $\hat{f}(n, \rho^2)$ as the same family of distributions, with parameters adjusted to approximate the posterior means and variances.

Defining $\tilde{\Psi} = (\Psi, n, \rho^2)$, where $\Psi = (\Pi, \Sigma)$, we can approximate the marginal likelihood by first approximating the following Bayes factor:

$$\frac{\pi(Y|M_1)}{\pi(Y|M_2)} = \int \frac{\pi(Y|\Psi, M_1)\pi(\Psi|M_1)\hat{f}(n, \rho^2)}{\pi(Y|\tilde{\Psi}, M_2)\pi(\Psi|M_2)\pi(n, \rho^2|M_2)}\pi(\tilde{\Psi}|Y, M_2)d\tilde{\Psi}$$

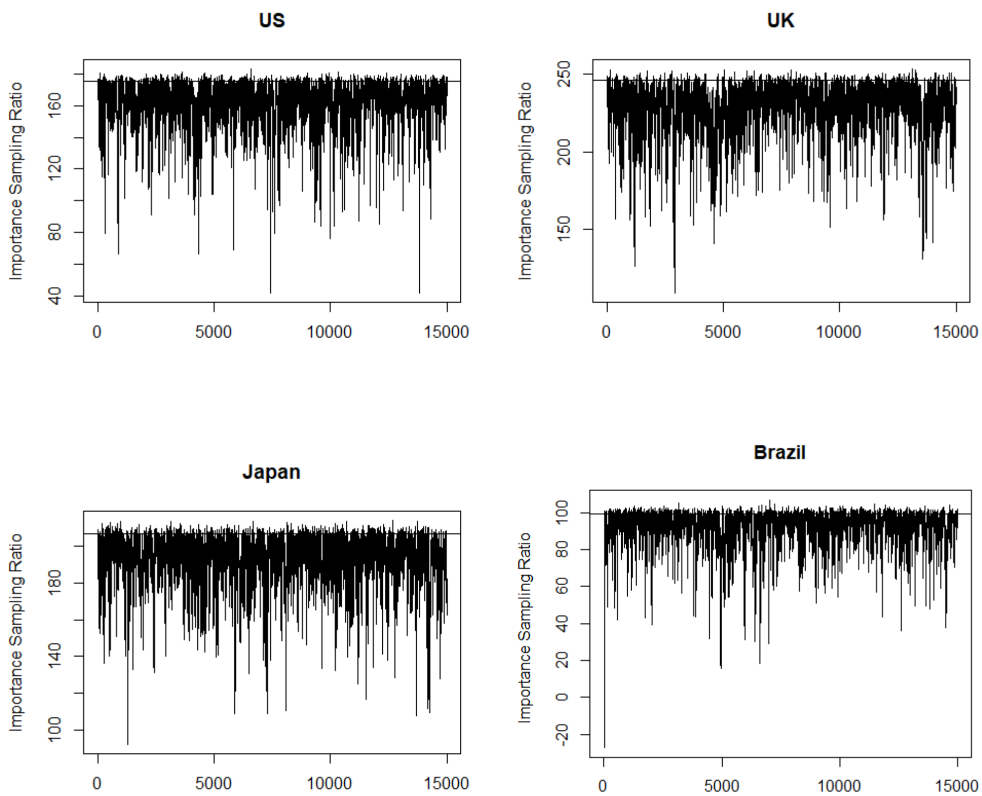
where $\pi(Y|\tilde{\Psi}, M_2)$ is the likelihood after integrating the volatilities, as given by Proposition 2.1. This Bayes factor can be calculated by importance sampling, where the weight for each value of $\tilde{\Psi}$ is thus defined as

$$W(\tilde{\Psi}_i) = \frac{\pi(Y|\Psi_i, M_1)\pi(\Psi_i|M_1)\hat{f}(n_i, \rho_i^2)}{\pi(Y|\tilde{\Psi}_i, M_2)\pi(\Psi_i|M_2)\pi(n_i, \rho_i^2|M_2)}$$

where each $\tilde{\Psi}_i = (\Psi_i, n_i, \rho_i^2)$ is obtained with the MCMC sampler for M_2. Thus, the Bayes Factor can be approximated with $\frac{1}{N}\sum_{i=1}^{N} W(\tilde{\Psi}_i)$, where N is the number of random draws from the posterior. The posterior in model M_2 is more spread than the posterior in M_1, as desired for importance sampling.

Figure 1 shows the importance sampling ratios obtained from 15000 iterations with a burn in of 1000 of the sampler using our approach for the macroeconomic data. The horizontal line indicates the estimated value of the log Bayes factor. Approximately 5% of the log weights go beyond the horizontal line indicating good performance.

Figure 1: Importance Sampling Ratios



Importance Sampling Ratios obtained from 15000 iterations.

3 Approximate Factor Model with a Common Multiplicative Factor

The CSV assumption that $var(e_t) = \sigma_t \Sigma$ is equivalent to assuming that there is a multiplicative heteroscedastic factor f_t that interacts with homoscedastic errors $\tilde{e}_t$, such that $e_t = f_t \tilde{e}_t$, with $var(\tilde{e}_t) = \Sigma$, $var(f_t) = \sigma_t$ and f_t being independent of $\tilde{e}_t$. Because this implies that all r linear combinations of $\tilde{e}_t$ are homoscedastic, we generalize the model by assuming that there are only $r - r_1$ homoscedastic linear combinations of $\tilde{e}_t$ with the remaining r_1 combinations being heteroscedastic, while keeping the assumption that the unconditional var-cov matrix of $\tilde{e}_t$ is an unrestricted positive definite symmetric matrix Σ.

We therefore assume $var(e_t) = \sigma_t \Sigma_t$ and for Σ_t we specify an approximate factor model structure as proposed by Chan et al. (2018). We assume that $\Sigma = E(\Sigma_t)$ exists and is finite, and that the vector $\tilde{e}_t = e_t/\sqrt{\sigma_t}$ 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 $r = r_1 + r_2$:

$$\tilde{e}_t = A_1 u_{1t} + A_2 u_{2t} = \begin{pmatrix} A_1 & A_2 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = A u_t. \tag{3.1}$$

where A is a $r \times r$ matrix and $u_t = (u_{1t}', u_{2t}')'$. As a normalization we fix $E(\Upsilon_t^{-1}) = I_{r_1}$ such that $\Sigma = E(\Sigma_t) = A_1 A_1' + A_2 A_2'$. To identify A_1, A_2 we use the eigenvalue decomposition of Σ. In particular, A_1, A_2 are restricted such that $A_1' A_2 = 0$, $A_1' A_1 = S_1$, $A_2' A_2 = S_2$, where S_1 and S_2 are diagonal matrices containing the eigenvalues of Σ in decreasing order. The eigenvalues in S_1 are larger than those in S_2. This normalization implies a one-to-one mapping between Σ and (A_1, A_2).

Therefore $var(\tilde{e}_t | x_t) = \Sigma_t$ can be written as:

$$\Sigma_t = A_1 \Upsilon_t^{-1} A_1' + A_2 A_2', \tag{3.2}$$

where Υ_t is a Wishart Autoregressive process of order 1 (WAR(1), Gourieroux et al. (2009)), normalized such that $E(\Upsilon_t^{-1}) = I_{r_1}$.

Identifying A_1, A_2 using the eigenvalue decomposition of Σ is natural if it is assumed that $e_t/\sqrt{\sigma_t}$ has an approximate factor structure (Chamberlain and Rothschild (1983)) with only r_1 heteroscedastic factors and r is large relative to r_1, because a factor structure implies that the first r_1 eigenvalues of Σ 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'). Hence we can interpret u_{1t} as the heteroscedastic factors and A_1 as the factor loadings. Because only the products $A_1 A_1'$ and $A_2 A_2'$ are identified (Chamberlain and Rothschild (1983)) we solved the indeterminacy by restricting A_1, A_2 such that $A_1' A_1$ and $A_2' A_2$ are diagonal matrices.

We specify priors directly on Σ and Π, which allows us to specify the same priors as in the CSV or homoscedastic VAR models, facilitating model comparison.

The WAR(1) can be described by first defining $K_t = Z_t' Z_t$, where Z_t is a $\tilde{n} \times r_1$ matrix distributed as a Gaussian AR(1) process:

$$Z_t = Z_{t-1} \tilde{\rho} + \varepsilon_t, \qquad vec(\varepsilon_t) \sim N(0, I_{r_1} \otimes I_{\tilde{n}}), \tag{3.3}$$

where $\tilde{\rho}$ is diagonal $r_1 \times r_1$ (with diagonal elements smaller than one in absolute value), $\otimes$

denotes the Kronecker product and we assume that $vec(Z_1)$ is drawn from the stationary distribution $N(0,(I_{r_1}-\tilde{\rho}^2)^{-1}\otimes I_{\tilde{n}})$. The parameter $\tilde{n}$ represents the degrees of freedom in the WAR(1) process and it will be estimated. This representation implies that $\tilde{n}$ is an integer, but as in the previous section, we will treat it as continuous because it is just the degrees of freedom parameter of a non-central Wishart density. Because $E(K_t^{-1})=(\tilde{n}-r_1-1)^{-1}(I-\tilde{\rho}^2)$, we normalize K_t^{-1} as $\Upsilon_t^{-1}=(\tilde{n}-r_1-1)(I-\tilde{\rho}^2)^{-1/2}K_t^{-1}(I-\tilde{\rho}^2)^{-1/2}$, so that $E(\Upsilon_t^{-1})=I_{r_1}$. We assume that $\tilde{n}>r_1+1$, such that $E(\Sigma_t)$ is finite

Regarding the posterior simulator, conditional on $k_{1:T}$ the parameters $\Pi,\Sigma,\tilde{n},\tilde{\rho}^2,K_{1:T}$ can be sampled as described in Chan et al. (2018). This algorithm generates $K_{1:T}$ using a conditional particle filter (Andrieu et al. (2010)), Π from its normal conditional posterior, and $(\Sigma,\tilde{n},\tilde{\rho}^2)$ using a Metropolis step. Thanks to the presence of homoscedastic errors, drawing Π only requires inverting matrices of order r_1r and k_x. Conditional on $K_{1:T}$ we can use the steps 3 and 4 outlined at the end of Section 2.1 to draw (ρ^2,n) and $k_{1:T}$, respectively. The Appendix in Section 6.4 shows the trace plots of this algorithm for the US data with $r=20, r_1=1$, showing very good convergence and mixing properties. A similar performance was obtained with the exchange rate data.

4 Empirical Application

To illustrate the efficiency and usefulness of our proposed model addition to the CSV literature, we provide two applications. The first application uses daily exchange rate returns in a VAR of 8 currencies. The second application uses 20 macroeconomic variables each for US, UK, Japan and Brazil.

