Supplementary material: Scalable inference for a full multivariate stochastic volatility model

Petros Dellaportas[*], Michalis K. Titsias[†], Katerina Petrova[‡] and Anastasios Plataniotis[§]

1 MSV in TVP-VAR: MCMC specifics

The stochastic volatility approach of the current paper can easily be extended to a Gaussian TVP-VAR model

$$y_t = (I_M \otimes x_t)\beta_t + \varepsilon_t, \quad \varepsilon_t|\Omega_t \sim iid\mathcal{N}(0_M, \Omega_t)$$

where $\beta_t = vec\,(B_0, B_1, ..., B_p)'$ and $x_t = \left(1, y'_{t-1}, ..., y'_{t-p}\right)'$. In order to estimate the variation in the autoregressive parameters, we use the closed form conditional quasi-posteriors from **?**. Conditional on $\Omega_{1:T}$, the VAR model can be transformed as:

$$\begin{aligned} \widetilde{y}_t &= \widetilde{x}_t\beta_t + \eta_t, \quad \eta_t \sim iid\mathcal{N}(0_M, I_M), \\ \widetilde{y}_t &:= \Omega_t^{-1/2} y_t, \;\; \widetilde{x}_t := \Omega_t^{-1/2}(I_M \otimes x_t). \end{aligned} \tag{1}$$

[*]Corresponding author. Department of Statistical Science, University College, Gower Street, London WC1E 6BT, UK and Department of Statistics, Athens University of Economics and Business, Patission 76, Athens 10434, Greece. email: p.dellaportas@ucl.ac.uk

[†]Department of Informatics, Athens University of Economics and Business, Greece

[‡]Barcelona graduate school of Economics, Pompeu Fabra University, Barcelona, Spain

[§]Ernst and Young, Athens, Greece

Assuming a $\mathcal{N}\left(\beta_{0t}, V_{0t}^{-1}\right)$ prior distribution for β_t, $t \in \{1, ..., T\}$, Petrova (2019) shows that the conditional quasi-posterior distribution of β_t is given in close form by $\mathcal{N}\left(\widetilde{\beta}_t, \widetilde{V}_t^{-1}\right)$ with posterior mean and variance

$$\widetilde{\beta}_t = \widetilde{V}_t^{-1}\left[\sum\nolimits_{j=1}^{T} \vartheta_{tj}\widetilde{x}_j'\widetilde{y}_j + V_{0t}\beta_{0t}\right], \quad \widetilde{V}_t = V_{0t} + \sum\nolimits_{j=1}^{T} \vartheta_{tj}\widetilde{x}_j'\widetilde{x}_j \tag{2}$$

for each $t \in \{1, ..., T\}$, where $\widetilde{x}_t$ and $\widetilde{y}_t$ are defined in (1), and ϑ_{tj} are kernel weights computed as

$$\vartheta_{tj} = \left[\sum_{j=1}^{T}\left(\frac{K\left(\left(t-j\right)/H\right)}{\sum_{j=1}^{T} K\left(\left(t-j\right)/H\right)}\right)^2\right]^{-1} \frac{K\left(\left(t-j\right)/H\right)}{\sum_{j=1}^{T} K\left(\left(t-j\right)/H\right)}$$

and $K\left(.\right)$ is a non-negative continuous kernel function with bandwidth parameter H. Conditional on a draw $\beta_{1:T}|\Omega_{1:T}$, $\Omega_{1:T}$ can be estimated using an MCMC draw from the algorithm of this paper on the residuals of the VAR: $\hat{\varepsilon}_t|\beta_t = y_t - (I_M \otimes x_t)\beta_t$.

1.1 Additional results

Figures 1,2,3 and 4 illustrate the estimated off-diagonal elements of the covariance matrix and the stimated pairwise correlations, for all orderings, for both constant and time-varying parameters, for MSV and Primiceri models.

