ÖREBRO UNIVERSITET

Vector autoregression models with skewness and heavy tails

Hoang Nguyen

joint with **Sune Karlsson, Stepan Mazur**
School of Business - Örebro University, Sweden

Sunday 6th June, 2021

Motivation

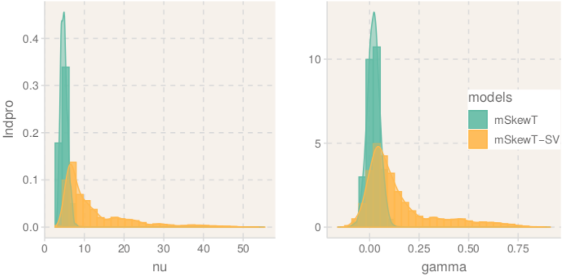
- The Gaussian vector autoregression (VAR) model has become one of the key macroeconomic models for policy makers and forecasters (Karlsson, 2013).
  - However, macroeconomic downturns during recessions and crisis can hardly be explained by a Gaussian structural shock (Mishkin, 2011; Acemoglu et al., 2017).
- There are several proposals for extending the Gaussian VAR model.
  - VAR models with Stochastic volatility (Uhlig, 1997; Clark, 2011).
  - VAR models with Time varying parameter (Primiceri, 2005; Cogley and Sargent, 2005).
  - VAR models with Heavy tail distributions (Clark and Ravazzolo, 2015; Chiu et al., 2017; Liu, 2019).

Related literature & Research questions

- The non-normality and fat tail characteristics of macroeconomic variables have been documented in many studies.
  - Fagiolo et al. (2008) analyze the output growth rates of OECD countries.
  - Ascari et al. (2015) shows findings on U.S. consumption, investment, employment, inflation and real wage.
  - Christiano (2007) found an evidence against the Gaussian assumption of an estimated VAR by inspecting the skewness and kurtosis properties of residuals.
  - Ni and Sun (2005) propose the VAR models with multivariate Student-t distribution (also Cúrdia et al. (2014) and Chib and Ramamurthy (2014)).
  - Panagiotelis and Smith (2008) first come up with an application of multivariate skew-t VAR models.
  - Liu (2019) estimates different fat tail and asymmetry distributions for macroeconomic variables, even though, the symmetric Student-t distribution is preferred among proposals for monthly data.
  - Cross and Poon (2016), Chiu et al. (2017) and Liu (2019) show that ignoring the stochastic volatility of the shocks will overestimate the fatness of the tail distribution.
- There is a gap in the literature on the combination of heavy tails, skewness and stochastic volatility for the VAR model.

Preliminary results

- We propose a general class of Generalized Hyperbolic Skew Student's-t distribution with stochastic volatility (Skew-t.SV) VAR.
- In general, fat tail models improve in-sample goodness of fit and out-of-sample forecast.
- Slight evidence of skewness.
- Ignoring the stochastic volatility of the shocks will overestimate the fatness of the tail distribution **and underestimate the skewness**.



1. VAR Models
   - Gaussian VAR Models
   - Orthogonal Skew-t VAR Models
   - Multi-Skew-t VAR Models
   - Comparison of the model implied distributions
2. Bayesian Inference
3. Empirical illustration
   - Numerical Performance and Convergence
   - Model comparison and Forecast metrics
4. Conclusion

Gaussian VAR Models

The Gaussian VAR model with stochastic volatility (Gaussian.SV) is discussed as following,

$$y_t = c + B_1 y_{t-1} + \ldots + B_p y_{t-p} + A^{-1} H_t^{1/2} \epsilon_t, \tag{1}$$

where

- y_t is a k-dimensional vector of endogenous variables that $y_t = [y_{1t}, \ldots, y_{kt}]'$;
- c is a k-dimensional vector of constant;
- B_j a k x k matrix of regression coefficients for $j = 1, \ldots, p$;
- A is a k x k lower triangular matrix that describes the contemporaneous interaction of the endogenous variables;
- H_t is a k x k diagonal matrix that captures the heteroskedastic volatility; the vector of Gaussian shock $\epsilon_t \sim Normal(0, I_k)$. And

$$\log h_{it} = \log h_{it-1} + \sigma_i \eta_{it} \tag{2}$$

for $i = 1, \ldots, k$ and $\eta_{it} \sim Normal(0, 1)$.