In both applications we compare our proposed IG models to the 9 model specifications listed in Table 1. In addition to the standard Bayesian VAR with Gaussian errors (e.g. Sims (1980)) we consider the CSV model (e.g. Pajor (2006), Yu and Meyer (2006), Carriero et al. (2016)) in which $var(e_t)=\sigma_t\Sigma$, with σ_t following a stationary log-normal autoregressive (LNAR) process. The third model (CSV-t) adds Student's t innovations by writing $var(e_t)=\lambda_t\sigma_t\Sigma$, with λ_t following an iid univariate inverse gamma distribution (e.g. Chan (2020)). The fourth model (CSV-MA), which was proposed by Chan (2020), introduces serial correlation by assuming that e_t has the following Moving Average (MA) structure: $e_t=\varepsilon_t+\psi_1\varepsilon_{t-1}$, with ε_t being iid, $var(\varepsilon_t)=\sigma_t\Sigma$, and σ_t is a LNAR process. The fifth model is like the fourth but with Student's t innovations (CSV-MA-t), such that $var(\varepsilon_t)=\lambda_t\sigma_t\Sigma$ (Chan (2020)), with λ_t being iid univariate inverse gamma. The sixth model is an exact heteroscedastic factor model (FSV-r_1) such that $e_t = A_1 u_1 + \tilde{u}_2$, $u_1 : r_1 \times 1$, $\tilde{u}_2 : r \times 1$, $cov(u_1, \tilde{u}_2) = 0$, both u_1 and $\tilde{u}_2$ Gaussian having diagonal var-cov matrices with diagonal elements that change with time according to independent LNAR processes (e.g. Chib et al. (2006)). The seventh model (FSV-t-r_1) is like the sixth, but the elements of $\tilde{u}_2$ follow independent Student's t distributions instead of Gaussian (e.g. Chib et al. (2006)). The eighth model assumes that the diagonal elements of the Cholesky decomposition of $var(e_t)$ follow unit roots LNAR processes, while the rest of elements are constant (e.g. Cogley and Sargent (2005)). The ninth model is a heteroscedastic approximate factor model, such that $e_t = A_1 u_1 + A_2 u_2$, $u_1 : r_1 \times 1$, $u_2 : r_2 \times 1$, $cov(u_1, u_2) = 0$, with the var-cov matrix of u_1 changing with time according to a WAR(1) process and u_2 being homoscedastic (Chan et al. (2018)). In the case of factor models, we estimate the models with several values of r_1.

We evaluate the performance of the models by using joint predictive likelihoods, which measure the one step ahead out of sample forecasting accuracy. We use joint predictive likelihoods because our purpose is not to predict individual variables, but to find a model that captures well the dynamics in the volatility matrix, and therefore helps to understand the relationships among variables. Because of the connection to the marginal likelihood, the one-step ahead joint predictive likelihood seems to be the best criterion for this purpose.

For a given period T_0 the predictive likelihood $\pi(Y_{(T_0+1):T}|Y_{1:T_0}, M)$ measures how well the model predicts the data $Y_{(T_0+1):T}$ given previous data, and it can be obtained from the marginal likelihood as follows (e.g. Geweke and Amisano (2010)):

$$\pi(Y_{(T_0+1):T}|Y_{1:T_0}, M) = \frac{\pi(Y_{1:T}|M)}{\pi(Y_{1:T_0}|M)}$$

where $\pi(Y_{1:T_0}|M)$ denotes the marginal likelihood for model M given data $Y_{1:T_0}$.

Taking logs the log predictive likelihood becomes the difference of log marginal likelihoods. Therefore, one way to obtain the log predictive likelihood is to obtain the log marginal likelihoods given data to the time periods above and obtain the difference. Another approach, which is more time consuming, is to estimate the model repeatedly for each sample size, and use the following relationship:

$$log(\pi(Y_{(T_0+1):T}|Y_{1:T_0}, M)) = \sum_{h=1}^{T-T_0} log(\pi(Y_{(T_0+h):T}|Y_{1:T_0+h-1}, M))$$

where each component of the sum is calculated using the posterior simulator with data up to $T_0 + h$.

To compare models we report the Average Log Predictive Likelihood (ALPL), which can be obtained by averaging over the number of periods, that is:

$$ALPL = \frac{log(\pi(Y_{(T_0+1):T}|Y_{1:T_0}, M))}{T - T_0}$$

A larger ALPL implies a better empirical fit. Whenever the prior is based on the data, for example in the Minnesota prior, we use only data up to T_0 to train the prior in all cases.

For comparison purposes, the priors for those parameters that are common to all models are the same. The prior for Π, Σ is a normal-inverse-Wishart prior with shrinkage parameters $k_1 = 0.04$, $k_2 = 100$, as defined in Chan (2020), with $r + 3$ degrees of freedom, and such that the prior mean of Σ is equal to a diagonal matrix whose elements are estimated with OLS residuals using data up to T_0.

The algorithm had good mixing and convergence properties. The Appendix in Section 6.4 shows trace plots for the US data, which are similar to those obtained with other datasets.

Table 1: Models for Comparison

Model	**Error Structure**
BVAR	Homoscedastic Gaussian errors
CSV	CSV
CSV-t	CSV and t innovations
CSV-MA	CSV and MA(1) innovations
CSV-MA-t	CSV MA(1) and t innovations
FSV-r_1	Factor SV model with r_1 factors
FSV-t-r_1	Factor SV model with r_1 factors and t innovations
BVAR-SV	Only diagonal elements of Cholesky change with time
AF-r_1	Approximate Factor Models with r_1 factors
CSV-IG	CSV and inverse gamma (IG) SV
CSV-IG-AF-r_1	CSV, IG-SV and additional heteroscedastic multiplicative factors

4.1 Exchange Rate Data application

We use 1000 observations of daily exchange rate data for 8 currencies to the USD that constitute the top trading partners for Zimbabwe in terms of both exports and imports, whose ISO Codes are: GBP, EUR, CNY, HKD, INR, ZAR, SGD, ZWD. The data for the first 7 currencies were obtained from the Board of Governors of the Federal Reserve and covers the period beginning 29 April 2019 and ending 28 April 2023, while the ZWD series was obtained from the Reserve Bank of Zimbabwe for the same period.

Figures 2 and 3 show the last 400 observations in levels and log first differences. We estimate a VAR in which Y_t represents the vector of log first differences, while x_t contains an intercept and one lag of Y_t. Table 2 shows the ALPL for the last 200 observations for 6 models ($T - T_0 = 200$, $T = 998$), calculated as the difference of two marginal likelihoods. The best model is the CSV-t (ALPL=33.25), followed by the CSV-MA-t (ALPL=33.23) and the CSV-IG (ALPL=33.20). When we calculate the ALPL by estimating the model recursively for each sample size we are able to draw a comparison with a larger number of models, 16, which is shown in Table 3. In this case we use the last 400 observations of the dataset ($T = 400$), and calculate the ALPL with 50 observations ($T - T_0 = 50$). We find that all CSV models are much superior to all types of SV and FSV models, which gives strong support to the presence of the multiplicative heteroscedastic factor in CSV models. Although the CSV-t (ALPL=35.469) is better than the CSV-IG (ALPL=35.301), the CSV-IG-AF-2 becomes the winner among all models (ALPL=35.50), showing that adding additional heteroscedastic factors while keeping the CSV structure is an effective way of improving the empirical performance of CSV models. The numerical standard errors are sufficiently small that all differences in ALPL are statistically highly significant, and note that a difference of 0.03 in ALPL per observation is not a small number, since it implies that with a sample of only 100 observations the log Bayes factor would become 3, implying that the best model is 20 times more likely than the second best.

Figure 4 presents the contribution of each observation to the ALPL for the CSV-IG-AF-1 model and the FSV-1 model (which is the best among the FSV models), showing that the former absolutely dominates the latter for every observation. Table 4 sheds some light on why the FSV-1 model does not perform well and shows the sample correlation matrix of the idiosyncratic errors, which should be close to the identity matrix if the assumptions of exact factor models were correct. However, more than half of the correlations are larger than 0.1 in absolute values, with correlations that are as large as -0.59 or 0.46. As the table shows, in the FSV-2 model these correlations decrease only slightly, with still half of them being larger than 0.1 in absolute value. Therefore, one reason for the good performance of the CSV models is that they allow for a more flexible correlation structure. Although the SV model also allows for a flexible correlation structure, it performs badly, indicating that another important reason for the good performance of CSV models is the parsimonious representation of heteroscedasticity by means of a common multiplicative heterocedastic factor.

Figure 2: Exchange Rates in Levels

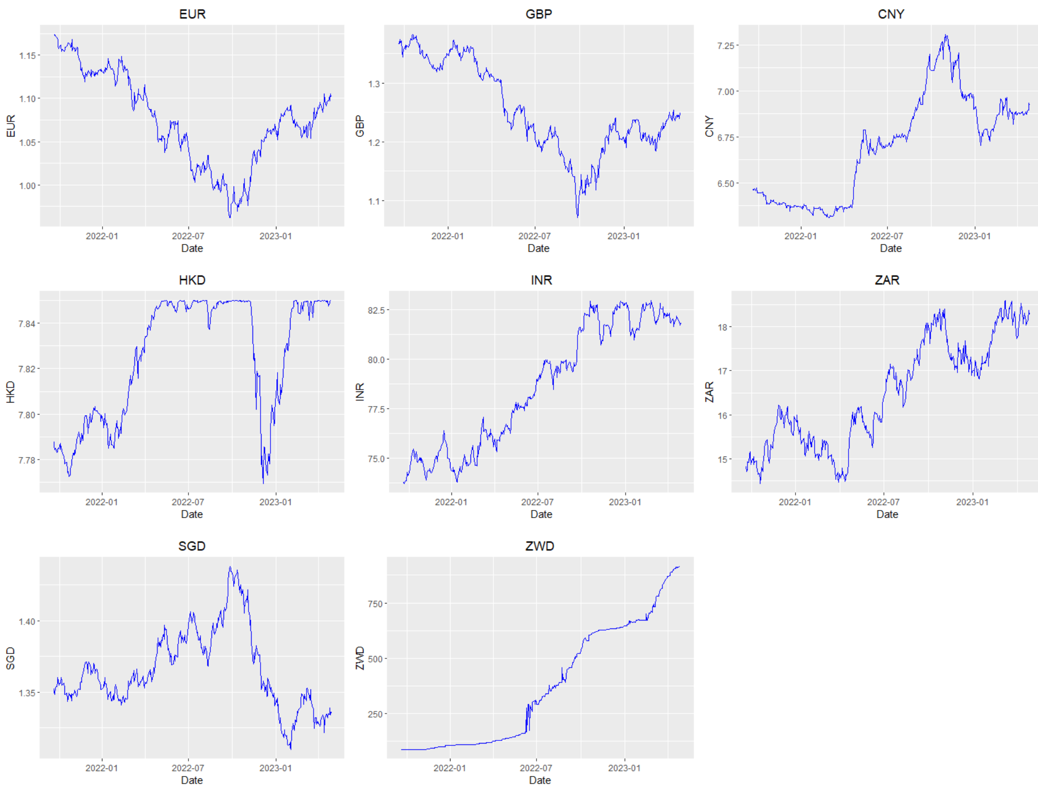


Figure 3: Exchange Rates in Log First Differences

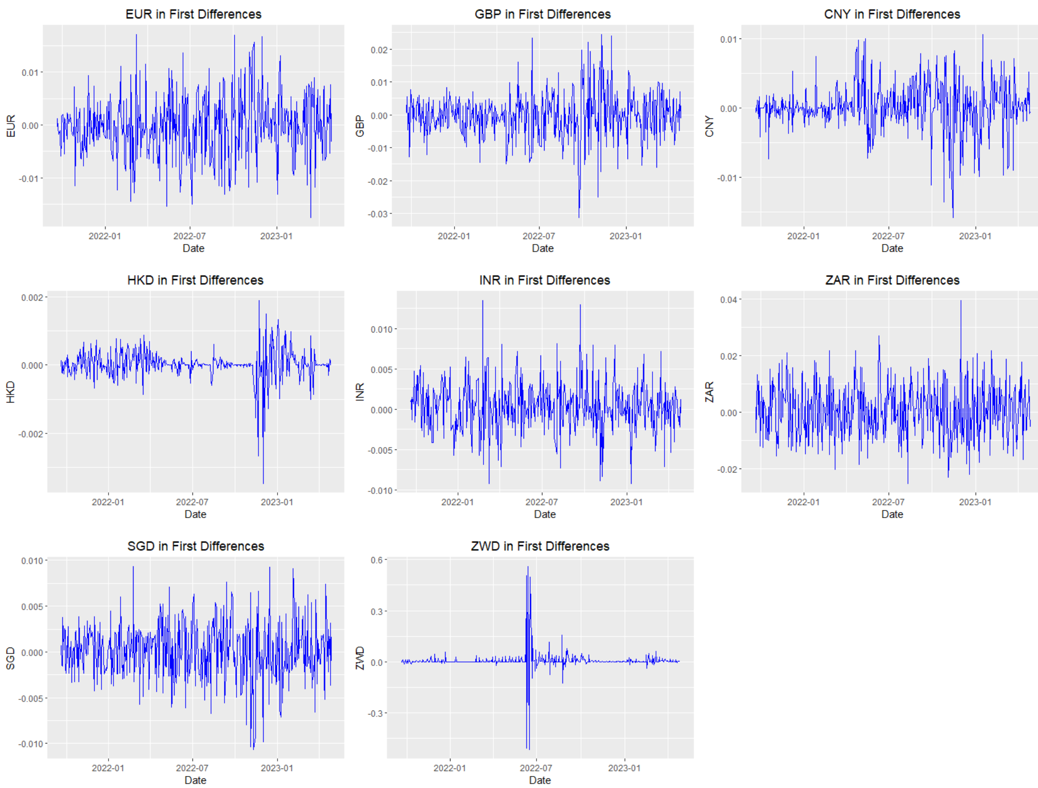


Table 2: Marginal Likelihood and ALPL for Exchange Rates VAR: $T - T_0 = 200$, $T = 998$

