Vector autoregression models with skewness and heavy tails

Sune Karlsson[a], Stepan Mazur[a,b] and Hoang Nguyen[a]

[a]Unit of Statistics, School of Business, Örebro University, Sweden
[b]Department of Economics and Statistics, School of Business and Economics, Linnaeus University, Sweden

October 21, 2022

Online Appendix

A Bayesian estimation

This section summarizes the model setup and MCMC samplers for the models where the specification or the MCMC sampler is not discussed in detail in the paper.

A.1 Gaussian VAR with constant variance

The model is $\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{A}^{-1}\mathbf{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$ with $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$ and $\mathbf{\Sigma} = \mathrm{diag}(\tau_1^2, \ldots, \tau_k^2)$.

A.1.1 Prior distribution

The priors for $\mathbf{B}$ and $\mathbf{A}$ are as described in section 3.1. The prior for τ_i^2 is inverse Gamma, $\mathcal{IG}(1/2, 1/2)$.

A.1.2 MCMC

1. Sample $\mathbf{B}$ from the full conditional normal posterior.
2. Sample $\mathbf{A}$ from the full conditional normal posterior as in step 2 of Section 3.2, setting $\widetilde{\mathbf{u}}_t = \mathbf{y}_t - \mathbf{B}\mathbf{x}_t$ and $\mathbf{A}\widetilde{\mathbf{u}}_t = \mathbf{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$
3. Sample τ_i^2 from the full conditional inverse Gamma posterior, $\mathcal{IG}((T+1)/2, (\sum_{t=1}^T \tilde{\tilde{u}}_{it}^2 + 1)/2)$ for $\widetilde{\widetilde{\mathbf{u}}}_t = \mathbf{A}\widetilde{\mathbf{u}}_t$

A.2 Gaussian VAR with SV

The model is $\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{A}^{-1}\mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t$ with $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$ with $\mathbf{H}_t = \text{diag}(h_{1t}, \ldots, h_{kt})$ and the log volatilities evolving as

$$\log h_{it} = \log h_{it-1} + \sigma_i \eta_{it}, \quad i = 1, \ldots, k,$$

where $\eta_{it} \sim \mathcal{N}(0, 1)$.

A.2.1 Prior distribution

The priors for $\mathbf{B}$, $\mathbf{A}$, $\boldsymbol{\sigma}^2$ and $\mathbf{h}_0$ are as described in section 3.1.

A.2.2 MCMC

1. Sample $\mathbf{B}$ from the full conditional normal posterior.
2. Sample $\mathbf{A}$ from the full conditional normal posterior as in step 2 of Section 3.2, setting $\widetilde{\mathbf{u}}_t = \mathbf{y}_t - \mathbf{B}\mathbf{x}_t$ and $\mathbf{A}\widetilde{\mathbf{u}}_t = \mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t$
3. Sample $\mathbf{h}_{0:T}$ as in step 3 of Section 3.2
4. Sample $\boldsymbol{\sigma}^2$ as in step 4 of Section 3.2

A.3 VAR with multivariate $t$distribution and constant variance

The model is $\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \sqrt{\xi_t}\mathbf{A}^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$ with $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$, $\boldsymbol{\Sigma} = \text{diag}(\tau_1^2, \ldots, \tau_k^2)$ and $\xi_t \sim \mathcal{IG}(\nu/2, \nu/2)$.

A.3.1 Prior distribution

The priors for $\mathbf{B}$ and $\mathbf{A}$ are as described in section 3.1. The prior for τ_i^2 is inverse Gamma, $\mathcal{IG}(1/2, 1/2)$ and the prior for the single ν is $\mathcal{G}(2, 0.1)$ as in section 3.1.

A.3.2 MCMC

1. Sample $\mathbf{B}$ from the full conditional normal posterior.
2. Sample $\mathbf{A}$ from the full conditional normal posterior as in step 2 of section 3.2, setting $\widetilde{\mathbf{u}}_t = (\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)/\sqrt{\xi_t}$ and $\mathbf{A}\widetilde{\mathbf{u}}_t = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$
3. Sample τ_i^2 from the full conditional inverse Gamma posterior, $\mathcal{IG}((T+1)/2, (\sum_{t=1}^T \tilde{\tilde{u}}_{it}^2 + 1)/2)$ for $\widetilde{\widetilde{\mathbf{u}}}_t = \mathbf{A}\widetilde{\mathbf{u}}_t$
4. The full conditional posterior for ν is proportional to $\mathcal{G}(2, 0.1) \prod_{t=1}^T \mathcal{IG}(\xi_t, \nu/2, \nu/2)$ and is sampled as in step 5 of section 3.2
5. Sample ξ_t from the full conditional inverse Gamma posterior, $\mathcal{IG}((\nu+k)/2, (\nu+\sum_{i=1}^k u_{it}^2/\tau_i^2)/2)$ for $\mathbf{u}_t = \mathbf{A}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)$

A.4 VAR with multivariate $t$distribution and SV

The model is $\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \sqrt{\xi_t}\mathbf{A}^{-1}\mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t$ with $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$, $\xi_t \sim \mathcal{IG}(\nu/2, \nu/2)$, $\mathbf{H}_t = \mathrm{diag}(h_{1t}, \ldots, h_{kt})$ and the log volatilities evolving as

$$\log h_{it} = \log h_{it-1} + \sigma_i \eta_{it}, \quad i = 1, \ldots, k,$$

where $\eta_{it} \sim \mathcal{N}(0, 1)$.

A.4.1 Prior distribution

The priors for $\mathbf{B}$, $\mathbf{A}$, $\mathbf{h}_0$ and $\boldsymbol{\sigma}^2$ are as described in section 3.1 and the prior for the single ν is $\mathcal{G}(2, 0.1)$ as in section 3.1.

A.4.2 MCMC

1. Sample $\mathbf{B}$ from the full conditional normal posterior.
2. Sample $\mathbf{A}$ from the full conditional normal posterior as in step 2 of section 3.2, setting $\widetilde{\mathbf{u}}_t = (\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)/\sqrt{\xi_t}$ and $\mathbf{A}\widetilde{\mathbf{u}}_t = \mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t$
3. Sample $\mathbf{h}_{0:T}$ as in step 3 of Section 3.2
4. Sample $\boldsymbol{\sigma}^2$ as in step 4 of Section 3.2
5. The full conditional posterior for ν is proportional to $\mathcal{G}(2, 0.1) \prod_{t=1}^{T} \mathcal{IG}(\xi_t, \nu/2, \nu/2)$ and is sampled as in step 5 of section 3.2
6. Sample ξ_t from the full conditional inverse Gamma posterior, $\mathcal{IG}((\nu+k)/2, (\nu+\sum_{i=1}^{k} u_{it}^2/h_{it})/2)$ for $\mathbf{u}_t = \mathbf{A}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)$

A.5 VAR with multivariate skew-t distribution and constant variance

The model is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + (\xi_t - \mu_\xi)\boldsymbol{\gamma} + \xi_t^{1/2}\mathbf{A}^{-1}\boldsymbol{\Sigma}^{1/2}\epsilon_t$$

with $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$, $\boldsymbol{\Sigma} = \mathrm{diag}(\tau_1^2, \ldots, \tau_k^2)$ and $\xi_t \sim \mathcal{IG}(\nu/2, \nu/2)$

A.5.1 Prior distribution

The priors for $\mathbf{B}$, $\boldsymbol{\gamma}$ and $\mathbf{A}$ are as described in section 3.1. The prior for τ_i^2 is inverse Gamma, $\mathcal{IG}(1/2, 1/2)$ and the prior for the single ν is $\mathcal{G}(2, 0.1)$ as in section 3.1.

A.5.2 MCMC

1. Sample $\mathbf{B}$ and $\boldsymbol{\gamma}$ jointly as $\boldsymbol{b} = (vec(\mathbf{B})', \boldsymbol{\gamma}')'$. Rewrite the model as
$$\begin{aligned} \xi_t^{-1/2}\mathbf{A}\mathbf{y}_t &= \xi_t^{-1/2}\mathbf{A}\mathbf{B}\mathbf{x}_t + \xi_t^{-1/2}(\xi_t - \mu_\xi)\mathbf{A}\boldsymbol{\gamma} + \boldsymbol{\Sigma}^{1/2}\epsilon_t \\ &= \left(\xi_t^{-1/2}\mathbf{x}_t' \otimes \mathbf{A} \quad \xi_t^{-1/2}(\xi_t - \mu_\xi)\mathbf{A}\right)\begin{pmatrix} vec(\mathbf{B}) \\ \gamma \end{pmatrix} + \boldsymbol{\Sigma}^{1/2}\epsilon_t \\ \widetilde{\mathbf{y}}_t &= \widetilde{\mathbf{X}}_t\boldsymbol{b} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t. \end{aligned} \tag{1}$$
a multivariate regression model with error variance $\boldsymbol{\Sigma}$ where $\widetilde{\mathbf{y}}_t = \xi_t^{-1/2}\mathbf{A}\mathbf{y}_t$ and $\widetilde{\mathbf{X}}_t = \left(\xi_t^{-1/2}\mathbf{x}_t' \otimes \mathbf{A} \quad \xi_t^{-1/2}(\xi_t - \mu_\xi)\mathbf{A}\right)$. $\boldsymbol{b}$ is then sampled from the full conjugate normal posterior.