Gaussian VAR Models

The Gaussian VAR model with stochastic volatility (Gaussian.SV) is discussed as following,

$$y_t = Bx_t + u_t, \tag{3}$$

where

- $B = [c, B_1, \dots, B_p]$ is a k x (1+kp) dimensional matrix,
- $x_t = [1, y'_{t-1}, \dots, y'_{t-p}]'$ is (1+kp) dimensional vector
- $u_t = A^{-1} H_t^{1/2} \epsilon_t$ is a k-dimensional vector of heteroskedastic shocks associated with the VAR equations.
- $Au_t = H_t^{1/2} \epsilon_t$ is a k-dimensional vector of orthogonal shocks.

Orthogonal Skew-t (OST) VAR Models

Following Cúrdia et al. (2014), Clark and Ravazzolo (2015), and Chiu et al. (2017), we account for the asymmetric heteroskedastic shocks in the orthogonal residuals $Au_t = A(y_t - Bx_t)$ of the VAR models by assuming that

$$\widetilde{\boldsymbol{\epsilon}}_t = \mathrm{A}\mathrm{u}_t = \mathrm{A}(\mathrm{y}_t - \mathrm{B}\mathrm{x}_t) = (\mathrm{W}_t - \overline{\mathrm{W}})\boldsymbol{\gamma} + \mathrm{W}_t^{1/2}\mathrm{H}_t^{1/2}\boldsymbol{\epsilon}_t, \tag{4}$$

where

- the mixing variable $W_t = diag(\xi_{1t}, \ldots, \xi_{kt})$ is a diagonal matrix that $\xi_{it} \sim InvGamma(\frac{\nu_i}{2}, \frac{\nu_i}{2})$ and ν_i is the degree of freedom for $i = 1, \ldots, k$ and $t = 1, \ldots, T$.
- $\gamma = [\gamma_1, \ldots, \gamma_k]'$ is a k-dimensional vector of the skewness parameters.
- Equation (4) represents the marginal distribution of the orthogonal shock Au_t as a vector of independent univariate generalized hyperbolic skew Student-t distributions, see McNeil et al. (2015).
- Orthogonal Student-t VAR model and Gaussian VAR model as special cases.

Multi-Skew-t (MST) VAR Models

We propose a class of the multi-Skew-t (MST) VAR model by assuming the residuals $u_t = (y_t - Bx_t)$ as,

$$\mathrm{u}_t = (\mathrm{W}_t - \overline{\mathrm{W}})\boldsymbol{\gamma} + \mathrm{W}_t^{1/2}\mathrm{A}^{-1}\mathrm{H}_t^{1/2}\boldsymbol{\epsilon}_t. \tag{5}$$

- Comparing to the OST VAR model, the MST VAR model imposes the fat tail and skewness of the structure shock directly in each VAR equation rather than in the idiosyncratic shock.
- The tail co-movement can be defined by restricting that $\xi_{1t} = \ldots = \xi_{kt}$ so the marginal distribution of u_t is a multivariate generalized hyperbolic skew Student-t (Skew-t) distribution as a special case, see McNeil et al. (2015).
- It becomes a multivariate Student-t (Student-t) distribution when $\xi_{1t} = \ldots = \xi_{kt}$ and $\gamma_1 = \ldots = \gamma_k = 0$.

Comparison of the model implied distributions

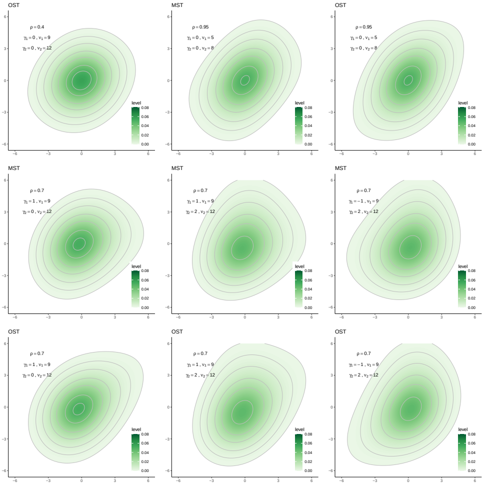


Figure: The density plots of the MST and OST distributions with different scale parameters h_{1t} and h_{2t}. It is assumed that $\rho = 0.5$, $\boldsymbol{\gamma} = (1, 2)$, and $\boldsymbol{\nu} = (9, 12)$.