	ML	ALPL
BVAR	32989.11	31.81
BVAR-CSV	34408.31	33.12
BVAR-CSV-t	34546.71**	33.25**
BVAR-CSV-MA	34427.03	33.18
BVAR-CSV-MA-t	34474.12	33.23*
BVAR-CSV-IG	34490.0*	33.20

The best model is marked by ** and the second best by *.

Table 3: ALPL for Exchange Rates VAR: $T - T_0 = 50$, $T = 400$

Model	**ALPL**
BVAR	34.2083
BVAR−SV	27.9892
BVAR−CSV	35.4629
BVAR−CSV−t	35.4699*
BVAR−FSV−1	27.2324
BVAR−FSV−2	26.6181
BVAR−FSV−3	25.8803
BVAR−FSV−t−1	27.1685
BVAR−FSV−t−2	26.5469
BVAR−FSV−t−3	25.8002
BVAR−AF−1	35.2097
BVAR−AF−2	35.1769
BVAR−AF−3	35.0710
BVAR−CSV−IG	35.3010
BVAR−CSV−IG−AF1	35.4606
BVAR−CSV−IG−AF2	35.5001**

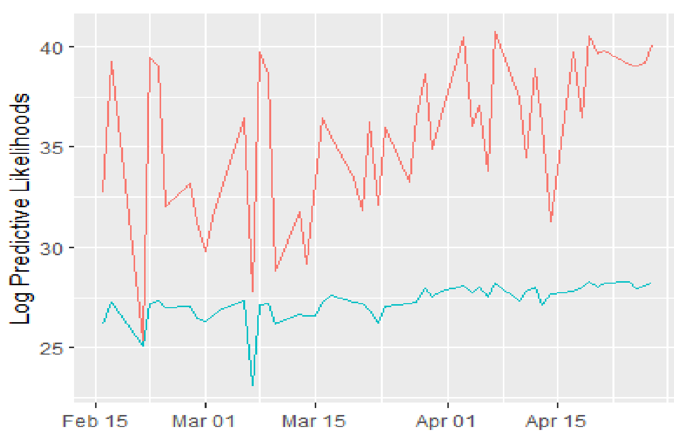


Figure 4: Contribution to the ALPL for CSV-IG-AF1 (above) versus FSV-1 (below) for the last 50 observations.

Table 4: Correlation matrix of idiosyncratic errors in SV exact factor models with one and two factors.

One factor model								Two factors model							
1.00	**0.32**	**-0.19**	-0.02	**-0.16**	**-0.21**	**-0.33**	-0.01	1.00	0.06	**-0.17**	-0.02	**-0.15**	**-0.20**	**-0.30**	-0.02
0.32	1.00	**-0.33**	-0.03	**-0.29**	**-0.39**	**-0.59**	**-0.11**	0.06	1.00	**-0.18**	-0.02	**-0.16**	**-0.22**	**-0.32**	-0.03
-0.19	**-0.33**	1.00	0.04	**0.18**	**0.28**	**0.44**	0.06	**-0.17**	**-0.18**	1.00	0.04	**0.17**	**0.27**	**0.43**	0.05
-0.02	-0.03	0.04	1.00	0.02	-0.01	0.03	0.00	-0.02	-0.02	0.04	1.00	0.02	-0.01	0.03	0.00
-0.16	**-0.29**	**0.18**	0.02	1.00	**0.25**	**0.32**	0.00	**-0.15**	**-0.16**	**0.17**	0.02	1.00	**0.24**	**0.31**	0.00
-0.21	**-0.39**	**0.28**	-0.01	**0.25**	1.00	**0.46**	0.01	**-0.20**	**-0.22**	**0.27**	-0.01	**0.24**	1.00	**0.45**	0.00
-0.33	**-0.59**	**0.44**	0.03	**0.32**	**0.46**	1.00	0.06	**-0.30**	**-0.32**	**0.43**	0.03	**0.31**	**0.45**	1.00	0.06
-0.01	**-0.11**	0.06	0.00	0.00	0.01	0.06	1.00	-0.02	-0.03	0.05	0.00	0.00	0.00	0.06	1.00

Correlations larger than 0.1 in absolute value are in bold.

4.2 Macroeconomic Application

For consistency with previous literature, the 20 macro variables that we choose for the US are the same as those used in e.g. Koop (2013), Carriero et al. (2016), Chan (2020) updated to 2022Q4. The variables include among others real output, personal consumption expenditures, investments, federal interest rates and the S&P500.

Similar variables were chosen for Japan (JP), UK and Brazil (BR). The comprehensive list with descriptions for the variables and the transformation employed is listed in Table 5[1].

Therefore we have 20 variables for each country ($r = 20$) and we include 4 lags and an intercept in each VAR. The last observation for each country is 2022Q4, and after constructing the lags the sample sizes become 248, 247, 247 and 103, for US, UK, JP and BR, respectively. For each country the ALPL is evaluated using the last 50 observations ($T - T_0 = 50$), such

[1]US macroeconomic data for the empirical application was obtained from the Federal Reserve Bank of Philadelphia, while the financial variables were sourced from the Federal Reserve Bank of St Louis.

Variables for Japan were obtained from the Federal Reserve Bank of St Louis with the exception of three variables that were obtained from CEIC data. That is, the foreign effective exchange rate and the monetary base, cited by CEIC as sourced from the Bank of Japan, while the industrial production index was sourced from the International Monetary Fund. All variables were chosen to closely match the 20 US variables, as such, the index of aggregate weekly hours for Japan represents hourly earnings for manufacturing whereas housing starts are obtained as data for work started on construction, dwellings or residential buildings as a total.

UK variables were obtained from the Federal Reserve Bank of St Louis. Long term government 10 year bond yields replace the 10 year treasury constant maturity rate. The Import price index for the UK is for all goods and services classified by origin. The variables for Brazil were also obtained from the Federal Reserve Bank of St Louis with the exception of 7 variables obtained from CEIC data sourced from various sources. The industrial production index, producer price index and the payroll index was cited as sourced from the Brazilian Institute of Geography and Statistics. The import price index was sourced from the Centre for Foreign Trade Studies Foundation. Government bond yields were sourced from the National Treasury Secretariat. The monetary base was sourced from the Central Bank of Brazil. Lastly, the Equity Market Index Sao Paulo Stock Exchange was calculated from the daily BOVESPA index.

Monthly data is converted to quarterly observations by obtaining their 3 monthly average values for the corresponding quarter.

that T_0 is 2010Q4.

According to Table 6, which shows the ALPL calculated as the difference of two marginal likelihoods, for every country the CSV-IG model is the best among all CSV models. The second best model is CSV-MA-t for the US, CSV for the UK, CSV-t or CSV-MA for Japan, and CSV-MA for Brazil[2].

In order to compare with a wider set that includes models for which the numerical calculation of the marginal likelihood is difficult, we calculate the ALPL by recursively estimating the model for different sample sizes, each time averaging the likelihood contribution over draws from the posterior (Geweke and Amisano (2010)). However, for this approach to work well, such average must not be dominated by a single or very few draws. Thus, in our calculations we require that at least 80 draws are above the calculated mean (which amounts to 5% of the iterations). When this requirement is not met, we consider the calculation as not reliable, and the result as missing. This happens for some observations for most of the models, because some observations during the COVID-19 crisis take extreme values, and the number of parameters to integrate out in the VAR is very large.

Table 7 shows the ALPL thus calculated for the US. Excluding the last 12 observations (2020Q1-2022Q4) from the evaluation period permits evaluating the ALPL for all models, with the best and second best models being the IG-CSV-AF1 and IG-CSV-AF2, respectively. Excluding only 9 observations from the evaluation period still allows comparison of all heteroscedastic models, and the ranking is the same. Excluding fewer observations implies that some models cannot be evaluated, but the winner models continue to be the same. Overall we can see that CSV models outperform exact factor models and the SV model.

Table 8 is for the UK and shows that whenever models can be compared, exact factor models are better, with the winner being the FSV-t-7. In this case the SV model is also better than CSV models.

Table 9 is for JP and shows results that are very similar to those of the UK. Exact factor models and the SV model perform better than CSV models, with the FSV-t-7 being the winner.

Table 10 is for BR and shows that the CSV-IG model can be evaluated with all observations, and is the winner in all cases. We can also see that all CSV models perform better than exact factor models and the SV model.

In summary, the best models are CSV-IG-AF1, FSV-t-7, FSV-t-7 and CSV-IG for US,

[2]Numerical standard errors for each log marginal likelihood calculation were at most 0.14, implying that the ALPL is accurate for more than two decimal points, so that all differences in ALPL values are significant.

UK, JP and BR, respectively. CSV models are better than exact factor models and the SV model in the US and BR, whereas the opposite happens in UK and JP.