2 Priors over θ_h, θ_δ and (B, V)

Recall that $\theta_h = \{(\phi_i^h, h_{i,0}, \sigma_i^h)_{i=1}^N\}$ and $\theta_\delta = \{(\phi_{ij}^\delta, \delta_{ij,0}, \sigma_{ij}^\delta)_{i<j}\}$. Each $(\sigma_i^h)^2$ and $(\sigma_i^\delta)^2$ is assigned an inverse Gamma prior IGa$((\sigma_i^h)^2|\alpha_0^\sigma, \beta_0^\sigma)$ where $\alpha_0^\sigma = 10$ and $\beta_0^\sigma = 0.1$ were used in the simulations. Each $h_{i,0}$ and $\delta_{ij,0}$ had an improper prior of the form $p(h_{i,0}) \propto 1$.

The prior over ϕ_i^h, $i = 1, \ldots, N$ (and similarly for ϕ_{ij}^δ) was constructed as follows. Based on the transformation

$$\hat{\phi}_i^h = \log\left(\frac{1+\phi_i^h}{1-\phi_i^h}\right),$$

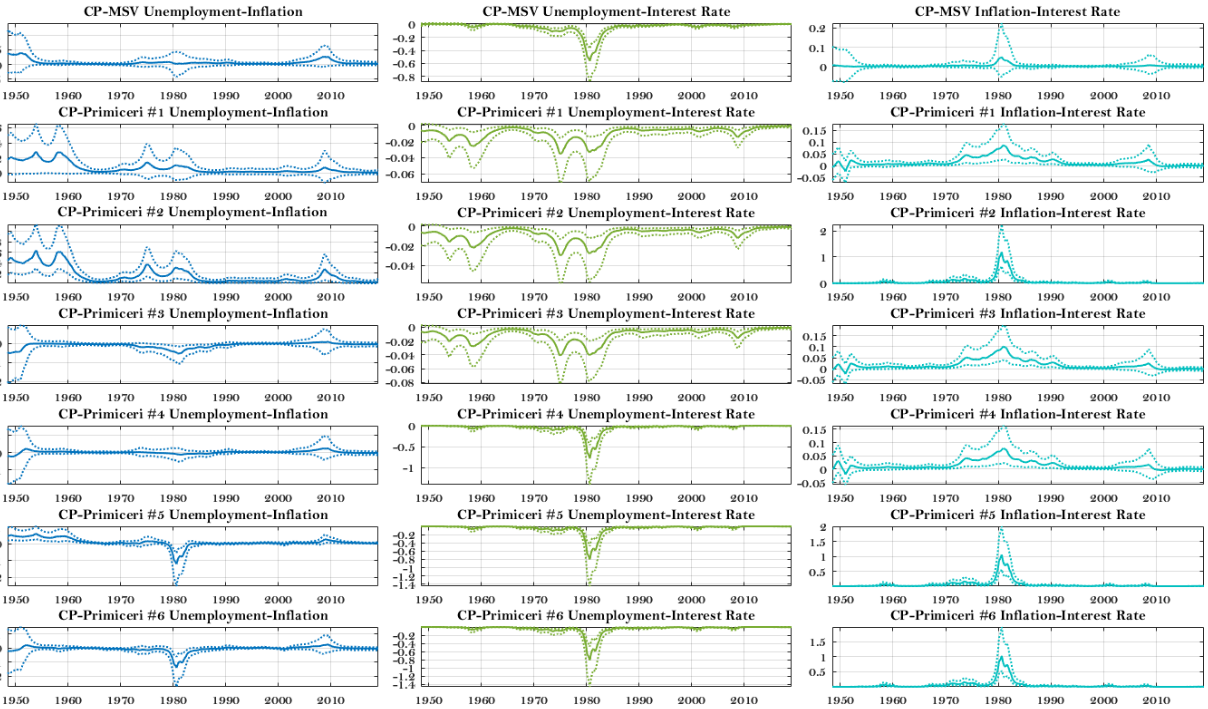

Figure 1: Estimated off-diagonal elements of the covariance matrix, all orderings; CP-MSV: constant parameter MSV model. CP-Primiceri: constant parameter Primiceri model.