Comparison of the model implied distributions

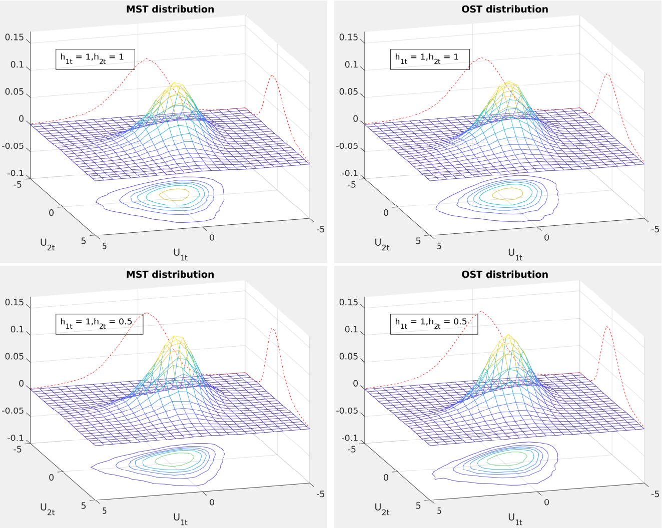


Figure: The density plots of the MST and OST distributions with different scale parameters h_{1t} and h_{2t}. It is assumed that $\rho = 0.5$, $\gamma = (1, 2)$, and $\boldsymbol{\nu} = (9, 12)$.

Bayesian Inference

- The set of the VAR-MST-SV model parameters $\boldsymbol{\theta} = \{\mathrm{B}, \mathrm{a}, \boldsymbol{\gamma}, \boldsymbol{\nu}, \boldsymbol{\sigma}^2, \xi_{1:T}, \mathrm{h}_{0:T}\}$.
- We use the Minnesota-style priors for the coefficients B, see Koop and Korobilis (2010), and vague prior distributions for other parameters.

  - $vec(\mathrm{B}) \sim \mathcal{N}(\mathrm{b}_0, \mathrm{V}_{\mathrm{b}_0})$
  - $a \sim \mathcal{N}(0_k, 10I_k)$
  - $\gamma \sim \mathcal{N}(0_k, \mathrm{I}_k)$
  - $\nu \sim \mathcal{G}amma(2, 0.1)$
  - $\sigma_i^2 \sim \mathcal{G}amma(1/2, 1/2V_\sigma)$
  - $\log h_{i0} \sim N\left(\log \hat{\Sigma}_{i,OLS}, 4\right)$

Gibb sampler scheme $\boldsymbol{\theta} = \{\mathrm{B}, \mathrm{a}, \boldsymbol{\gamma}, \boldsymbol{\nu}, \boldsymbol{\sigma}^2, \xi_{1:T}, \mathrm{h}_{0:T}\}$

Let's Ψ be a set of conditional parameters except the one that we sample from.

MST VAR Model: $\mathrm{y}_t = \mathrm{B}\mathrm{x}_t + (\mathrm{W}_t - \overline{\mathrm{W}})\boldsymbol{\gamma} + \mathrm{W}_t^{1/2}\mathrm{A}^{-1}\mathrm{H}_t^{1/2}\boldsymbol{\epsilon}_t$.

- Sample $\pi(\boldsymbol{b}|\Psi)$ where $\boldsymbol{b} = (\mathrm{B}', \boldsymbol{\gamma}')'$.