2. Sample $\mathbf{A}$ from the full conditional normal posterior as in step 2 of section 3.2, setting $\widetilde{\mathbf{u}}_t = \xi_t^{-1/2}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t - \xi_t^{-1/2}(\xi_t - \mu_\xi)\boldsymbol{\gamma})$ and $\mathbf{A}\widetilde{\mathbf{u}}_t = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$

3. Sample τ_i^2 from the full conditional inverse Gamma posterior, $\mathcal{IG}((T+1)/2, (\sum_{t=1}^T \tilde{\tilde{u}}_{it}^2 + 1)/2)$ for $\widetilde{\widetilde{\mathbf{u}}}_t = \mathbf{A}\widetilde{\mathbf{u}}_t$

4. The full conditional posterior for ν is proportional to $\mathcal{G}(2, 0.1)\prod_{t=1}^T \mathcal{IG}(\xi_t, \nu/2, \nu/2)$ and is sampled as in step 5 of section 3.2

5. To sample from the full conditional posterior for ξ_t, write
$$\mathbf{u}_t = \mathbf{y}_t - \mathbf{B}\mathbf{x}_t + \mu_\xi\gamma = \xi_t\gamma + \xi_t^{1/2}\mathbf{A}^{-1}\boldsymbol{\Sigma}^{1/2}\epsilon_t.$$
$\mathbf{u}_t$ is thus conditionally normal, $\mathbf{u}_t \sim \mathcal{N}(\xi_t\boldsymbol{\gamma}, \xi_t\mathbf{A}^{-1}\mathbf{H}_t\mathbf{A}^{-1\prime})$ and the likelihood contribution is
$$\begin{aligned} &\xi_t^{-k/2}\exp\left\{-\frac{1}{2\xi_t}(\mathbf{u}_t - \xi_t\boldsymbol{\gamma})'\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{u}_t - \xi_t\boldsymbol{\gamma})\right\} \\ &\propto \xi_t^{-k/2}\exp\left\{\frac{1}{2}\left(\frac{\mathbf{u}_t'\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A}\mathbf{u}_t}{\xi_t} + \xi_t\boldsymbol{\gamma}'\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A}\boldsymbol{\gamma}\right)\right\}. \end{aligned} \tag{2}$$
With the independent inverse Gamma distribution for ξ_t we have conditional independence and the full conditional posterior is generalized inverse Gaussian (GIG)
$$\begin{aligned} \pi(\xi_t|\boldsymbol{\Psi}) &\propto \xi_t^{-k/2}\exp\left\{-\frac{1}{2}\left(\frac{q_t^2}{\xi_t} + \xi_t p_t^2\right)\right\}\xi_t^{-(\nu/2+1)}\exp\left\{-\frac{\nu}{2\xi_t}\right\} \\ &= \xi_t^{\lambda-1}\exp\left\{-\frac{1}{2}\left(\frac{\chi}{\xi_t} + \psi\xi_t\right)\right\} \end{aligned}$$
where $q_t^2 = \mathbf{u}_t'\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A}\mathbf{u}_t$ and $p_t^2 = \boldsymbol{\gamma}'\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A}\boldsymbol{\gamma}$. The conditional distribution of $\pi(\xi_t|\boldsymbol{\Psi})$ is $GIG(\lambda, \psi, \chi)$ with $\lambda = -(\nu + k)/2$, $\chi = q_t^2 + \nu$ and $\psi = p_t^2$.

A.6 VAR with multivariate skew-t distribution and SV

The model is
$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + (\xi_t - \mu_\xi)\gamma + \xi_t^{1/2}\mathbf{A}^{-1}\mathbf{H}_t^{1/2}\epsilon_t$$
with $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$, $\xi_t \sim \mathcal{IG}(\nu/2, \nu/2)$, $\mathbf{H}_t = \text{diag}(h_{1t}, \ldots, h_{kt})$ and the log volatilities evolving as
$$\log h_{it} = \log h_{it-1} + \sigma_i\eta_{it}, \quad i = 1, \ldots, k,$$
where $\eta_{it} \sim \mathcal{N}(0, 1)$.

A.6.1 Prior distribution

The priors for $\mathbf{B}$, $\boldsymbol{\gamma}$, $\mathbf{A}$, $\mathbf{h}_0$ and $\boldsymbol{\sigma}^2$ are as described in section 3.1 and the prior for the single ν is $\mathcal{G}(2, 0.1)$ as in section 3.1.

A.6.2 MCMC

1. Sample Sample $\mathbf{B}$ and $\boldsymbol{\gamma}$ jointly from the full conditional normal posterior. To this end, rewrite the model as in (1) to yield $\widetilde{\mathbf{y}}_t = \widetilde{\mathbf{X}}_t \boldsymbol{b} + \mathbf{H}_t^{1/2} \boldsymbol{\epsilon}_t$, a heteroskedastic multivariate regression.
2. Sample $\mathbf{A}$ from the full conditional normal posterior as in step 2 of section 3.2, setting $\widetilde{\mathbf{u}}_t = \xi_t^{-1/2}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t - \xi_t^{-1/2}(\xi_t - \mu_\xi)\boldsymbol{\gamma})$ and $\mathbf{A}\widetilde{\mathbf{u}}_t = \mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t$
3. Sample $\mathbf{h}_{0:T}$ as in step 3 of Section 3.2
4. Sample $\boldsymbol{\sigma}^2$ as in step 4 of Section 3.2
5. The full conditional posterior for ν is proportional to $\mathcal{G}(2, 0.1) \prod_{t=1}^{T} \mathcal{IG}(\xi_t, \nu/2, \nu/2)$ and is sampled as in step 5 of section 3.2

   (d) Sample from $\pi(\xi_t|\boldsymbol{\Psi})$ for $t = 1, \ldots, T$
6. To sample ξ_t from the full conditional posterior, write

$$\mathbf{u}_t = \mathbf{y}_t - \mathbf{B}\mathbf{x}_t + \mu_\xi \gamma = \xi_t \gamma + \xi_t^{1/2} \mathbf{A}^{-1} \mathbf{H}_t^{1/2} \epsilon_t.$$

   and replace $\boldsymbol{\Sigma}$ with $\mathbf{H}_t$ in (2). The full conditional posterior is thus generalized inverse Gaussian, $GIG(\lambda, \psi, \chi)$ with $\lambda = -(\nu + k)/2$, $\chi = q_t^2 + \nu$ and $\psi = p_t^2$ for $q_t^2 = \mathbf{u}_t' \mathbf{A}' \mathbf{H}_t^{-1} \mathbf{A} \mathbf{u}_t$ and $p_t^2 = \boldsymbol{\gamma}' \mathbf{A}' \mathbf{H}_t^{-1} \mathbf{A} \boldsymbol{\gamma}$.

A.7 VAR with MT distribution and constant variance

The model is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{W}_t^{1/2} \mathbf{A}^{-1} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\epsilon}_t$$

with $\mathbf{W}_t = \mathrm{diag}(\xi_{1t}, \ldots, \xi_{kt})$, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$, $\boldsymbol{\Sigma} = \mathrm{diag}(\tau_1^2, \ldots, \tau_k^2)$ and $\xi_{it} \sim \mathcal{IG}(\nu_i/2, \nu_i/2)$

A.7.1 Prior

The priors for $\mathbf{B}$, ν_i and $\mathbf{A}$ are as described in section 3.1 and the prior for τ_i^2 is inverse Gamma, $\mathcal{IG}(1/2, 1/2)$.