we assign the prior

$$\begin{aligned} \hat{\phi}_i^h | \mu_h, \lambda_h &\sim \mathcal{N}(\mu_h, \lambda_h^{-1}), \\ \mu_h | \lambda_h &\sim \mathcal{N}(\mu_0, (k_0 \lambda_h)^{-1}), \\ \lambda_h &\sim \text{Ga}(\alpha_0, \beta_0). \end{aligned} \tag{3}$$

In other words, we have

$$\prod_{i=1}^{N} \mathcal{N}(\hat{\phi}_i^h | \mu_h, \lambda_h^{-1}) \mathcal{N}(\mu_h | \mu_0, (k_0 \lambda_h)^{-1}) \text{Ga}(\lambda_h | \alpha_0, \beta_0)$$

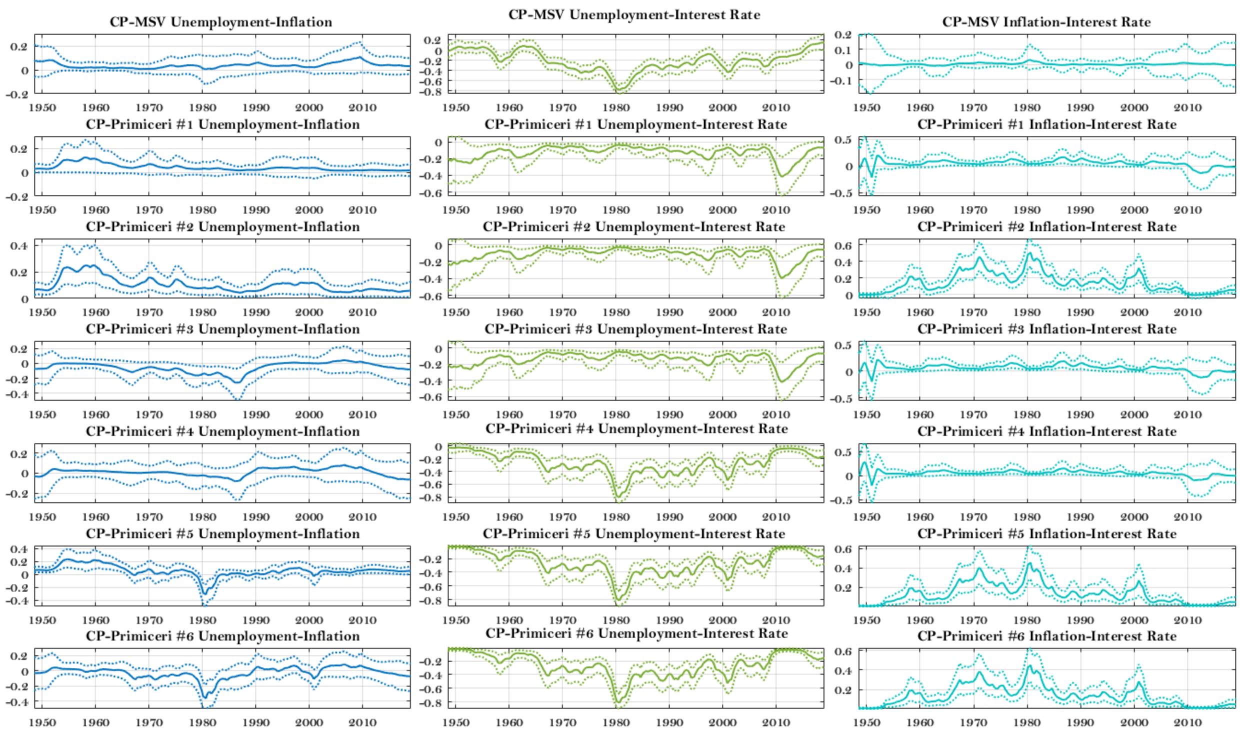

Figure 2: Estimated pairwise correlations, all orderings; CP-MSV: constant parameter MSV model. CP-Primiceri: constant parameter Primiceri model.