Table 5: Variables Description

Variables Description	**Transformation**	**US**	**UK**	**JP**	**BR**
Real GNP/GDP	400 Δ log	$\circ$	$\circ$	$\circ$	$\circ$
Real Personal Consumption Expenditure	400 Δ log	$\circ$	$\circ$	$\circ$	$\circ$
Real Gross Private Domestic Investments:Nonresidential	400 Δ log	$\circ$			
Real Gross Private Domestic Investments:Residential	400 Δ log	$\circ$	$\circ$	$\circ$	
Real Net Exports of Goods and Services	None	$\circ$	$\circ$	$\circ$	$\circ$
Nominal Personal Income	400 Δ log	$\circ$			
Industrial Production Index	400 Δ log	$\circ$	$\circ$	$\circ$	$\circ$
Unemployment Rate	None	$\circ$	$\circ$	$\circ$	$\circ$
Nonfarm Payroll Employment	400 Δ log	$\circ$			$\circ$
Indexes of Aggregate Weekly Hours:Total	400 Δ log	$\circ$	$\circ$	$\circ$	
Housing Starts	400 Δ log	$\circ$	$\circ$	$\circ$	
Price Index for Personal Consumption Expenditures, Constructed	400 Δ log	$\circ$		$\circ$	$\circ$
Price Index for Imports of Goods and Services	400 Δ log	$\circ$	$\circ$		$\circ$
Effective Federal Funds Rate	None	$\circ$			$\circ$
1 Year Treasury Constant Maturity Rate	None	$\circ$	$\circ$		
10 Year Treasury Constant Maturity Rate	None	$\circ$	$\circ$		$\circ$
Moody's Seasoned Baa Corporate Bond Minus Federel Funds Rate	None	$\circ$			
ISM Manufacturing PMI Composite Index	None	$\circ$			
ISM Manufacturing New Orders Index	None	$\circ$			
S&P500	400 Δ log	$\circ$			
Producer Production Index	400 Δ log		$\circ$	$\circ$	$\circ$
Consumer Price Index	400 Δ log		$\circ$	$\circ$	$\circ$
Interest Rates ,Government Securities, Government Bonds	None		$\circ$	$\circ$	
Spot Exchange Rates	400 Δ log		$\circ$	$\circ$	$\circ$
M1	400 Δ log		$\circ$	$\circ$	$\circ$
M2	400 Δ log		$\circ$	$\circ$	$\circ$
Foreign Effective Exchange Rate	400 Δ log		$\circ$	$\circ$	$\circ$
Total Share Prices for All Shares	400 Δ log		$\circ$	$\circ$	$\circ$
Basic Discount Rate	None		$\circ$	$\circ$	$\circ$
Monetary Base	400 Δ log			$\circ$	$\circ$
Nikkei225	400 Δ log			$\circ$	
Equity Market Index Sao Paulo Stock Exchange	400 Δ log				$\circ$

Table 6: ALPL for Macro Variables: $T - T_0 = 50$

Model	**US**	**UK**	**Japan**	**Brazil**
BVAR	-44.99	-50.80	-42.61	-65.80
BVAR-CSV	-37.58	-43.82*	-41.61	-64.55
BVAR-CSV-t	-37.34	-43.95	-41.60*	-64.27
BVAR-CSV-MA	-37.48	-43.93	-41.60*	-64.20**
BVAR-CSV-MA-t	-37.11*	-43.88	-41.74	-64.31
BVAR-CSV-IG	-36.91**	-43.75**	-41.51**	-64.20**

The best model is marked by ** and the second best by *.

Table 7: ALPL for US Macro variables: $T - T_0 = 50$

Model	**US(−9)**	**US(−8)**	**US(−7)**	**US(−6)**	**US(−4)**	**US(−1)**	**US(−12)**
BVAR							-31.426
BVAR−SV	-33.083		-33.448				-32.159
BVAR−CSV	-31.541	-31.805	-31.801	-32.314	-33.266	-35.539	-30.593
BVAR−CSV−t	-31.402	-31.679	-31.682	-32.210	-33.257		-30.468
BVAR−FSV−5	-33.150	-33.815	-33.466	-33.466			-32.406
BVAR−FSV−6	-33.054		-33.248				-32.438
BVAR−FSV−7	-32.935		-32.202				-32.358
BVAR−FSV−t−5	-33.009		-33.245				-32.424
BVAR−FSV−t−6	-33.004	-33.673	-33.222	-34.063			-32.476
BVAR−FSV−t−7	-32.754		-32.963				-32.261
BVAR−AF−5	-32.164	-32.422					-31.309
BVAR−AF−6	-32.022	-32.281					-31.136
BVAR−CSV−IG	-31.507	-31.781	-31.781	-32.301	-33.28	-35.608	-30.582
BVAR−CSV−IG−AF1	-31.091**	-31.432**	-31.361**	-31.965**	-32.891**	-35.226**	-30.338**
BVAR−CSV−IG−AF2	-31.177*	-31.508*	-31.434*	-32.022*	-32.997*	-35.303*	-30.388*
Left out	39-46,48	40-46,48	39-45,48	40-45	40-43	40	39-50

The best model is marked by ** and the second best by *. Observations that are excluded from the ALPL calculation are indicated in the row labeled as 'Left out'. US(-x) means that x observations were excluded. Observation 50 is the last in the evaluation period, and corresponds to 2022Q4.

Table 8: ALPL for UK Macro Variables: $T - T_0 = 50$

Model	UK(−13)	UK(−11)	UK(−10)	UK(−8)	UK(−7)	UK(−3)	UK(−2)	UK(0)	UK(−26)
BVAR	-38.962								-39.426
SV	-35.996	-36.497	-36.607						-36.350
CSV	-37.501	-38.197	-38.511				-42.226		-38.093
CSVt	-37.607	-38.371	-38.675	-39.657	-39.958	-41.592			-38.092
FSV5	-35.696	-36.079	-36.095	-36.758					-36.250
FSV6	-34.958	-35.348	-35.352	-35.905	-36.193**				-35.215
FSV7	-34.591*	-34.955*	-35.019*						-34.932
FSVt5	-35.724	-36.078	-36.076	-36.580					-36.116
FSVt6	-34.658	-35.023	-35.043	-35.511**					-34.891*
FSVt7	-34.543**	-34.940**	-34.969**	-35.534*					-34.620**
AF5	-38.582								-38.875
AF6	-38.693								-39.095
IG	-37.581	-38.292	-38.607	-39.618	-39.906	-41.575	-42.441	-43.850	-38.144
IGAF1	-36.903	-37.608	-37.916	-38.952	-38.279*	-41.013*	-41.851*	-43.298*	-37.530
IGAF2	-36.744	-37.279	-37.611	-38.610	-38.947	-40.713**	-41.551**	-42.965**	-37.464
Left out	13,25-26 39-45 47-48,50	13,25 39-45 47-48	25,39-45 47-48	39-45,47	40-45,47	40,41,43	25,40		25-50

Same notes as in Table 7.

Table 9: ALPL for JP Macro Variables: $T - T_0 = 50$

Model	**JP(−4)**	**JP(−2)**	**JP(−1)**	**JP(0)**	**JP(−11)**
BVAR	-41.1531	-41.4989	-41.7639		-40.2718
BVAR−SV	-37.8063				
BVAR−CSV	-40.1506	-40.6503	-40.9002		-39.1914
BVAR−CSV−t	-40.2664	-40.7527	-40.0054*		-39.2812
BVAR−FSV−5	-37.5361	-37.9137			-36.5313
BVAR−FSV−6	-37.4138	-37.9362			-36.4892
BVAR−FSV−7	-37.4262	-37.8502			-36.3275*
BVAR−FSV−t−5	-37.4325	-37.788	-37.0406**		-36.5316
BVAR−FSV−t−6	-37.2689*	-37.6987*			-36.335
BVAR−FSV−t−7	-37.1979**	-37.5429**			-36.1206**
BVAR−AF−5	-40.37682	-40.74981			-39.48199
BVAR−AF−6	-40.51447	-40.9131	-40.17646		-39.66288
BVAR−CSV−IG	-40.1445	-40.67838	-40.92979	-41.5288	-39.21498
BVAR−CSV−IG−AF1	-40.03713	-40.56063	-40.80496	-41.4039*	-39.13826
BVAR−CSV−IG−AF2	-39.8466	-40.35255	-40.59195	-41.1802**	-38.96878
Left out	16, 40, 42, 44	40,42	40		40-50

Same notes as in Table 7.

Table 10: ALPL for Brazil Macro Variables: $T - T_0 = 50$

Model	BR(−6)	BR(−5)	BR(−3)	BR(−1)	BR(0)	BR(−11)
BVAR						-61.7335
BVAR−SV	-64.2319	-64.5394	-64.9309	-65.6374		-63.6324
BVAR−CSV	-62.2592	-62.4676	-62.9552	-63.5669		-61.4065
BVAR−CSV−t	-62.2325	-62.4328	-62.9268*	-63.5837		-61.3828*
BVAR−FSV−5	-64.9037	-65.1331	-65.4392	-66.17		-64.2763
BVAR−FSV−6	-64.8513	-65.1139	-65.419			-64.2278
BVAR−FSV−7	-64.9138	-65.1633	-65.4911	-66.2781		-64.3289
BVAR−FSV−t−5	-64.8923	-65.1664	-65.4394			-64.3269
BVAR−FSV−t−6	-64.9237	-65.1472	-65.455			-64.3282
BVAR−FSV−t−7	-64.9457	-65.1767	-65.4764			-64.3673
BVAR−AF−5	-62.66618	-62.95486				-61.65061
BVAR−AF−6	-62.59818					-61.57513
BVAR−CSV−IG	-62.13638**	-62.359**	-62.83997**	-63.44601**	-64.1829**	-61.27249**
BVAR−CSV−IG−AF1	-62.21758*	-62.43772*	-62.93142	-63.51524*	-64.2580*	-61.39914
BVAR−CSV−IG−AF2	-62.25439	-62.46895	-62.95001	-63.54592	-64.3026	-61.39902
Left out	40,41,43 44,45,48	40,41 43-45	40,41,44 45	40		40-50

Same notes as in Table 7.

5 Conclusion

We proposed a novel inverse gamma CSV model that implies fat tails for the observed data. We obtained an analytic expression for the likelihood, which facilitates the calculation of the marginal likelihood, permits Maximum Likelihood estimation and exact sampling from the posterior of the volatilities.

We generalized the CSV model by developing an approximate factor model with a common multiplicative factor. This model captures the commonality in volatilities through the common multiplicative factor, but allows for more general patterns by parsimoniously adding other heteroscedastic factors in an approximate factor model structure.

Using data on exchange rates we found that all CSV models greatly outperformed other SV models, and that the best model was the CSV-IG-AF.

Using macro data from four countries we found that the CSV-IG was the best among the CSV models, and that the best models were the CSV-IG-AF, CSV-IG, FSV-t and FSV-t for the US, BR, UK and JP, respectively.

Therefore, we provide further evidence in favor of the CSV structure, as well as methods to effectively improve these models through the fat tails induced by the inverse gamma SV and the additional heteroscedastic factors in the CSV-IG-AF models.

References

Abramowitz, M., Stegun, I. A. and Romer, R. H. (1988) *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* American Association of Physics

Teachers.

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.

Arias, J.E., Rubio-Ramirez, J.F. and Shin, M. (2023) "Macroeconomic forecasting and variable ordering in multivariate stochastic volatility models," *Journal of Econometrics*, 235, 2, 1054–1086.

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

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

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

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

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

Chan, J.C.C. (2023) "Comparing stochastic volatility specifications for large Bayesian VARs," *Journal of Econometrics*, 235, 1419-1446.

Chan, J., Doucet, A., Leon-Gonzalez, R. and R. W. Strachan (2018) "Multivariate Stochastic Volatility with Co-Heteroscedasticity," *GRIPS Discussion Papers 18-12, National Graduate Institute for Policy Studies.*

Chan, J.C.C., Koop, G. and Yu, X. (2023) "Large Order-Invariant Bayesian VARs with Stochastic Volatility" *Journal of Business Economics & Statistics*, 00 (0), 1-13.

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 E. Mertens (2023) "Stochastic volatility in Bayesian vector autoregressions," in *Oxford Research Encyclopedia of Economics and Finance.* Oxford University Press.

Cogley, T., and Sargent, T. J. (2005) "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US," *Review of Economic Dynamics*, 8, 262–302.

Cross, J. and Poon, A. (2016) "Forecasting structural change and fat-tailed events in Australian macroeconomic variables" *Economic Modelling*, 58 , 34-51.

Engle, R.F. (1982) "Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kindom Inflation," *Econometrica*, 50 (4), 987-1007.

Geweke, J. (2004). "Getting it Right: Joint Distribution Tests of Posterior Simulators." *Journal of the American Statistical Association*, 99 (467), 799-804.

Geweke, J. and Amisano, G. (2010). Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns. *International Journal of Forecasting*, 26 , 216-230.

Götz, T. B., and Hauzenberger, K. (2021) "Large mixed-frequency vars with a parsimonious time-varying parameter structure" *The Econometrics Journal*, 24 (3), 442-461.