A.7.2 MCMC

1. Sample $\mathbf{B}$ from the full conditional normal posterior.
2. Sample $\mathbf{A}$ from the full conditional normal posterior as in step 2 of section 3.2, setting $\widetilde{\mathbf{u}}_t = \mathbf{W}_t^{-1/2}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)$ and $\mathbf{A}\widetilde{\mathbf{u}}_t = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$
3. Sample τ_i^2 from the full conditional inverse Gamma posterior, $\mathcal{IG}((T+1)/2, (\sum_{t=1}^{T} \tilde{\tilde{u}}_{it}^2 + 1)/2)$ for $\widetilde{\widetilde{\mathbf{u}}}_t = \mathbf{A}\widetilde{\mathbf{u}}_t$

4. The full conditional posterior for ν_i is proportional to $\mathcal{G}(2, 0.1) \prod_{t=1}^{T} \mathcal{IG}(\xi_{it}, \nu_i/2, \nu_1/2)$ and is sampled as in step 5 of section 3.2

5. Sample ξ_{it} as in step 6 of Section 3.2, with

$$\pi(\mathbf{W}_t|\mathbf{\Psi}) \propto \prod_{i=1}^{k} \xi_{it}^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)' \mathbf{\Omega}_t^{-1}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)\right) \mathcal{IG}\left(\xi_{it}; \frac{\nu_i}{2}, \frac{\nu_i}{2}\right).$$

where $\mathbf{\Omega}_t = \mathbf{W}_t^{1/2}\mathbf{A}^{-1}\mathbf{\Sigma}\mathbf{A}^{-1\prime}\mathbf{W}_t^{1/2}$.

A.8 VAR with MT distribution and SV

The model is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{W}_t^{1/2}\mathbf{A}^{-1}\mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t$$

with $\mathbf{W}_t = \text{diag}(\xi_{1t}, \ldots, \xi_{kt})$, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$, $\xi_{it} \sim \mathcal{IG}(\nu_i/2, \nu_i/2)$ $\mathbf{H}_t = \text{diag}(h_{1t}, \ldots, h_{kt})$ and the log volatilities evolving as

$$\log h_{it} = \log h_{it-1} + \sigma_i \eta_{it}, \quad i = 1, \ldots, k,$$

where $\eta_{it} \sim \mathcal{N}(0, 1)$.

A.8.1 Prior

The priors for $\mathbf{B}$, ν_i, $\mathbf{A}$ and $\mathbf{h}_0$ are as described in section 3.1.

A.8.2 MCMC

1. Sample $\mathbf{B}$ from the full conditional normal posterior.

2. Sample $\mathbf{A}$ from the full conditional normal posterior as in step 2 of section 3.2, setting $\widetilde{\mathbf{u}}_t = \mathbf{W}_t^{-1/2}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)$ and $\mathbf{A}\widetilde{\mathbf{u}}_t = \mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t$

3. Sample $\mathbf{h}_{0:T}$ as in step 3 of Section 3.2

4. Sample $\boldsymbol{\sigma}^2$ as in step 4 of Section 3.2

5. The full conditional posterior for ν_i is proportional to $\mathcal{G}(2, 0.1) \prod_{t=1}^{T} \mathcal{IG}(\xi_{it}, \nu_i/2, \nu_1/2)$ and is sampled as in step 5 of section 3.2

6. Sample ξ_{it} as in step 6 of Section 3.2, with

$$\pi(\mathbf{W}_t|\mathbf{\Psi}) \propto \prod_{i=1}^{k} \xi_{it}^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)' \mathbf{\Omega}_t^{-1}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)\right) \mathcal{IG}\left(\xi_{it}; \frac{\nu_i}{2}, \frac{\nu_i}{2}\right).$$

where $\mathbf{\Omega}_t = \mathbf{W}_t^{1/2}\mathbf{A}^{-1}\mathbf{H}_t\mathbf{A}^{-1\prime}\mathbf{W}_t^{1/2}$.

A.9 VAR with MST distribution and constant variance

The model is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + (\mathbf{W}_t - \overline{\mathbf{W}})\boldsymbol{\gamma} + \mathbf{W}_t^{1/2}\mathbf{A}^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$$

with $\mathbf{W}_t = \mathrm{diag}(\xi_{1t}, \ldots, \xi_{kt})$, $\overline{\mathbf{W}} = \mathrm{diag}(\mu_{\xi,1} \ldots, \mu_{\xi,k})$, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$, $\boldsymbol{\Sigma} = \mathrm{diag}(\tau_1^2, \ldots, \tau_k^2)$ and $\xi_{it} \sim \mathcal{IG}(\nu_i/2, \nu_i/2)$

A.9.1 Prior

The priors for $\mathbf{B}$, $\boldsymbol{\gamma}$, ν_i and $\mathbf{A}$ are as described in section 3.1 and the prior for τ_i^2 is inverse Gamma, $\mathcal{IG}(1/2, 1/2)$.

A.9.2 MCMC

1. Sample $\mathbf{B}$ and $\boldsymbol{\gamma}$ jointly from the full conditional normal posterior as in step 1 of Section 3.2, replacing $\mathbf{H}_t$ with $\boldsymbol{\Sigma}$.
2. Sample $\mathbf{A}$ as in step 2 of Section 3.2, replacing $\mathbf{H}_t$ with $\boldsymbol{\Sigma}$.
3. Sample τ_i^2 from the full conditional inverse Gamma posterior, $\mathcal{IG}((T+1)/2, (\sum_{t=1}^T \tilde{\tilde{u}}_{it}^2 + 1)/2)$ for $\tilde{\tilde{\mathbf{u}}}_t = \mathbf{A}\tilde{\mathbf{u}}_t$ and $\tilde{\mathbf{u}}_t = \mathbf{W}_t^{-1/2}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t - (\mathbf{W}_t - \overline{\mathbf{W}})\gamma)$
4. The full conditional posterior for ν_i is proportional to $\mathcal{G}(2, 0.1) \prod_{t=1}^T \mathcal{IG}(\xi_{it}, \nu_i/2, \nu_1/2)$ and is sampled as in step 5 of section 3.2
5. Sample ξ_{it} as in step 6 of Section 3.2, replacing $\mathbf{H}_t$ with $\boldsymbol{\Sigma}$ in the expression for $\boldsymbol{\Omega}_t$

A.10 VAR with OT distribution and constant variance

The model is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{A}^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$$

with $\mathbf{W}_t = \mathrm{diag}(\xi_{1t}, \ldots, \xi_{kt})$, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$, $\boldsymbol{\Sigma} = \mathrm{diag}(\tau_1^2, \ldots, \tau_k^2)$ and $\xi_{it} \sim \mathcal{IG}(\nu_i/2, \nu_i/2)$

A.10.1 Prior

The priors for $\mathbf{B}$, ν_i and $\mathbf{A}$ are as described in section 3.1 and the prior for τ_i^2 is inverse Gamma, $\mathcal{IG}(1/2, 1/2)$.

A.10.2 MCMC

1. Sample $\mathbf{B}$ from the full conditional normal posterior.
2. Let $\mathbf{u}_t = \mathbf{y}_t - \mathbf{B}\mathbf{x}_t$ we then have a triangular equation system

$$\mathbf{A}\mathbf{u}_t = \mathbf{W}_t^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$$

   with $\xi_{it}\tau_i^2$ for $i = 2, \ldots, k$. Rows $2, \ldots, k$ in $\mathbf{A}$ are then sampled from the full conditional posteriors as in step 2 of Section 3.2.

3. Sample τ_i^2 from the full conditional inverse Gamma posterior, $\mathcal{IG}((T+1)/2, (\sum_{t=1}^{T} \tilde{\tilde{u}}_{it}^2/\xi_{it} + 1)/2)$ for $\tilde{\tilde{\mathbf{u}}}_t = \mathbf{A}\mathbf{u}_t$

4. The full conditional posterior for ν_i is proportional to $\mathcal{G}(2, 0.1) \prod_{t=1}^{T} \mathcal{IG}(\xi_{it}, \nu_i/2, \nu_i/2)$ and is sampled as in step 4 of section 3.2

5. Sample ξ_{it} from the full conditional inverse Gamma distribution, $\mathcal{IG}((\nu_i + \tilde{\tilde{u}}_{it}^2/\tau_i)/2, (\nu_i + 1)/2)$ for $\tilde{\tilde{\mathbf{u}}}_t = \mathbf{A}\mathbf{u}_t$

A.11 VAR with OT distribution and SV

The model is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{A}^{-1}\mathbf{W}_t^{1/2}\mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t.$$

with $\mathbf{W}_t = \operatorname{diag}(\xi_{1t}, \ldots, \xi_{kt})$, $\xi_{it} \sim \mathcal{IG}(\nu_i/2, \nu_i/2)$, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$, $\mathbf{H}_t = \operatorname{diag}(h_{1t}, \ldots, h_{kt})$ and the log volatilities evolving as

$$\log h_{it} = \log h_{it-1} + \sigma_i \eta_{it}, \quad i = 1, \ldots, k,$$

where $\eta_{it} \sim \mathcal{N}(0, 1)$.

A.11.1 Prior distribution

The priors for $\mathbf{B}$, $\boldsymbol{\gamma}$, ν_i, $\mathbf{A}$, $\mathbf{h}_0$ and $\boldsymbol{\sigma}^2$ are as described in section 3.1.