and by marginalizing out μ_h and λ_h we obtain

$$p(\hat{\phi}_1^h, \ldots, \hat{\phi}_N^h | \mu_0, k_0, \alpha_0, \beta_0) = \frac{\Gamma(\alpha_N)}{\Gamma(\alpha_0)} \frac{\beta_0^{\alpha_0}}{\beta_N^{\alpha_N}} \left(\frac{k_0}{k_N}\right)^{\frac{1}{2}} (2\pi)^{-\frac{N}{2}},$$

where

$$k_N = k_0 + N, \;\; \alpha_N = \alpha_0 + \frac{N}{2}, \;\; \beta_N = \beta_0 + \frac{1}{2}\sum_{i=1}^{N}(\hat{\phi}_i^h - \bar{\phi}_i^h)^2 + \frac{k_0 N(\bar{\phi}_i^h - \mu_0)^2}{2(k_0 + N)}.$$

In all simulations the hyperparameters took the fixed values $\mu_0 = 0, k_0 = 1, \alpha_0 = 1, \beta_0 = 1$.

Each element of the factor loadings matrix B is assigned an independent Gaussian prior with mean zero and variance equal to $\sigma_b^2 = 2$. Finally, the prior over the noise variances parameter v_i were given an inverse Gamma prior with both hyperparameters set to 0.001, i.e. $\alpha_0^v = \beta_0^v = 0.001$.

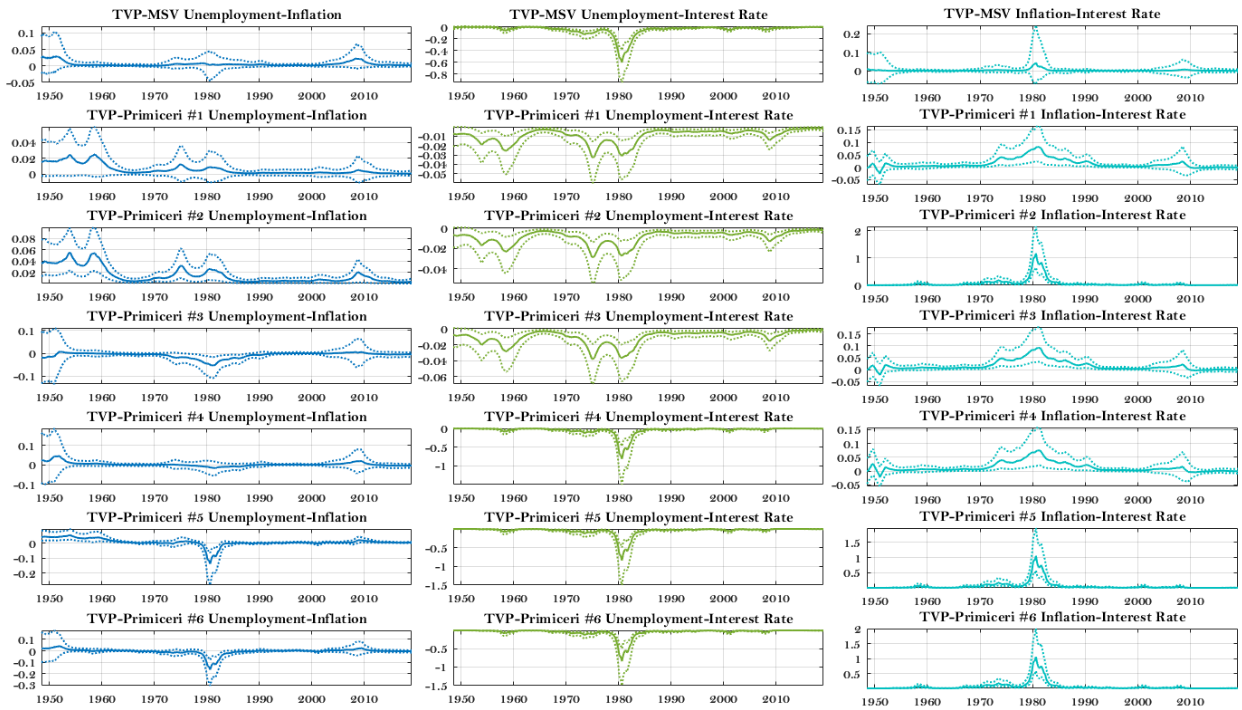

Figure 3: Estimated off-diagonal elements of the covariance matrix, all orderings; TVP-MSV: time-varying parameter MSV model. TVP-Primiceri: time-varying parameter Primiceri model.