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

Hartwig, B. (2022) "Bayesian VARs and prior calibration in times of COVID-19," *Studies in Nonlinear Dynamics & Econometrics*, 28(1), 1–24

Hou, C., Nguyen, B. and Zhang, B. (2023) "Real-time forecasting of the Australian macroeconomy using flexible bayesian VARs" *Journal of Forecasting*, 42 (2), 418-451.

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.

Koop, G. M. (2013) "Forecasting with medium and large Bayesian VARs," *Journal of Applied Econometrics*, 28 (2), 177-203.

León-González, R. (2019), "“Efficient Bayesian Inference in Generalized Inverse Gamma Processes for Stochastic Volatility," *Econometric Reviews*, 38, 899-920.

Leon-Gonzalez, R. and Majoni, B. (2023) "Exact Likelihood for Inverse Gamma Stochastic Volatility Models," *Working Paper Series* 23-11, Rimini Centre for Economic Analysis.

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

Mumtaz, H. (2016) "The evolving transmission of uncertainty shocks in the United Kingdom," *Econometrics*, ISSN 2225-1146, MDPI, Basel, Vol. 4, Iss. 1, pp. 1-18.

Mumtaz, H. (2018) "A generalised stochastic volatility in mean VAR", *Economic Letters*, 173, 10-14.

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

Poon, A. (2018) "Assessing the synchronicity and nature of Australian state business cycles," *Economic Record*, 94 (307), 372-390.

Shephard, N. (1994) "Local scale models: State space alternative to integrated GARCH processes," *Journal of Econometrics*, 60 (1-2), 181-202.

Sims, C.A. (1980) "Macroeconomics and reality," *Econometrica*, 48 (1), 1-48.

Sundararajan, R. R. and Barreto-Souza, W. (2023) "Student-t stochastic volatility model with composite likelihood EM-algorithm," *Journal of Time Series Analysis*, 44, 125-147.

Wu, P. and Koop, G. (2022). "Fast, Order-Invariant Bayesian Inference in VARs using the Eigendecomposition of the Error Covariance Matrix," *Working Papers 2310, University of Strathclyde Business School, Department of Economics.*

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

6 Appendix

6.1 Proof of Proposition 2.1

To derive the likelihood we will make use of the following lemma, which is a slightly modified version of Theorem 7.3.4. in Muirhead (2005).

Lemma 6.1.

$$\int |K|^{\frac{n+r-2}{2}} \exp\left(-\frac{1}{2}AK\right){}_0F_1\left(\frac{n}{2};\frac{1}{4}BK\right)dK =$$
$$\Gamma\left(\frac{n+r}{2}\right)\left|\frac{1}{2}A\right|^{-\frac{n+r}{2}} {}_1F_1\left(\frac{n+r}{2};\frac{n}{2};\frac{1}{2}BA^{-1}\right)$$

where ${}_0F_1(.)$ and ${}_1F_1(.)$ are hypergeometric series and A, B, K, n, r are positive scalars.

Proof. The integral is a gamma multiplied by a hypergeometric function. Therefore, the integral is very standard so we can use the properties of hypergeometric functions. We apply Theorem 7.3.4 in Muirhead (2005) to get the result. Thus, we transform the functions by applying a change of variables. Let $X = \frac{1}{4}BK$ such that $K = 4XB^{-1}$ and we have:

$${}_0F_1\left(\frac{n}{2};\frac{1}{4}BK\right) = {}_0F_1\left(\frac{n}{2};X\right)$$

Then the integral can be written as:

$$\int |X|^{\frac{n+r-2}{2}}|4B^{-1}|^{\frac{n+r-2}{2}} \exp\left(-\frac{1}{2}4AXB^{-1}\right){}_0F_1\left(\frac{n}{2};X\right)dK$$

We use the Jacobian $dK = |4B^{-1}|dX$ to integrate with respect to X:

$$\int |X|^{\frac{n+r-2}{2}} \exp(-2XB^{-1}A){}_0F_1\left(\frac{n}{2};X\right)dX|4B^{-1}|^{\frac{n+r}{2}}$$

This integral is the same as in the theorem, therefore, when we integrate out X we get the following:

$$\int |X|^{\frac{n+r-2}{2}} \exp(-X2B^{-1}A){}_0F_1\left(\frac{n}{2};X\right)dX|4B^{-1}|^{\frac{n+r}{2}} =$$
$$\Gamma\left(\frac{n+r}{2}\right)\left|\frac{1}{2}A\right|^{-\frac{n+r}{2}} {}_1F_1\left(\frac{n+r}{2};\frac{n}{2};\frac{1}{2}BA^{-1}\right)$$

□

Using this lemma Proposition 2.1 can be proved as follows.

Proof. To obtain the likelihood for the first observation, we have that k_1 is a gamma, Bauwens et al. (2000) gives the prior density for k_1 as:

$$|k_1|^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2} tr\big(k_1(1-\rho^2)\big)\Big) \frac{1}{c_0} \tag{6.1}$$

where $c_0 = \frac{\Gamma(\frac{n}{2})}{(\frac{1-\rho^2}{2})^{\frac{n}{2}}}$, is a constant and Γ is a gamma function. Let $V_1^{-1} = (1-\rho^2)$, thus, the likelihood for the first observation is as follows:

$$\begin{aligned} L(Y_1) &= \int L(Y_1 \mid k_1)\pi(k_1) dk_1 \\ &= \int (2\pi)^{-\frac{r}{2}} |\Sigma|^{-\frac{1}{2}} k_1^{\frac{r}{2}} \exp\Big(-\frac{1}{2}\varepsilon_1^2 k_1\Big) k_1^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}(1-\rho^2)k_1\Big)\frac{1}{c_0} dk_1 \end{aligned} \tag{6.2}$$

The integral is with respect to k_1, so after rearranging and combining like terms we have:

$$L(Y_1) = \int (2\pi)^{-\frac{r}{2}} |\Sigma|^{-\frac{1}{2}} k_1^{\frac{n+r-2}{2}} \exp\Big(-\frac{1}{2}(\varepsilon_1^2 + V_1^{-1})k_1\Big)\frac{1}{c_0} dk_1$$

where $k_1^{\frac{n+r-2}{2}} \exp(-\frac{1}{2}(\varepsilon_1^2 + V_1^{-1})k_1)$ is the kernel of a gamma with $n+r$ degrees of freedom. Let $\widetilde{V}_2 = (\varepsilon_1^2 + V_1^{-1})^{-1}$, therefore, the density of $k_1|Y_1$ is:

$$\pi(k_1|Y_1) = k_1^{\frac{n+r-2}{2}} \exp\Big(-\frac{1}{2}k_1 \widetilde{V}_2^{-1}\Big)\frac{1}{\bar{c_0}} \tag{6.3}$$

with $\bar{c_0} = \frac{\Gamma(\frac{n+r}{2})}{(\frac{\widetilde{V}_2^{-1}}{2})^{\frac{n+r}{2}}}$. Thus, we have the likelihood as:

$$L(Y_1) = (2\pi)^{-\frac{r}{2}} |\Sigma|^{-\frac{1}{2}} \Gamma\Big(\frac{n+r}{2}\Big) \Big|\frac{\varepsilon_1^2 + V_1^{-1}}{2}\Big|^{-\frac{n+r}{2}} \frac{1}{c_0}$$

Taking into account c_0 we can write the likelihood for $t = 1$ as:

$$L(Y_1) = (2\pi)^{-\frac{r}{2}} |\Sigma|^{-\frac{1}{2}} 2^{\frac{r}{2}} \frac{\Gamma(\frac{n+r}{2})}{\Gamma(\frac{n}{2})} |\varepsilon_1^2 + V_1^{-1}|^{-\frac{n+r}{2}} V_1^{-\frac{n}{2}}$$

Define $k_{1:2} = (k_1, k_2)$, then we have the likelihood for the second observation as:

$$L(Y_2|Y_1) = \int L(Y_2|k_{1:2}, Y_1)\pi(k_{1:2}|Y_1) dk_{1:2}$$

where $\pi(k_{1:2}|Y_1) = \pi(k_1|Y_1)\pi(k_2|k_1, Y_1)$. The prior for k_t unconditionally is a gamma. However, $k_t|k_{t-1}$ is a non central chi-squared. Muirhead (2005, p. 442) gives this non central chi-squared density as follows:

$$\pi(k_t|k_{t-1}) = k_t^{\frac{n-2}{2}} \exp\Big(-\frac{1}{2}k_t\Big)\,{}_0F_1\Big(\frac{n}{2};\frac{1}{4}\rho^2 k_{t-1}k_t\Big)\exp\Big(-\frac{1}{2}\rho^2 k_{t-1}\Big)\Big(\Gamma\Big(\frac{n}{2}\Big)\Big)^{-1}\frac{1}{c} \quad (6.4)$$

where ${}_0F_1$ is a hypergeometric function, $\rho^2 k_{t-1}$ is the non-centrality parameter and $c = 2^{\frac{n}{2}}$. Then we can write the likelihood for the second observation conditional on the first as:

$$L(Y_2|Y_1) = \int (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}} k_2^{\frac{r}{2}} \exp\Big(-\frac{1}{2}\varepsilon_2^2 k_2\Big)\pi(k_{1:2}|Y_1)dk_{1:2} \quad (6.5)$$

We integrate first with respect to k_1. Define l_2 as representing all the elements in (6.4) that do not depend on k_1 as follows:

$$l_2 = \Big(k_2^{\frac{n-2}{2}}\exp\Big(-\frac{1}{2}k_2\Big)\Big)^{-1}\Big(\frac{1}{\Gamma(\frac{n}{2})}\Big)^{-1}\Big(\frac{1}{c}\Big)^{-1} \quad (6.6)$$

Given that $\pi(k_2|k_1, Y_1) = \pi(k_2|k_1)$, and given (6.4) and (6.3), we can write $\pi(k_2|Y_1)$ as follows:

$$\pi(k_2|Y_1) = \int \pi(k_2|k_1, Y_1)\pi(k_1|Y_1)dk_1 =$$
$$\frac{1}{\bar{c}_0}\int k_1^{\frac{n+r-2}{2}} \exp\Big(-\frac{1}{2}(\widetilde{V}_2^{-1}k_1)\Big)\exp\Big(-\frac{1}{2}(\rho^2 k_1)\Big)\,{}_0F_1\Big(\frac{n}{2};\frac{1}{4}\rho^2 k_1 k_2\Big)\frac{1}{l_2}dk_1$$

where we have used the expression for $\pi(k_1|Y_1)$ in (6.3). We can write the above integral more compactly as:

$$\int \pi(k_2|k_1, Y_1)\pi(k_1|Y_1)dk_1 = \int \frac{1}{\bar{c}_0}k_1^{\frac{n+r-2}{2}}\exp\Big(-\frac{1}{2}(\widetilde{V}_2^{-1}+\rho^2)k_1\Big)\,{}_0F_1\Big(\frac{n}{2};\frac{1}{4}\rho^2 k_1 k_2\Big)\frac{1}{l_2}dk_1$$

Applying Lemma 6.1 the solution to this integral is:

$$\pi(k_2|Y_1) = \int \pi(k_2|k_1, Y_1)\pi(k_1|Y_1)dk_1 =$$
$$\frac{1}{\bar{c}_0}\Gamma\Big(\frac{n+r}{2}\Big)\Big|\frac{\widetilde{V}_2^{-1}+\rho^2}{2}\Big|^{-\frac{n+r}{2}}\,{}_1F_1\Big(\frac{n+r}{2};\frac{n}{2};\frac{1}{2}k_2\rho^2(\widetilde{V}_2^{-1}+\rho^2)^{-1}\Big)\frac{1}{l_2} \quad (6.7)$$