A.11.2 MCMC

1. Sample $\mathbf{B}$ from the full conditional normal posterior.

2. Sample $\mathbf{A}$ as in step 2 of Section A.10.2 with $\xi_{it} h_{it}$ as the known error variance of equations $i = 2, \ldots, k$ of the triangular equation system.

3. Sample $\mathbf{h}_{0:T}$ as in step 3 of Section 3.2 with $\tilde{\tilde{\mathbf{u}}}_t = \mathbf{W}_t^{-1/2}\mathbf{A}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t)$

4. Sample $\boldsymbol{\sigma}^2$ as in step 4 of Section 3.2

5. The full conditional posterior for ν_i is proportional to $\mathcal{G}(2, 0.1) \prod_{t=1}^{T} \mathcal{IG}(\xi_{it}, \nu_i/2, \nu_1/2)$ and is sampled as in step 5 of section 3.2

6. Sample ξ_{it} from the full conditional inverse Gamma distribution, $\mathcal{IG}((\nu_i + \tilde{\tilde{u}}_{it}^2/h_{it})/2, (\nu_i + 1)/2)$ for $\tilde{\tilde{\mathbf{u}}}_t = \mathbf{A}\mathbf{u}_t$

A.12 VAR with OST distribution and constant variance

The model is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{A}^{-1}\left((\mathbf{W}_t - \overline{\mathbf{W}})\boldsymbol{\gamma} + \mathbf{W}_t^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t\right).$$

with $\mathbf{W}_t = \operatorname{diag}(\xi_{1t}, \ldots, \xi_{kt})$, $\overline{\mathbf{W}} = \operatorname{diag}(\mu_{\xi,1} \ldots, \mu_{\xi,k})$, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$, $\boldsymbol{\Sigma} = \operatorname{diag}(\tau_1^2, \ldots, \tau_k^2)$ and $\xi_{it} \sim \mathcal{IG}(\nu_i/2, \nu_i/2)$

A.12.1 Prior

The priors for $\mathbf{B}$, $\boldsymbol{\gamma}$, ν_i and $\mathbf{A}$ are as described in section 3.1 and the prior for τ_i^2 is inverse Gamma, $\mathcal{IG}(1/2, 1/2)$.

A.12.2 MCMC

1. Sample $\mathbf{B}$ and $\boldsymbol{\gamma}$ jointly as $\boldsymbol{b} = (\text{vec}(\mathbf{B})', \boldsymbol{\gamma}')'$. Rewrite the model as

$$\begin{aligned}\mathbf{A}\mathbf{y}_t =& \mathbf{A}\mathbf{B}\mathbf{x}_t + (\mathbf{W}_t - \overline{\mathbf{W}})\gamma + \mathbf{W}_t^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t \\ =& \begin{pmatrix}\mathbf{x}_t' \otimes \mathbf{A} & (\mathbf{W}_t - \overline{\mathbf{W}})\end{pmatrix}\begin{pmatrix}\text{vec}(\mathbf{B}) \\ \gamma\end{pmatrix} + \mathbf{W}_t^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t \\ \widetilde{\mathbf{y}}_t =& \widetilde{\mathbf{X}}_t\boldsymbol{b} + \mathbf{W}_t^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t.\end{aligned} \tag{3}$$

a heteroskedastic multivariate regression model with error variance-covariance $\text{diag}(\xi_{1t}\tau_1^2, \ldots, \xi_{kt}\tau_k^2)$ where $\widetilde{\mathbf{y}}_t = \mathbf{A}\mathbf{y}_t$ and $\widetilde{\mathbf{X}}_t = \begin{pmatrix}\mathbf{x}_t' \otimes \mathbf{A} & (\mathbf{W}_t - \overline{\mathbf{W}})\end{pmatrix}$. $\boldsymbol{b}$ is then sampled from the full conditional normal posterior.

2. Let $\mathbf{u}_t = \mathbf{y}_t - \mathbf{B}\mathbf{x}_t$ we then have a triangular equation system

$$\mathbf{A}\mathbf{u}_t = (\mathbf{W}_t - \overline{\mathbf{W}})\gamma + \mathbf{W}_t^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$$

with $\tilde{u}_{it} = u_{it} - (\xi_{it} - \mu_{\xi,i})\gamma_i$ as dependent variable, $(-u_{1,t}, \ldots, -u_{i-1,t})$ as explanatory variables and know variance $\xi_{it}\tau_i^2$ for $i = 2, \ldots, k$. Rows $2, \ldots, k$ in $\mathbf{A}$ are then sampled from the full conditional posteriors.

3. Sample τ_i^2 from the full conditional inverse Gamma posterior, $\mathcal{IG}((T+1)/2, (\sum_{t=1}^T \tilde{\tilde{u}}_{it}^2/\xi_{it} + 1)/2)$ for $\widetilde{\widetilde{\mathbf{u}}}_t = \mathbf{A}\mathbf{u}_t - (\mathbf{W}_t - \overline{\mathbf{W}})\boldsymbol{\gamma}$

4. The full conditional posterior for ν_i is proportional to $\mathcal{G}(2, 0.1)\prod_{t=1}^T \mathcal{IG}(\xi_{it}, \nu_i/2, \nu_i/2)$ and is sampled as in step 4 of section 3.2

5. To sample from the full conditional distribution for ξ_{it} write

$$\mathbf{u}_t = \mathbf{A}\left(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t\right) + \overline{\mathbf{W}}\boldsymbol{\gamma} = \mathbf{W}_t\gamma + \mathbf{W}_t^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t.$$

That is, the conditional distribution of $\mathbf{u}_t$ is normal, $\mathbf{u}_t \sim \mathcal{N}(\mathbf{W}_t\boldsymbol{\gamma}, \mathbf{W}_t\boldsymbol{\Sigma})$ and the likelihood contribution is

$$|\mathbf{W}_t| \exp\left(-\frac{1}{2}(\mathbf{u}_t - \mathbf{W}_t\gamma)'\boldsymbol{\Sigma}^{-1}\mathbf{W}_t^{-1}(\mathbf{u}_t - \mathbf{W}_t\gamma)\right) \tag{4}$$

with

$$(\mathbf{u}_t - \mathbf{W}_t\gamma)'\boldsymbol{\Sigma}^{-1}\mathbf{W}_t^{-1}(\mathbf{u}_t - \mathbf{W}_t\gamma) = \sum_{i=1}^k \frac{u_{it}^2/\tau_i^2}{\xi_{it}} - 2\sum_{i=1}^k \frac{u_{it}\gamma_i}{\tau_i^2} + \sum_{i=1}^k \frac{\gamma_i^2}{\tau_i^2}\xi_{it}.$$

With the independent inverse Gamma distribution for ξ_{it} we have conditional independence and the full conditional posterior is generalized inverse Gaussian (GIG),

$$\begin{aligned}\pi(\xi_{it}|\boldsymbol{\Psi}) &\propto \xi_{it}^{-1/2}\exp\left\{-\frac{1}{2}\left(\frac{q_{it}^2}{\xi_{it}} + \xi_{it}p_{it}^2\right)\right\}\xi_{it}^{-(\nu_i/2+1)}\exp\left\{-\frac{\nu_i}{2\xi_{it}}\right\} \\ &= \xi_{it}^{\lambda-1}\exp\left\{-\frac{1}{2}\left(\frac{\chi}{\xi_{it}} + \psi\xi_{it}\right)\right\}\end{aligned}$$

for $q_{it}^2 = u_{it}^2/\tau_i^2$, $p_{it}^2 = \gamma_i^2/\tau_i^2$, $\lambda = -(\nu_i+1)/2$, $\chi = q_{it}^2+\nu_i$ and $\psi = p_{it}^2$. That is a $GIG(\lambda, \psi, \chi)$ distribution.

A.13 VAR with OST distribution and SV

The model is

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \mathbf{A}^{-1}\left((\mathbf{W}_t - \overline{\mathbf{W}})\boldsymbol{\gamma} + \mathbf{W}_t^{1/2}\mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t\right).$$

with $\mathbf{W}_t = \mathrm{diag}(\xi_{1t}, \ldots, \xi_{kt})$, $\xi_{it} \sim \mathcal{IG}(\nu_i/2, \nu_i/2)$, $\overline{\mathbf{W}} = \mathrm{diag}(\mu_{\xi,1} \ldots, \mu_{\xi,k})$, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{I})$, $\mathbf{H}_t = \mathrm{diag}(h_{1t}, \ldots, h_{kt})$ and the log volatilities evolving as

$$\log h_{it} = \log h_{it-1} + \sigma_i \eta_{it}, \quad i = 1, \ldots, k,$$

where $\eta_{it} \sim \mathcal{N}(0, 1)$.

A.13.1 Prior distribution

The priors for $\mathbf{B}$, $\boldsymbol{\gamma}$, ν_i, $\mathbf{A}$, $\mathbf{h}_0$ and $\boldsymbol{\sigma}^2$ are as described in section 3.1.