3 Sampling moves for $(\theta_h, \theta_\delta)$ and (B, V)

We start by describing the Gibbs sampling steps for the parameters (B, V) in the factor MSV model. The conditional posterior distribution over the factor loading matrix B factorizes across rows so that for the ith row takes the form

$$\mathcal{N}(B_i|(\Phi\Phi^T + \frac{v_i}{\sigma_b^2}I)^{-1}\Phi R_i, v_i(\Phi\Phi^T + \frac{v_i}{\sigma_b^2}I)^{-1}),$$

where Φ is $K \times T$ matrix containing the f_ts as columns and R_i is T-dimensional vector containing all returns for the stock price i. The posterior conditional over the noise variance v_i takes the form of the following inverse Gamma

$$\text{IGa}(v_i|\alpha_0^v + T/2, \beta_0^v + \frac{1}{2}||R_i - \Phi^T B_i||^2).$$

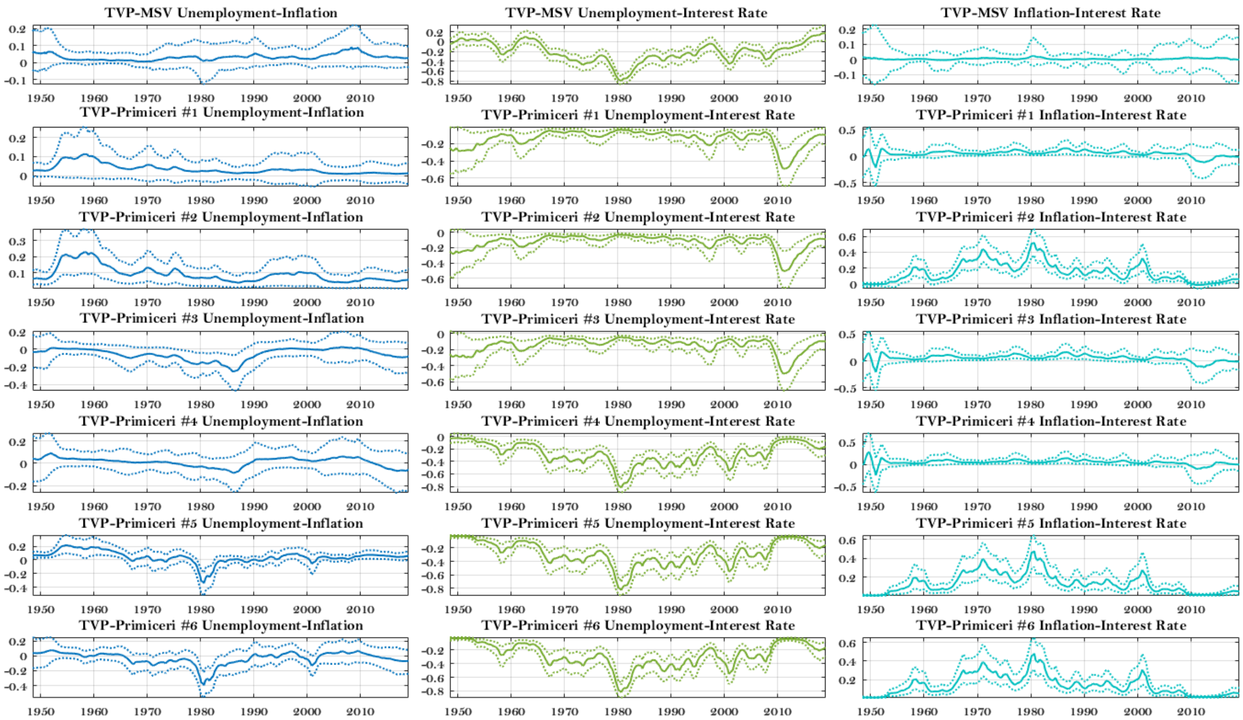

Figure 4: Estimated pairwise correlations, all orderings; TVP-MSV: time-varying parameter MSV model. TVP-Primiceri: time-varying parameter Primiceri model.