Given (6.6) and (6.7), the distribution of $k_2|Y_1$ is a mixture of gammas as follows:

$$\pi(k_2|Y_1) \propto k_2^{\frac{n-2}{2}} \exp\Bigg(-\frac{1}{2}k_2\Bigg)\,{}_1F_1\Bigg(\frac{n+r}{2};\frac{n}{2};\frac{1}{2}k_2\rho^2(\widetilde{V}_2^{-1}+\rho^2)^{-1}\Bigg) \tag{6.8}$$

The normalising constant for this density function can be obtained in closed form by applying Muirhead (2005, p. 260):

$$\int k_2^{\frac{n-2}{2}} \exp\Bigg(-\frac{1}{2}k_2\Bigg)\,{}_1F_1\Bigg(\frac{n+r}{2};\frac{n}{2};\frac{1}{2}k_2\delta_2\Bigg)dk_2 = \Gamma\Bigg(\frac{n}{2}\Bigg)2^{\frac{n}{2}}\,{}_2F_1\Bigg(\frac{n+r}{2},\frac{n}{2};\frac{n}{2};\delta_2\Bigg) \tag{6.9}$$

where $\delta_2 = \rho^2(\widetilde{V}_2^{-1}+\rho^2)^{-1}$. This ${}_2F_1\big(\frac{n+r}{2},\frac{n}{2};\frac{n}{2};\delta_2\big)$ function has the same terms in the denominator and the numerator thus they cancel out and we have:

$${}_2F_1\Bigg(\frac{n+r}{2},\frac{n}{2};\frac{n}{2};\delta_2\Bigg) = {}_1F_0\Bigg(\frac{n+r}{2};\delta_2\Bigg) \tag{6.10}$$

This function simplifies to a known solution for $|\delta_2| < 1$, see Muirhead (2005, p. 261) .

$${}_1F_0\Bigg(\frac{n+r}{2};\delta_2\Bigg) = (1-\delta_2)^{-\frac{n+r}{2}} \tag{6.11}$$

The normalising constant becomes:

$$\Gamma\Bigg(\frac{n}{2}\Bigg)2^{\frac{n}{2}}\,{}_1F_0\Bigg(\frac{n+r}{2};\delta_2\Bigg) = \Gamma\Bigg(\frac{n}{2}\Bigg)2^{\frac{n}{2}}(1-\delta_2)^{-\frac{n+r}{2}}$$

Given this normalising constant, we have the density for $\pi(k_2|Y_1)$ from (6.8) as follows:

$$\pi(k_2|Y_1) = \frac{1}{c_1}k_2^{\frac{n-2}{2}} \exp\Bigg(-\frac{1}{2}k_2\Bigg)\,{}_1F_1\Bigg(\frac{n+r}{2};\frac{n}{2};\frac{1}{2}k_2\rho^2(\widetilde{V}_2^{-1}+\rho^2)^{-1}\Bigg)$$

where $c_1 = \Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}(1-\delta_2)^{-\frac{n+r}{2}}$. Thus, the likelihood for the second observation is as follows:

$$\begin{aligned} L(Y_2|Y_1) &= \int \pi(Y_2|k_2,Y_1)\pi(k_2|Y_1)dk_2 \\ &= \int (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}k_2^{\frac{n+r-2}{2}} \exp\Bigg(-\frac{1}{2}(\varepsilon_2^2+1)k_2\Bigg)\frac{1}{c_1}\,{}_1F_1\Bigg(\frac{n+r}{2};\frac{n}{2};\frac{1}{2}k_2\rho^2(\widetilde{V}_2^{-1}+\rho^2)^{-1}\Bigg)dk_2 \end{aligned}$$

Using Muirhead (2005, p. 261) and taking into account c_1, the likelihood for the second observation is:

$$L(Y_2|Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}} \frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}} \frac{\Gamma\big(\frac{n+r}{2}\big)}{\Gamma\big(\frac{n}{2}\big)} \frac{(\varepsilon_2^2+1)^{-\frac{n+r}{2}}}{(1-\delta_2)^{-\frac{n+r}{2}}} {}_2F_1\bigg(\frac{n+r}{2}, \frac{n+r}{2}; \frac{n}{2}; (\varepsilon_2^2+1)^{-1}\delta_2\bigg)$$

Thus we get a Gauss hypergeometric function which can be evaluated easily. Let $Z_2 = (\varepsilon_2^2+1)^{-1}\delta_2$. This series converges because $|Z_2| < 1$ (Abramowitz et al. (1988)). To accelerate the convergence of this series we apply the Euler transformation as in Abramowitz et al. (1988, p. 559) and thus we get:

$$ {}_2F_1\bigg(\frac{n+r}{2}, \frac{n+r}{2}; \frac{n}{2}; Z_2\bigg) = (1-Z_2)^{-\frac{n+2r}{2}} {}_2F_1\bigg(-\frac{r}{2}, -\frac{r}{2}; \frac{n}{2}; Z_2\bigg) \tag{6.12}$$

Thus $\hat{C}_2 = (1-Z_2)^{-\frac{n+2r}{2}} {}_2F_1\big(-\frac{r}{2}, -\frac{r}{2}; \frac{n}{2}; Z_2\big)$, then we can write the $L(Y_2|Y_1)$ as follows:

$$L(Y_2|Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}} \frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}} \frac{\Gamma\big(\frac{n+r}{2}\big)}{\Gamma\big(\frac{n}{2}\big)} \frac{(\varepsilon_2^2+1)^{-\frac{n+r}{2}}}{(1-\delta_2)^{-\frac{n+r}{2}}} \hat{C}_2$$

The density of k_t for the third observation is given by:

$$\pi(k_3|Y_2, Y_1) = \int \pi(k_3|k_2)\pi(k_2|Y_2, Y_1) dk_2$$

where $\pi(k_2|Y_2, Y_1) \propto \pi(k_2|Y_1)L(Y_2|k_2, Y_1)$. The distribution for $\pi(k_2|Y_1)$ in (6.8) can be written as follows:

$$\pi(k_2|Y_1) \propto \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} k_2^{\frac{n+2h_2-2}{2}} \exp\bigg(-\frac{1}{2}k_2\bigg)$$

where $\widetilde{C}_{2,h_2} = \frac{[(n+r)/2]_{h_2}}{[n/2]_{h_2}} \big(\frac{1}{2}\rho^2(\widetilde{V}_2^{-1} + \rho^2)^{-1}\big)^{h_2} \frac{1}{h_2!}$. Thus we have:

$$\pi(k_2|Y_2, Y_1) \propto \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} k_2^{\frac{n+r+2h_2-2}{2}} \exp\bigg(-\frac{1}{2}k_2(\varepsilon_2^2+1)\bigg) \tag{6.13}$$

Given (6.4) and (6.13) we have:

$$\begin{aligned}\pi(k_3|Y_2, Y_1) \propto \int & k_3^{\frac{n-2}{2}} \exp\bigg(-\frac{1}{2}k_3\bigg) {}_0F_1\bigg(\frac{n}{2}; \frac{1}{4}\rho^2 k_2 k_3\bigg) \exp\bigg(-\frac{1}{2}\rho^2 k_2\bigg) \\ & \times \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} k_2^{\frac{n+r+2h_2-2}{2}} \exp\bigg(-\frac{1}{2}k_2(\varepsilon_2^2+1)\bigg) \frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}} dk_2\end{aligned}$$

which simplifies to:

$$\pi(k_3|Y_2,Y_1) \propto \int k_3^{\frac{n-2}{2}} \exp\left(-\frac{1}{2}k_3\right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4}\rho^2 k_2 k_3\right) \exp\left(-\frac{1}{2}(\varepsilon_2^2+1+\rho^2)k_2\right)$$
$$\sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} k_2^{\frac{n+r+2h_2-2}{2}} \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} dk_2$$

Using Lemma 6.1 the density of $k_3|Y_2, Y_1$ is thus:

$$\begin{aligned}\pi(k_3|Y_2,Y_1) = \frac{1}{c_3} k_3^{\frac{n-2}{2}} \exp\left(-\frac{1}{2}k_3\right) \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} \Gamma\left(\frac{n+r+2h_2}{2}\right) \\ {}_1F_1\left(\frac{n+r+2h_2}{2}; \frac{n}{2}; \frac{1}{2}k_3\rho^2 S_3\right)(2S_3)^{\frac{n+r+2h_2}{2}} \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}}\end{aligned} \tag{6.14}$$

where $S_3 = (\varepsilon_2^2+1+\rho^2)^{-1}$ and c_3 is the normalising constant as in (6.9) as follows:

$$c_3 = \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} \Gamma\left(\frac{n+r+2h_2}{2}\right)(2S_3)^{\frac{n+r+2h_2}{2}} {}_2F_1\left(\frac{n+r+2h_2}{2}, \frac{n}{2}; \frac{n}{2}; \rho^2 S_3\right)$$

Similar to (6.10) and (6.11), the hypergeometric function simplifies to get:

$$c_3 = \sum_{h_2=0}^{\infty} \widetilde{C}_{2,h_2} \Gamma\left(\frac{n+r+2h_2}{2}\right)(2S_3)^{\frac{n+r+2h_2}{2}} (1-\rho^2 S_3)^{-\frac{n+r+2h_2}{2}}$$

Collecting terms dependent on h_2 we can write c_3 as

$$c_3 = \left(\sum_{h_2=0}^{\infty} \frac{[(n+r)/2]_{h_2}}{[n/2]_{h_2}} [(n+r)/2]_{h_2} \frac{\delta_3^{h_2}}{h_2!}\right) \Gamma\left(\frac{n+r}{2}\right)(1-\rho^2 S_3)^{-\frac{n+r}{2}}(2S_3)^{\frac{n+r}{2}}$$

where $\delta_3 = \left((1-\rho^2 S_3)^{-1} S_3 \rho^2 (\widetilde{V}_2^{-1} + \rho^2)^{-1}\right)$. This can be written as:

$$c_3 = {}_2F_1\left(\frac{n+r}{2}, \frac{n+r}{2}; \frac{n}{2}; \delta_3\right) \Gamma\left(\frac{n+r}{2}\right)(1-\rho^2 S_3)^{-\frac{n+r}{2}}(2S_3)^{\frac{n+r}{2}}$$

Using Euler's acceleration in (6.12) we have therefore:

$$c_3 = (1-\delta_3)^{-\frac{n+2r}{2}} {}_2F_1\left(-\frac{r}{2}, -\frac{r}{2}; \frac{n}{2}; \delta_3\right) \Gamma\left(\frac{n+r}{2}\right)(1-\rho^2 S_3)^{-\frac{n+r}{2}}(2S_3)^{\frac{n+r}{2}}$$

Therefore the likelihood for $t = 3$ is as follows:

$$L(Y_3|Y_2, Y_1) = \int \pi(Y_3|k_3, Y_2, Y_1)\pi(k_3|Y_2, Y_1)dk_3$$

Thus we have from (6.14)

$$L(Y_3|Y_2, Y_1) = \int (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_3}k_3^{\frac{n+r-2}{2}} \exp\bigg(-\frac{1}{2}k_3(\varepsilon_3^2+1)\bigg)\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\Gamma\bigg(\frac{n+r+2h_2}{2}\bigg)$$
$$ {}_1F_1\bigg(\frac{n+r+2h_2}{2};\frac{n}{2};\frac{1}{2}k_3\rho^2S_3\bigg)(2S_3)^{\frac{n+r+2h_2}{2}}\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}dk_3$$

and we get (Muirhead (2005, p. 260)):

$$L(Y_3|Y_2, Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_3}\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\Gamma\bigg(\frac{n+r+2h_2}{2}\bigg)(2S_3)^{\frac{n+r+2h_2}{2}}\Gamma\bigg(\frac{n+r}{2}\bigg)2^{\frac{n+r}{2}}$$
$$(\varepsilon_3^2+1)^{-\frac{n+r}{2}}{}_2F_1\bigg(\frac{n+r+2h_2}{2},\frac{n+r}{2};\frac{n}{2};(\varepsilon_3^2+1)^{-1}\rho^2S_3\bigg)\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}$$