A.13.2 MCMC

1. Sample Sample $\mathbf{B}$ and $\boldsymbol{\gamma}$ jointly from the full conditional normal posterior. To this end, rewrite the model as in (3) to yield $\widetilde{\mathbf{y}}_t = \widetilde{\mathbf{X}}_t \boldsymbol{b} + \mathbf{W}_t^{1/2}\mathbf{H}_t^{1/2}\boldsymbol{\epsilon}_t$, a heteroskedastic multivariate regression.
2. Sample $\mathbf{A}$ as in step 2 of Section A.12.2 with $\xi_{it} h_{it}^2$ as the known error variance of equations $i = 2, \ldots, k$ of the triangular equation system.
3. Sample $\mathbf{h}_{0:T}$ as in step 3 of Section 3.2 with $\widetilde{\widetilde{\mathbf{u}}}_t = \mathbf{W}_t^{-1/2}\left[\mathbf{A}(\mathbf{y}_t - \mathbf{B}\mathbf{x}_t) - (\mathbf{W}_t - \overline{\mathbf{W}})\boldsymbol{\gamma}\right]$
4. Sample $\boldsymbol{\sigma}^2$ as in step 4 of Section 3.2
5. The full conditional posterior for ν_i is proportional to $\mathcal{G}(2, 0.1) \prod_{t=1}^T \mathcal{IG}(\xi_{it}, \nu_i/2, \nu_1/2)$ and is sampled as in step 5 of section 3.2
6. Sample ξ_{it} as in step 5 of Section A.12.2 with $q_{it}^2 = u_{it}^2/h_{it}$ and $p_{it}^2 = \gamma_i^2/h_{it}$

B Numerical Performance and Convergence

As the models we propose are relatively complicated with many latent variables the numerical performance of the MCMC algorithm and it's convergence properties are of interest. Here, we briefly report on these issues. In Table 1 we report on the run times for the MCMC algorithms relative to the base case of the VAR with Gaussian stochastic volatility. The Gaussian, Student-t and Skew-t without SV makes use of conditional conjugacy. For the OST model, the simulation of the generalized inverse Gaussian distribution for the mixing variables, ξ_{it}, is relatively time consuming.

Table 1: Relative time for the MCMC algorithm for the different VAR models

	Gaussian	Student-t	Skew-t	OT	MT	OST	MST
Non SV	0.64	0.67	1.26	0.73	1.30	3.30	1.35
SV	1.00	1.08	1.67	1.14	1.75	3.83	1.78

Times relative to the Gaussian VAR with stochastic volatility. For the Gaussian VAR with stochastic volatility 10 000 draws takes about 1 minute on an Intel Core i7-8700 processor (8 cores at 3.2 GHz).

Regarding converge, Table 2 reports on the convergence of the slowest mixing parameters , σ_i, the standard deviations of the innovations to the log volatilites, γ_i, the skewness parameters, and ν_i, the degrees of freedom, for the MST-SV and OST-SV models. The table shows the posterior mean and standard deviations along with the upper confidence 95% limit of the Gelman and Rubin (1992) $\hat{R}$ statistic. In no case do the statistics indicate a lack of convergence.

Table 2: Convergence diagnostic of the MST-SV model and the OST-SV model for the parameters σ, γ and ν

	Mean	Sd.	$\hat{R}$	Mean	Sd.	$\hat{R}$
		MST-SV			OST-SV	
σ_1	0.015	0.015	1.034	0.047	0.045	1.065
σ_2	0.059	0.024	1.001	0.062	0.025	1.002
σ_3	0.002	0.002	1.047	0.002	0.001	1.004
σ_4	0.003	0.002	1.006	0.003	0.002	1.005
γ_1	0.114	0.148	1.049	0.519	0.499	1.065
γ_2	0.029	0.087	1.001	0.029	0.092	1.001
γ_3	0.015	0.074	1.003	-0.003	0.071	1.001
γ_4	0.123	0.036	1.006	0.122	0.037	1.002
ν_1	10.579	5.805	1.054	23.593	15.475	1.055
ν_2	27.455	14.343	1.001	30.536	14.427	1.004
ν_3	39.589	16.200	1.006	37.701	14.801	1.003
ν_4	11.028	3.012	1.010	10.968	3.027	1.005

The table shows the estimation of the parameters σ, γ, ν, and the Gelman and Rubin's convergence statistics ($\hat{R}$ statistics, Gelman and Rubin (1992)). We calculate the $\hat{R}$ based on five chains of 100,000 posterior samples with 10,000 draws as burn-in, and thinned at every 10 iterations. The 95% upper confidence limits statistics of $\hat{R}$ are reported and the values are close to 1 indicate the convergence.

C Posterior comparison

Figures 1 and 2 compares the posterior distribution of the skewness, γ_i, and heavy tail, ν_i parameters for the MST and OST specifications with and without stochastic volatilities.

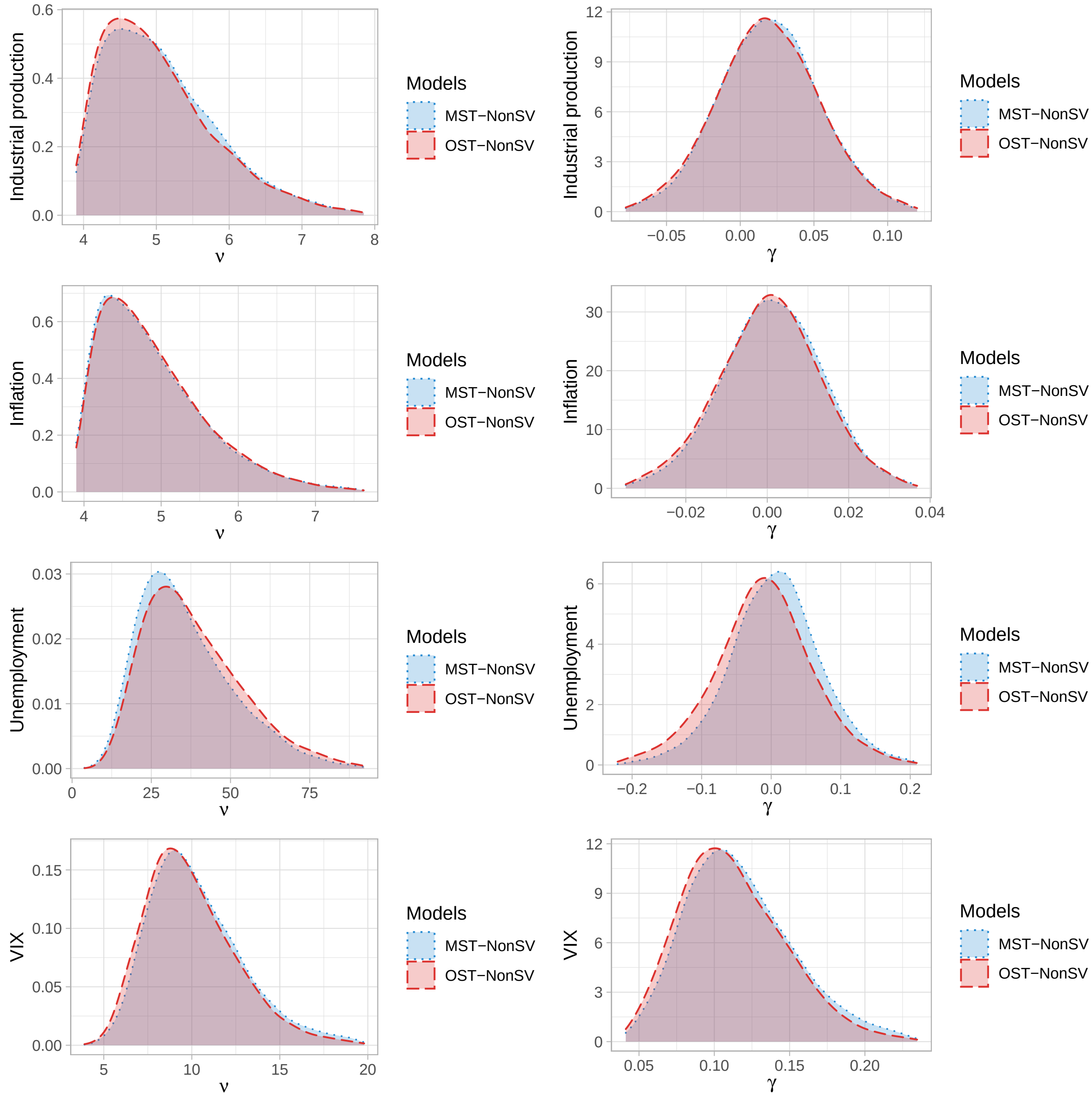

Figure 1: The plots show the posterior samples of the heavy tail and skewness parameters of the VAR models with MST and OST innovations.