We now discuss the sampling moves for $\theta_h = \{(\phi_i^h, h_{i,0}, \sigma_i^h)_{i=1}^N\}$. The conditional posterior over $h_{i,0}$ is the Gaussian

$$\mathcal{N}(h_{i,0}|m_i, s_i^2)$$

where $m_i = \frac{1}{1-(\phi_i^h)^2+(T-1)(1-\phi_i^h)^2}((1-(\phi_i^h)^2)h_{i,1}+(1-\phi_i^h)\sum_{t=2}^T(h_{i,t+1}-\phi_i h_{i,t}))$, and $s_i^2 = \frac{(\sigma_i^h)^2}{1-(\phi_i^h)^2+(T-1)(1-\phi_i^h)^2}$. Therefore, Gibbs step for $h_{i,0}$ was based on simulating from this Gaussian. For $(\sigma_i^h)^2$ the sampling step is conjugate, where the conditional posterior is the following inverse Gamma

$$\mathrm{IGa}((\sigma_i^h)^2|\alpha_0^\sigma + T/2, \beta_0^\sigma + \frac{1}{2}(1-(\phi_i^h)^2)(h_{i,1}-h_{i,0})^2 + \frac{1}{2}\sum_{t=1}^{T-1}(h_{i,t+1}-h_{i,0}-\phi_i^h(h_{i,t}-h_{i,0}))^2).$$

All parameters $(\hat{\phi}_1^h, \ldots, \hat{\phi}_N^h)$ following the prior $p(\hat{\phi}_1^h, \ldots, \hat{\phi}_N^h|\mu_0, k_0, \alpha_0, \beta_0)$ are sampled jointly by using Metropolis-within Gibbs step based on a Gaussian proposal distribution with a spherical covariance matrix

δI and where δ was adapted during burn-in to achieve an acceptance rate around $20 - 30\%$.

Finally, the sampling moves for the parameters θ_δ were exactly analogous to the θ_h case.

4 Pseudo-code for the algorithm for computing the partial derivatives with respect to the rotation angles

Initialize $v_t = r_t$.
For $i = 1$ to $K - 1$
  For $j = i + 1$ to K
    Set $c = \cos(\omega_{ij,t}),\ s = \sin(\omega_{ij,t})$
    Set $t_1 = v_t[i],\ t_2 = v_t[j]$
    Set $v_t[i] \leftarrow c * t_1 - s * t_2$
    Set $v_t[j] \leftarrow s * t_1 + c * t_2$
    Set $\beta_1[i, j] \leftarrow -v_t[j]$
    Set $\beta_2[i, j] \leftarrow v_t[i]$
  End For
End For
$\alpha = v_t \circ \text{diag}(\Lambda_t^{-1})$
For $i = 1$ to $K - 1$
  For $j = i + 1$ to K
    Set $grad_t[i, j] \leftarrow -\beta_1[i, j] * \alpha[i] - \beta_2[i, j] * \alpha[j]$
    Set $c = \cos(\omega_{ij,t}),\ s = \sin(\omega_{ij,t})$
    Set $t_1 = \alpha[i],\ t_2 = \alpha[j]$
    Set $\alpha[i] \leftarrow c * t_1 + s * t_2$
    Set $\alpha[j] \leftarrow -s * t_1 + c * t_2$
  End For
End For

Algorithm 1: Recursive algorithm for computing the partial derivatives with respect to the rotation angles in $O(K^2)$ time. For memory savings all $\alpha_{ij,t}$ (for $i < j$) vectors are recursively computed and stored in the same vector α.