Letting $Z_3 = (\varepsilon_3^2+1)^{-1}\rho^2S_3$, and defining $\hat{C}_3 = {}_2F_1\bigg(\frac{n+r+2h_2}{2},\frac{n+r}{2};\frac{n}{2};Z_3\bigg)$, we have that:

$$L(Y_3|Y_2, Y_1) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_3}\sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\frac{\Gamma\bigg(\frac{n+r+2h_2}{2}\bigg)}{(\varepsilon_3^2+1)^{\frac{n+r}{2}}}(2S_3)^{\frac{n+r+2h_2}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\big(\frac{n+r}{2}\big)}{\Gamma\big(\frac{n}{2}\big)}\hat{C}_3$$

The density for the fourth observation is given by:

$$\pi(k_4|Y_3, Y_2, Y_1) = \int \pi(k_4|k_3, Y_1, Y_2, Y_3)\pi(k_3|Y_3, Y_2, Y_1)dk_3 \tag{6.15}$$

with $\pi(K_3|Y_3, Y_2, Y_1) \propto \pi(K_3|Y_2, Y_1)L(Y_3|Y_2, Y_1)$. Let:

$$\widetilde{C}_{3,h_3} = \sum_{h_2=0}^{\infty}\widetilde{C}_{2,h_2}\Gamma\bigg(\frac{n+r+2h_2}{2}\bigg)\frac{[(n+r)/2+h_2]_{h_3}}{[n/2]_{h_3}}\bigg(\frac{1}{2}\rho^2S_3\bigg)^{h_3}\frac{1}{h_3!}(2S_3)^{\frac{n+r+2h_2}{2}} \tag{6.16}$$

Then from (6.14) we have:

$$\pi(k_3|Y_2, Y_1) \propto \sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}k_3^{\frac{n+2h_3-2}{2}}\exp\bigg(-\frac{1}{2}k_3\bigg)$$

As before, when we include the third observation, the distribution of $k_3|Y_3, Y_2, Y_1$ is a mixture of gammas and can be written as follows:

$$\pi(k_3|Y_3, Y_2, Y_1) \propto \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} k_3^{\frac{n+r+2h_3-2}{2}} \exp\left(-\frac{1}{2}k_3(\varepsilon_3^2+1)\right)$$

Let $\widetilde{V_4}^{-1} = (\varepsilon_3^2 + 1)$. Then, using (6.15) and (6.4), we have the distribution of $k_4|Y_3, Y_2, Y_1$ as follows:

$$\begin{aligned}\pi(k_4|Y_3, Y_2, Y_1) \propto \int k_4^{\frac{n-2}{2}} \exp\left(-\frac{1}{2}k_4\right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4}\rho^2 k_3 k_4\right) \exp\left(-\frac{1}{2}\rho^2 k_3\right) \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} \\ \times \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} k_3^{\frac{n+r+2h_3-2}{2}} \exp\left(-\frac{1}{2}k_3\widetilde{V_4}^{-1}\right) dk_3\end{aligned} \tag{6.17}$$

Taking this integral with respect to k_3 we get:

$$\begin{aligned}\pi(k_4|Y_3, Y_2, Y_1) \propto k_4^{\frac{n-2}{2}} \exp\left(-\frac{1}{2}k_4\right) \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} {}_1F_1\left(\frac{n+r+2h_3}{2}; \frac{n}{2}; \frac{1}{2}\rho^2 k_4 (\widetilde{V_4}^{-1} + \rho^2)^{-1}\right) \\ \Gamma\left(\frac{n+r+2h_3}{2}\right)(2S_4)^{\frac{n+r+2h_3}{2}} \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}}\end{aligned}$$

where $S_4 = (\widetilde{V_4}^{-1} + \rho^2)^{-1} = (\varepsilon_3^2 + 1 + \rho^2)^{-1}$. Let c_4 be the normalising constant, that is:

$$\begin{aligned}c_4 = \int k_4^{\frac{n-2}{2}} \exp\left(-\frac{1}{2}k_4\right) \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} {}_1F_1\left(\frac{n+r+2h_3}{2}; \frac{n}{2}; \frac{1}{2}\rho^2 k_4 (\widetilde{V_4}^{-1} + \rho^2)^{-1}\right) \\ \Gamma\left(\frac{n+r+2h_3}{2}\right)(2S_4)^{\frac{n+r+2h_3}{2}} \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} dk_4\end{aligned}$$

Thus we get:

$$c_4 = \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} {}_2F_1\left(\frac{n+r+2h_3}{2}, \frac{n}{2}; \frac{n}{2}; \rho^2 S_4\right)\Gamma\left(\frac{n+r+2h_3}{2}\right)(2S_4)^{\frac{n+r+2h_3}{2}}$$

Using (6.10) and (6.11), this simplifies to:

$$c_4 = \sum_{h_3=0}^{\infty} \widetilde{C}_{3,h_3} (1 - \rho^2 S_4)^{-\frac{n+r+2h_3}{2}} \Gamma\left(\frac{n+r+2h_3}{2}\right)(2S_4)^{\frac{n+r+2h_3}{2}}$$

Thus,

$$\pi(k_4|Y_3,Y_2,Y_1) = \frac{1}{c_4}k_4^{\frac{n-2}{2}} \exp\bigg(-\frac{1}{2}k_4\bigg)\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}\,{}_1F_1\bigg(\frac{n+r+2h_3}{2};\frac{n}{2};\frac{1}{2}\rho^2k_4(\widetilde{V}_4^{-1}+\rho^2)^{-1}\bigg)$$
$$\Gamma\bigg(\frac{n+r+2h_3}{2}\bigg)(2S_4)^{\frac{n+r+2h_3}{2}}\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}$$

Therefore the likelihood for $t=4$ is as follows:

$$L(Y_4|Y_3,Y_2,Y_1) = \int \pi(Y_4|k_4,Y_3,Y_2,Y_1)\pi(k_4|Y_3,Y_2,Y_1)dk_4$$

Thus we have:

$$L(Y_4|Y_3,Y_2,Y_1) = \int (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_4}k_4^{\frac{n+r-2}{2}}\exp\bigg(-\frac{1}{2}k_4(\varepsilon_4^2+1)\bigg)\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}\Gamma\bigg(\frac{n+r+2h_3}{2}\bigg)$$
$${}_1F_1\bigg(\frac{n+r+2h_3}{2};\frac{n}{2};\frac{1}{2}k_4\rho^2S_4\bigg)(2S_4)^{\frac{n+r+2h_3}{2}}\frac{1}{\Gamma\big(\frac{n}{2}\big)2^{\frac{n}{2}}}dk_4$$

This is similar to $t=3$ therefore we have:

$$L(Y_4|Y_3,Y_2,Y_1,\Sigma) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_4}\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h_3}\frac{\Gamma\bigg(\frac{n+r+2h_3}{2}\bigg)}{(\varepsilon_4^2+1)^{\frac{n+r}{2}}}(2S_4)^{\frac{n+r+2h_3}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\big(\frac{n+r}{2}\big)}{\Gamma\big(\frac{n}{2}\big)}\hat{C}_4$$

and the likelihood for any t is:

$$L(Y_t|Y_{1:t-1}) = (2\pi)^{-\frac{r}{2}}|\Sigma|^{-\frac{1}{2}}\frac{1}{c_t}\sum_{h_{t-1}=0}^{\infty}\widetilde{C}_{t-1,h_{t-1}}\frac{\Gamma\bigg(\frac{n+r+2h_{t-1}}{2}\bigg)}{(\varepsilon_t^2+1)^{\frac{n+r}{2}}}(2S_t)^{\frac{n+r+2h_{t-1}}{2}}\frac{2^{\frac{n+r}{2}}}{2^{\frac{n}{2}}}\frac{\Gamma\big(\frac{n+r}{2}\big)}{\Gamma\big(\frac{n}{2}\big)}\hat{C}_t$$

where for $t \geq 4$:

$$\delta_t = \Big((1-\rho^2 S_t)^{-1} S_t \rho^2 (\widetilde{V}_{t-1}^{-1} + \rho^2)^{-1} \Big)$$
$$Z_t = (\varepsilon_t^2 + 1)^{-1} S_t \rho^2$$
$$\hat{C}_t = {}_2F_1\left(\frac{n+r+2h_{t-1}}{2}, \frac{n+r}{2}; \frac{n}{2}; Z_t \right)$$
$$\widetilde{V}_t^{-1} = 1 + \varepsilon_{t-1}^2$$
$$S_t = (\varepsilon_{t-1}^2 + 1 + \rho^2)^{-1} = (\widetilde{V}_t^{-1} + \rho^2)^{-1}$$
$$c_t = \sum_{h_{t-1}=0}^{\infty} \widetilde{C}_{t-1,h_{t-1}} (1-\rho^2 S_t)^{-\frac{n+r+2h_{t-1}}{2}} \Gamma\left(\frac{n+r+2h_{t-1}}{2} \right) (2S_t)^{\frac{n+r+2h_{t-1}}{2}}$$

$$\widetilde{C}_{t-1,h_{t-1}} =$$
$$\sum_{h_{t-2}=0}^{\infty} \widetilde{C}_{t-2,h_{t-2}} \Gamma\left(\frac{n+r+2h_{t-2}}{2} \right) \frac{[(n+r)/2 + h_{t-2}]_{h_{t-1}}}{[n/2]_{h_{t-1}}} \left(\frac{1}{2}\rho^2 S_{t-1} \right)^{h_{t-1}} \frac{(2S_{t-1})^{\frac{n+r+2h_{t-2}}{2}}}{h_{t-1}!}$$

□

6.2 Proof of Proposition 2.2

Proof. Combining the prior density for k_1 in (6.1) with the transition equation in (6.4) and the likelihood, we get:

$$\begin{aligned} \pi(k_1 | k_{2:T}, Y_{1:T}) &\propto |k_1|^{\frac{n+r-2}{2}} \exp\left(-\frac{1}{2} S_2^{-1} k_1 \right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4}\rho^2 k_1 k_2 \right) \\ &= |k_1|^{\frac{n+r-2}{2}} \exp\left(-\frac{1}{2} S_2^{-1} k_1 \right) \sum_{h=0}^{\infty} \left(C_{1,h} |k_1|^h \right) \end{aligned} \tag{6.18}$$

with $C_{1,h} = \frac{1}{h!} \frac{1}{[n/2]_h} \left(\frac{1}{4}\rho^2 k_2 \right)^h$. The integral of (6.18) with respect to k_1 is proportional to:

$${}_1F_1\left(\frac{n+r}{2}; \frac{n}{2}; \frac{1}{2}\rho^2 k_2 S_2 \right)$$

and therefore:

$$\begin{aligned} &\pi(k_2 | k_{3:T}, Y_{1:T}) \\ &\propto |k_2|^{\frac{n+r-2}{2}} \exp\left(-\frac{1}{2} S_3^{-1} k_2 \right) {}_1F_1\left(\frac{n+r}{2}; \frac{n}{2}; \frac{1}{2}\rho^2 k_2 S_2 \right) {}_0F_1\left(\frac{n}{2}; \frac{1}{4}\rho^2 k_3 k_2 \right) \end{aligned} \tag{6.19}$$

where we have used that $S_3^{-1} = \varepsilon_2^2 + 1 + \rho^2$. Combining the series we get that:

$$
\begin{aligned}
&{}_1F_1\left(\frac{n+r}{2};\frac{n}{2};\frac{1}{2}\rho^2 k_2 S_2\right){}_0F_1\left(\frac{n}{2};\frac{1}{4}\rho^2 k_3 k_2\right) = \\
&\left(\sum_{h_1=0}^{\infty}\frac{[(n+r)/2]_{h_1}}{[n/2]_{h_1}}\frac{(\frac{1}{2}\rho^2 S_2)^{h_1}k_2^{h_1}}{h_1!}\right)\left(\sum_{h_2=0}^{\infty}\frac{1}{h_2!}\frac{1}{[n/2]_{h_2}}\left(\frac{1}{4}\rho^2 k_3\right)^{h_2}k_2^{h_2}\right)
\end{aligned}
\tag{6.20}
$$