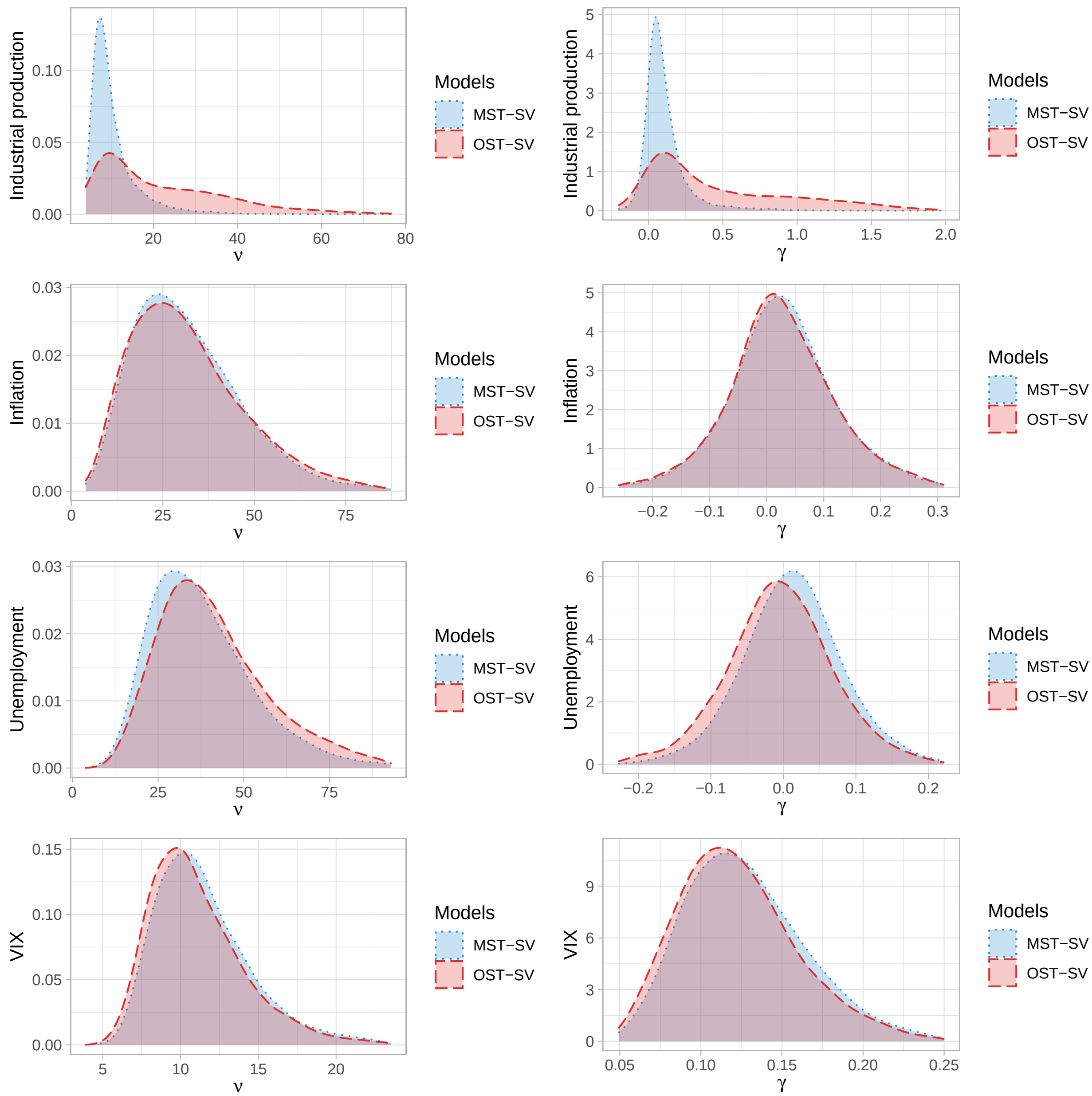

Figure 2: The plots show the posterior samples of the heavy tail and skewness parameters of the the VAR models with MST-SV and OST-SV innovations.

D Sensitivity to the order of variables

Figures 3 and 4 compare the posterior distribution of the skewness (γ_i) and heavy tail (ν_i) for two orderings of the variables for the MST and OST specifications where the results may be sensitive to the ordering. The orderings are the original order used in the paper; Industrial production, Inflation, Unemployment, VIX, and the alternative ordering; Inflation, VIX, Industrial production, Unemployment. The posterior distribution is largely unaffected by the ordering.

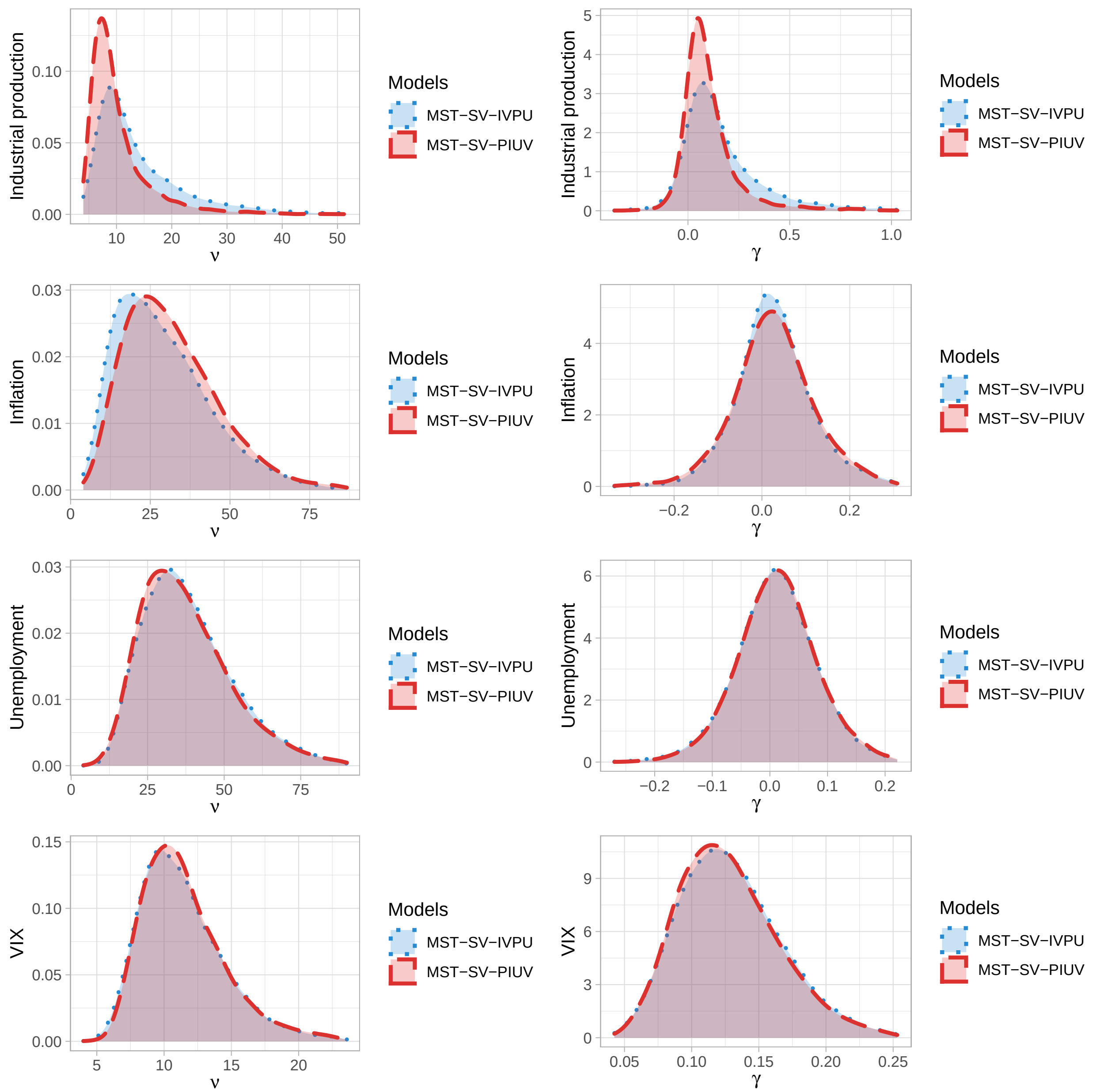

Figure 3: Posterior distribution of the heavy tail (left column) and skewness (right column) parameters of the VAR models with MST-SV. The order PIUV stands for Industrial production, Inflation, Unemployment, VIX and The order IVPU stands for Inflation, VIX, Industrial production, Unemployment.

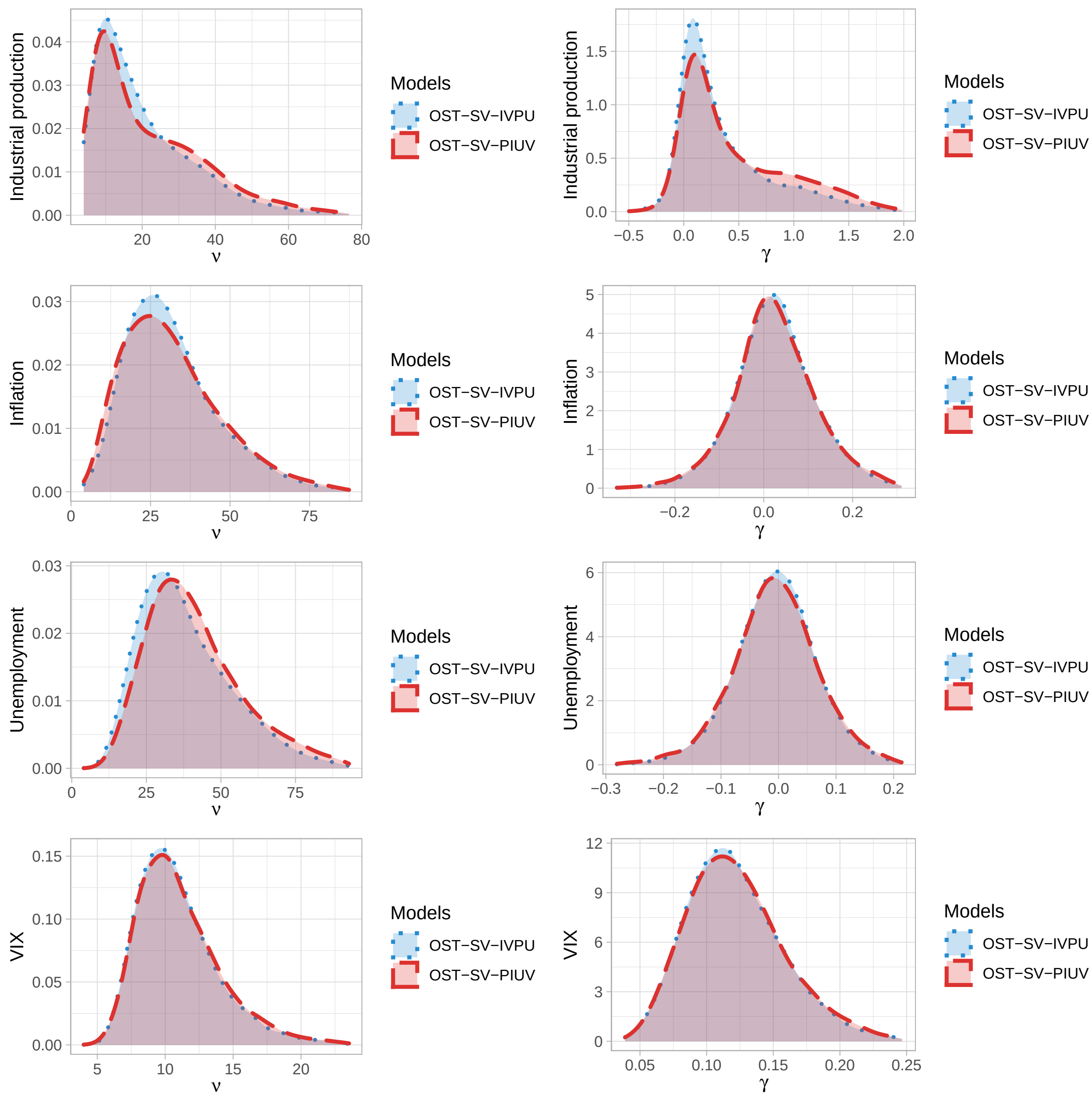

Figure 4: Posterior distribution of the heavy tail (left column) and skewness (right column) parameters of the VAR models with OST-SV. The order PIUV stands for Industrial production, Inflation, Unemployment, VIX and The order IVPU stands for Inflation, VIX, Industrial production, Unemployment.

E Forecast metrics

We compare the forecast accuracy using the mean square forecast error (MSFE) for the point forecast, the log predictive density (LP), and the continuous rank probability score (CRPS) of the posterior predictive distribution for the density forecast.

Let T_0 be the last observation in the first estimation sample and T_1 the last observation on variable i. The MSFE of variable i at h step ahead, for $h = 1, \dots, H$, is then obtained as,

$$\text{MSFE}_{i,h} = \frac{1}{T_1 - T_0 - h + 1} \sum_{t=T_0}^{T_1-h} \left(\bar{y}_{i,t+h|t} - y^o_{i,t+h}\right)^2,$$

where $\bar{y}_{i,t+h|t}$ is the mean of the posterior predictive samples using all data up to time t and $y^o_{i,t+h}$ is the observed outcome of variable i at h steps ahead. The model with a smaller MSFE is preferred.

The LP of the posterior predictive distribution is computed as,

$$\begin{aligned}
\text{LP}_{i,h} &= \frac{1}{T_1 - T_0 - h + 1} \sum_{t=T_0}^{T_1-h} \left[\log p(y^o_{i,t+h}|\mathbf{y}_{1:t}) \right] \\
&= \frac{1}{T_1 - T_0 - h + 1} \sum_{t=T_0}^{T_1-h} \left[\log \int_{\boldsymbol{\theta}} p(y^o_{i,t+h}|\boldsymbol{\theta}, \mathbf{y}_{1:t}) p(\boldsymbol{\theta}|\mathbf{y}_{1:t}) d\boldsymbol{\theta} \right]
\end{aligned}$$

where $p(y^o_{i,t+h}|\mathbf{y}_{1:t})$ is the h-step ahead posterior predictive density function evaluated at the realization of the variable. Following Andersson and Karlsson (2008), the LP of the posterior predictive distribution is computed using the Rao-Blackwellization idea which is more stable than the kernel density estimator for extreme observations. In particular, it is evaluated as,

$$\text{LP}_{i,h} = \frac{1}{T_1 - T_0 - h + 1} \sum_{t=T_0}^{T_1-h} \left[\log \sum_{r=1}^{R} \frac{1}{R} p(y^o_{i,t+h}|\boldsymbol{\theta}^{(r)}, \mathbf{y}_{1:t}) \right]$$

where $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(R)}$ are the posterior samples of the VAR model. The possibly high dimensional integral over intermediate observations implicit in $p(y^o_{i,t+h}|\boldsymbol{\theta}^{(r)}, \mathbf{y}_{1:t})$ can be approximated by the Monte Carlo approach. For each sample from the posterior we simulate a new path $\mathbf{y}^{(r)}_{(t+1):(t+h-1)|t}$ using the data generating process for the model and calculate $p(\mathbf{y}_{i,t+h|t}|\boldsymbol{\theta}^{(r)}, \mathbf{y}_{1:t}, \mathbf{y}^{(r)}_{(t+1):(t+h-1)|t})$. A higher LP value indicates a better density forecasting performance of the model.

The continuous rank probability score (CRPS) is also commonly used to rank the density forecasts. CRPS is obtained as the quadratic difference between the predictive cumulative distribution function and the empirical distribution of the variable (Gneiting and Raftery, 2007). As Clark and Ravazzolo (2015) noted the CRPS is less sensitive to outliers than the LP and rewards more for values of the predictive density that are close to the outcome.

$$CRPS_{i,h} = \frac{1}{T_1 - T_0 - h + 1} \sum_{t=T_0}^{T_1-h} \left[-E_f \left| y_{i,t+h|t} - y^o_{i,t+h} \right| + 0.5 E_f \left| y_{i,t+h|t} - y'_{i,t+h|t} \right| \right],$$

where f is the predictive density of the variable $y_{i,t+h|t}$, and $(y_{i,t+h|t}, y'_{i,t+h|t})$ are independent random draws from the predictive density f. We apply the Monte Carlo method to simulate 10,000 draws from the predictive density f and compute the expectation.

F Forecast evaluation

Table 3 reports the relative improvements in CRPS over the Gaussian VAR models where entries greater than 0 indicate that the given model is better. We confirm the previous conclusion by comparing the CRPS among models. However, the effect of heavy tails and skewness is smaller as the CRPS is less sensitive to outliers (Clark and Ravazzolo, 2015). Focusing on the models with stochastic volatility, skewness and heavy tailed specifications improve significantly on the Gaussian model 6 and 12 month forecast of industrial production and the 12 month forecast for the unemployment rate. For inflation and the VIX the forecast performance differs very little between the specifications.