By making the change of variables $h = h_1 + h_2$ we get that (6.20) can be written as:

$$
\sum_{h=0}^{\infty}\sum_{h_2=0}^{h}\left(\left(\frac{[(n+r)/2]_{h-h_2}}{[n/2]_{h-h_2}}\frac{(\frac{1}{2}\rho^2 S_2)^{h-h_2}}{(h-h_2)!}\right)\frac{1}{h_2!}\frac{1}{[n/2]_{h_2}}\left(\frac{1}{4}\rho^2\right)^{h_2}k_3^{h_2}\right)k_2^h = \sum_{h=0}^{\infty}C_{2,h}k_2^h \tag{6.21}
$$

where:

$$
C_{2,h} = \sum_{h_2=0}^{h}\widetilde{C}_{2,h-h_2}\frac{1}{h_2!}\frac{1}{[n/2]_{h_2}}\left(\frac{1}{4}\rho^2\right)^{h_2}k_3^{h_2}
$$

and $\widetilde{C}_{2,h-h_2}$ has been defined in proposition 3.1 as:

$$
\widetilde{C}_{2,h-h_2} = \frac{[(n+r)/2]_{h-h_2}}{[n/2]_{h-h_2}}\frac{(\frac{1}{2}\rho^2 S_2)^{h-h_2}}{(h-h_2)!}
$$

Using (6.21) we obtain that:

$$
\pi(k_2|k_{3:T}, Y_{1:T}) \propto |k_2|^{\frac{n+r-2}{2}}\exp\left(-\frac{1}{2}S_3^{-1}k_2\right)\sum_{h=0}^{\infty}\left(C_{2,h}k_2^h\right) \tag{6.22}
$$

as we wanted to prove. The integral of (6.22) with respect to k_2 is proportional to:

$$
\sum_{h=0}^{\infty}\left(C_{2,h}\frac{\Gamma\left(\frac{n+r+2h}{2}\right)}{(S_3^{-1}/2)^{\frac{n+r+2h}{2}}}\right) = \sum_{h=0}^{\infty}\left(\sum_{h_2=0}^{h}\widetilde{C}_{2,h-h_2}\frac{1}{h_2!}\frac{1}{[n/2]_{h_2}}\left(\frac{1}{4}\rho^2\right)^{h_2}k_3^{h_2}\right)\frac{\Gamma\left(\frac{n+r+2h}{2}\right)}{(S_3^{-1}/2)^{\frac{n+r+2h}{2}}} \tag{6.23}
$$

Making the change of variables $h_1 = h - h_2$, equation (6.23) can be written as:

$$
\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left(\widetilde{C}_{2,h_1}\frac{1}{h_2!}\frac{1}{[n/2]_{h_2}}\left(\frac{1}{4}\rho^2\right)^{h_2}k_3^{h_2}\right)\frac{\Gamma\left(\frac{n+r}{2}+h_1+h_2\right)}{(S_3^{-1}/2)^{\frac{n+r}{2}+h_1+h_2}} \tag{6.24}
$$

Note that $\Gamma\left(\frac{n+r}{2}+h_1+h_2\right)=\Gamma\left(\frac{n+r+2h_1}{2}\right)\left[\frac{n+r+2h_1}{2}\right]_{h_2}$. Then (6.24) can be written as:

$$\sum_{h_2=0}^{\infty}\sum_{h_1=0}^{\infty}\widetilde{C}_{2,h_1}\Gamma\left(\frac{n+r+2h_1}{2}\right)\frac{\left[(n+r)/2+h_1\right]_{h_2}}{[n/2]_{h_2}}\left(\frac{1}{2}\rho^2 S_3\right)^{h_2}\frac{1}{h_2!}(2S_3)^{\frac{n+r+2h_1}{2}}k_3^{h_2} \tag{6.25}$$

Using the definition of $\widetilde{C}_{3,h_2}$ in proposition 3.1, we can write (6.25) as:

$$\sum_{h_2=0}^{\infty}\widetilde{C}_{3,h_2}k_3^{h_2}$$

Recall that the transition density is in (6.4). Therefore, we have:

$$\pi(k_3|k_{4:T},Y_{1:T})\propto\left(\sum_{h_2=0}^{\infty}\widetilde{C}_{3,h_2}k_3^{h_2}\right){}_0F_1\left(\frac{n}{2};\frac{1}{4}\rho^2 k_3k_4\right)|k_3|^{\frac{n+r-2}{2}}\exp\left(-\frac{1}{2}S_4^{-1}k_3\right)$$

with $S_4^{-1}=\varepsilon_3^2+1+\rho^2$. As before, we can multiply the two series as follows:

$$\left(\sum_{h_2=0}^{\infty}\widetilde{C}_{3,h_2}k_3^{h_2}\right){}_0F_1\left(\frac{n}{2};\frac{1}{4}\rho^2 k_3k_4\right)=\left(\sum_{h_2=0}^{\infty}\widetilde{C}_{3,h_2}k_3^{h_2}\right)\left(\sum_{h_3=0}^{\infty}\frac{1}{[n/2]_{h_3}}\left(\frac{1}{4}\rho^2k_3\right)^{h_3}k_4^{h_3}\frac{1}{h_3!}\right)$$
$$=\sum_{h=0}^{\infty}\sum_{h_3=0}^{h}|k_3|^h\widetilde{C}_{3,h-h_3}\frac{1}{[n/2]_{h_3}}\left(\frac{1}{4}\rho^2\right)^{h_3}k_4^{h_3}\frac{1}{h_3!}=\sum_{h=0}^{\infty}|k_3|^hC_{3,h}$$

where

$$C_{3,h}=\sum_{h_3=0}^{\infty}\widetilde{C}_{3,h-h_3}\frac{1}{[n/2]_{h_3}}\left(\frac{1}{4}\rho^2\right)^{h_3}\frac{k_4^{h_3}}{h_3!}$$

and therefore, $\pi(k_3|k_{4:T},Y_{1:T})$ can be written as:

$$\pi(k_3|k_{4:T},Y_{1:T})\propto|k_3|^{\frac{n+r-2}{2}}\exp\left(-\frac{1}{2}S_4^{-1}k_3\right)\sum_{h=0}^{\infty}|k_3|^hC_{3,h} \tag{6.26}$$

as we wanted to prove. Since $\pi(k_3|k_{4:T},Y_{1:T})$ in (6.26) and $\pi(k_2|k_{3:T},Y_{1:T})$ in (6.22) have the same structure, and, since the transition density of k_t is always the same, we get analogous results for any $t<T$, as we wanted to prove. For $t=T$ the only difference is that there is no transition density from k_T to k_{T+1}. For this reason we do not need to multiply two series, and hence $C_{T,h}=\widetilde{C}_{T,h}$ and $S_{T+1}=(\varepsilon_T^2+1)^{-1}$. $\square$

6.3 Geweke Test

To test the implementation of the algorithm that we use to draw the unknown time varying volatilities $k_{1:T}$, we use the test proposed by Geweke (2004) as follows:

Step 1. Draw $k_{1:T}$ from the prior $\pi(k_{1:T})$, which are gamma distributions.
Step 2. Draw the data from the normal distribution $\pi(Y_{1:T}|k_{1:T})$.
Step 3. Draw $k_{1:T}|Y_{1:T}$ using the posterior simulator

We repeat the 3 steps many times independently, so that we obtain many independent draws for $k_{1:T}$. If our algorithm is sampling from the true posterior, we expect by checking the simulation that the distribution of $k_{1:T}$ should be the same in step 1 and in step 3. We use the Z-test to test whether the mean for the time varying volatility $k_{1:T}$ drawn from the prior in step 1 (labelled as X) is significantly different from the one calculated using the posterior simulator in step 3 (labelled as Y). Using the US Macro data, and setting the fixed parameters equal to their posterior means, we obtain the following Z statistic which under the null hypothesis that the posterior simulator is correct verifies $Z \backsim N(0,1)$:

Table 11: Z test statistic for the mean

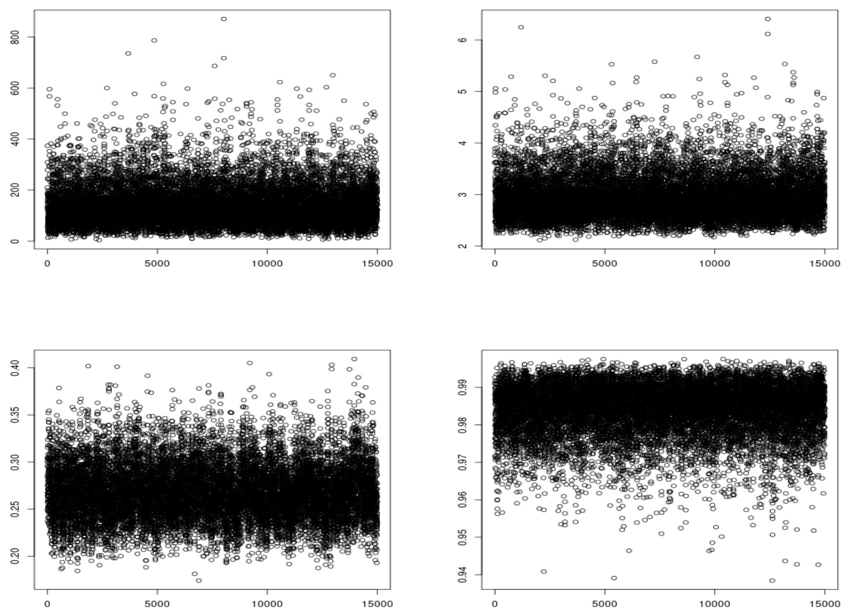
	X	**Y**
Standard Deviation	19.56542	19.91627
Mean	36.79686	36.65944
Mean Difference = 0.13742		
Number of Replications = 500		
Critical value (two tails) at 0.05 significance level = 1.96		
Z = 0.895		

The Z statistic is less than the critical value corresponding to the 0.05 significance level of 1.96, thus we fail to reject the null hypothesis that the posterior simulator is sampling from the true posterior distribution.

6.4 Trace Plots

Figure 5 shows 15000 iterations trace plots for $\tilde{\rho}$, $\tilde{n}$, the $(10, 10)$ element of Σ and the volatility at the middle of the sample[3] $K_{T/2}$, for the CSV-IG-AF-1 model with the US macro data. The conditional particle filter used 130 particles. The graphs show very good convergence and mixing properties. A similar performance was found for the exchange rate data with 150 particles.

Figure 5: Trace Plots for the CSV-IG-AF-1 Model



upper left: $K_{T/2}$, upper right: $\tilde{n}$, lower left: Σ, lower right: $\tilde{\rho}$

[3]Because the algorithm samples alternatively in both natural and reverse ordering, the middle of the sample is potentially the most sticky point.