Table 3: Improvement in CRPS over the Gaussian VAR models

	1M	3M	6M	12M	1M	3M	6M	12M
	(a) Industrial Production				(c) Unemployment rate			
Gaussian	-0.344	-0.357	-0.382	-0.391	-0.172	-0.204	-0.212	-0.220
Student-t	0.008*	0.006*	0.007*	0.003	0.003*	-0.001	-0.003*	-0.004*
Skew-t	0.007*†	0.007*	0.010*†	0.008*†	0.002*	0.000†	-0.001†	-0.003*
OT	0.009*	0.008*	0.005*	0.002	0.003*	-0.002	-0.005*	-0.010*
MT	0.009*	0.006*	0.004	0.001	0.003*	-0.001	-0.005*	-0.009*
OST	0.009*	0.008*	0.007*†	0.007*†	0.003*	-0.001	-0.004*†	-0.008*†
MST	0.008*	0.005*	0.005	0.004*†	0.003*	0.000	-0.004*	-0.008*†
Gaussian-SV	-0.317*	-0.324*	-0.351*	-0.366	-0.146*	-0.174*	-0.171*	-0.171*
Student-t-SV	0.001	0.001	0.004	0.006	0.000	0.001	0.002*	0.004*
Skew-t-SV	-0.000	0.000	0.004*	0.008*	0.000	0.000	0.001	0.001†
OT-SV	-0.001	0.001	0.003	0.007	-0.000	-0.000	0.000	0.001*
MT-SV	-0.000	0.001	0.003	0.008	-0.000	-0.000	0.000	0.001*
OST-SV	-0.000	0.002	0.003*	0.008*	-0.001†	-0.000	0.000	0.001*
MST-SV	-0.000	0.001	0.004*	0.010*	-0.000	-0.000	0.001	0.001*
	(b) Inflation				(d) VIX			
Gaussian	-0.114	-0.196	-0.321	-0.562	-0.123	-0.185	-0.217	-0.250
Student-t	-0.002*	-0.008*	-0.013*	-0.021*	-0.000	-0.005*	-0.008*	-0.009*
Skew-t	-0.002*†	-0.007*†	-0.007*†	-0.006†	0.001*†	-0.001†	-0.003*†	-0.005*
OT	-0.001*	-0.002*	-0.003	-0.009*	0.000	-0.003*	-0.006*	-0.008
MT	-0.001*	-0.003*	-0.004*	-0.012*	0.000	-0.003*	-0.006*	-0.007
OST	-0.001*†	-0.001†	0.001†	-0.001†	0.003*†	0.004*†	0.004*†	0.004*†
MST	-0.000*	-0.002*†	-0.001†	-0.004†	0.002*†	0.003*†	0.002†	0.002†
Gaussian-SV	-0.081*	-0.141*	-0.232*	-0.446*	-0.096*	-0.153*	-0.181*	-0.214*
Student-t-SV	0.000	-0.001	-0.002	-0.005	0.001*	0.000	0.001	-0.001
Skew-t-SV	0.000	-0.000†	-0.001	-0.003	0.001	-0.000	-0.000	-0.003
OT-SV	-0.000	-0.001	-0.002	-0.003	0.001*	0.000	0.000	-0.001
MT-SV	0.000	-0.001	-0.002	-0.004	0.001*	0.000	0.000	-0.000
OST-SV	-0.000	-0.000	-0.002	-0.005	0.000	-0.000	-0.001	-0.003
MST-SV	-0.000	-0.001*	-0.003	-0.007	0.000	-0.001	-0.001	-0.004

Each panel reports the CRPS of the models relative to the Gaussian VAR model with (and without) stochastic volatility. The relative improvements over the Gaussian models are computed as the difference between the CRPS of alternative specifications and the Gaussian models during 2000-2019. We perform a two-sided Diebold and Mariano (1995) test. * denotes that the corresponding model is significantly different from the Gaussian VAR at the 10% level. † denotes that the skew Student model significantly different from the corresponding Student at the 10% level. The entries greater than 0 indicate that the given model is better.

Figure 5 shows the cumulative log Bayes factors of the predictive density for the 3-month forecast horizon between the MST-SV and OST-SV models. As the OST-SV model allows the co-movement of variables in extreme events, the out-of-sample forecast during the 2008-2009 recessions is better than the MST-SV model. It suggests that an appropriate distribution of the VAR model's innovations needs to take into account not only the heavy tails and skewness of each marginal but also the joint co-movement of variables which is helpful during extreme events.

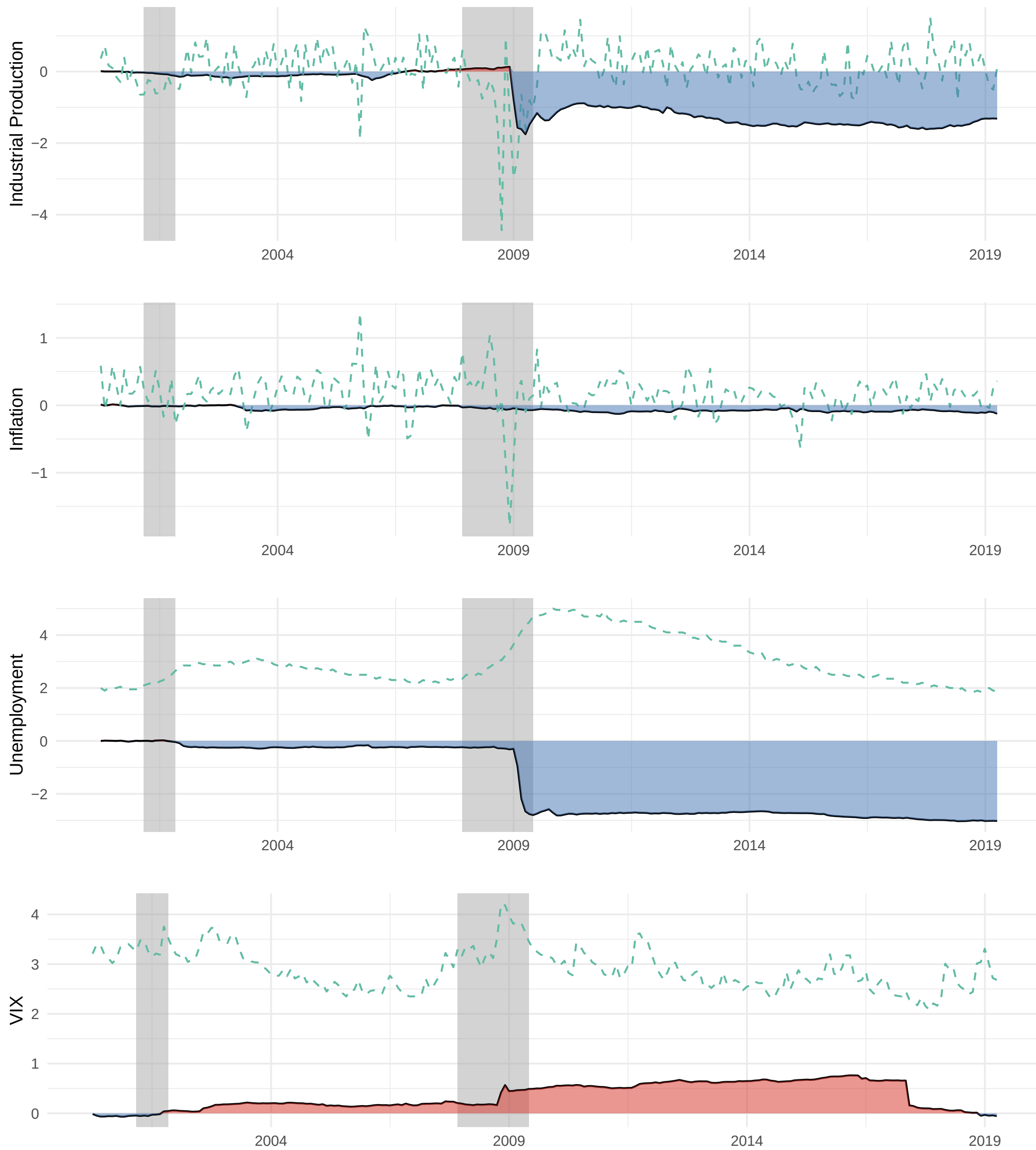

Figure 5: Cumulative log Bayes factors of the predictive density for the 3-month ahead forecast between the MST-SV and OST-SV models.

Positive values (red) means the MST-SV predicts better and negative values (blue) means that the OST-SV model does better. The dashed lines illustrate the scaled values of the original variables. See Geweke and Amisano (2010) for details.

References

Andersson, M. K. and Karlsson, S. (2008). Bayesian forecast combination for VAR models. In Chib, S. and Griffiths, W., editors, *Bayesian econometrics*, volume 23, pages 501–524. Emerald Group Publishing.

Clark, T. E. and Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30(4):551–575.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):134–144.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, pages 457–472.

Geweke, J. and Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26(2):216–230.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378.