$$
\begin{aligned}
\mathrm{y}_t &= \mathrm{B}\mathrm{x}_t + (\mathrm{W}_t - \overline{\mathrm{W}})\boldsymbol{\gamma} + \mathrm{W}_t^{1/2}\mathrm{A}^{-1}\mathrm{H}_t^{1/2}\boldsymbol{\epsilon}_t, \\
\mathrm{A}\mathrm{W}_t^{-1/2}\mathrm{y}_t &= \mathrm{A}\mathrm{W}_t^{-1/2}\mathrm{B}\mathrm{x}_t + \mathrm{A}\mathrm{W}_t^{-1/2}(\mathrm{W}_t - \overline{\mathrm{W}})\boldsymbol{\gamma} + \mathrm{H}_t^{1/2}\boldsymbol{\epsilon}_t \\
&= \mathrm{x}_t' \otimes \mathrm{A}\mathrm{W}_t^{-1/2}\vec{(}\mathrm{B}) + \mathrm{A}\mathrm{W}_t^{-1/2}(\mathrm{W}_t - \overline{\mathrm{W}})\boldsymbol{\gamma} + \mathrm{H}_t^{1/2}\boldsymbol{\epsilon}_t \\
&= (\mathrm{x}_t' \otimes \mathrm{A}\mathrm{W}_t^{-1/2} \quad \mathrm{A}\mathrm{W}_t^{-1/2}(\mathrm{W}_t - \overline{\mathrm{W}}))\boldsymbol{b} + \mathrm{H}_t^{1/2}\boldsymbol{\epsilon}_t, \\
\widetilde{\mathrm{y}}_t &= \widetilde{\mathrm{X}}_t\boldsymbol{b} + \mathrm{H}_t^{1/2}\boldsymbol{\epsilon}_t, \\
\pi(\boldsymbol{b}|\Psi) &\sim \mathcal{N}(\mathrm{b}^*, \mathrm{V}_\mathrm{b}^*)
\end{aligned}
$$

Gibb sampler scheme $\boldsymbol{\theta} = \{\mathrm{B}, \mathrm{a}, \boldsymbol{\gamma}, \boldsymbol{\nu}, \boldsymbol{\sigma}^2, \xi_{1:T}, \mathrm{h}_{0:T}\}$

MST VAR Model: $\mathrm{y}_t = \mathrm{B}\mathrm{x}_t + (\mathrm{W}_t - \overline{\mathrm{W}})\boldsymbol{\gamma} + \mathrm{W}_t^{1/2}\mathrm{A}^{-1}\mathrm{H}_t^{1/2}\boldsymbol{\epsilon}_t$.

- Sample $\pi(\mathrm{a}|\Psi)$ following Cogley and Sargent (2005) and use that (5) is a triangular model for the reduced form residuals,

$$\mathrm{A}\widetilde{\mathrm{u}}_t = \mathrm{H}_t^{1/2}\boldsymbol{\epsilon}_t,$$

  where $\widetilde{\mathrm{u}}_t = \mathrm{W}_t^{-1/2}(\mathrm{y}_t - \mathrm{B}\mathrm{x}_t - (\mathrm{W}_t - \overline{\mathrm{W}})\boldsymbol{\gamma})$.

- Sample $\pi(h_0, \mathrm{H}_{1:T}|\Psi)$ and $\pi(\boldsymbol{\sigma}^2|\Psi) = \pi(\boldsymbol{\sigma}^2|h_0, \mathrm{H}_{1:T})$,

  Let $\tilde{\tilde{\mathrm{u}}}_t = \mathrm{A}\tilde{\mathrm{u}}_t$, for each series $i = 1, \ldots, k$, we have that $\log \tilde{\tilde{u}}_{it}^2 = \log h_{it} + \log \epsilon_t^2$.

  Kim et al. (1998) approximated the distribution of $\log(\epsilon_t^2)$ as $\log(\chi^2)$ using a mixture of 7 Gaussian components.

Gibb sampler scheme $\boldsymbol{\theta} = \{\mathrm{B}, \mathrm{a}, \boldsymbol{\gamma}, \boldsymbol{\nu}, \boldsymbol{\sigma}^2, \xi_{1:T}, \mathrm{h}_{0:T}\}$

MST VAR Model: $\mathrm{y}_t = \mathrm{B}\mathrm{x}_t + (\mathrm{W}_t - \overline{\mathrm{W}})\boldsymbol{\gamma} + \mathrm{W}_t^{1/2}\mathrm{A}^{-1}\mathrm{H}_t^{1/2}\boldsymbol{\epsilon}_t$.

- Sample $\pi(\nu_i|\Psi) \propto \mathcal{G}(\nu_i; 2, 0.1)\prod_{t=1}^{T}\mathcal{IG}\left(\xi_{it}; \frac{\nu_i}{2}, \frac{\nu_i}{2}\right)$ using RW Metropolis Hastings.
- Sample $\pi(\mathrm{W}_t|\Psi)$ for $t = 1, \ldots, T$ using the independent Metropolis Hastings.

Empirical illustration

- Data: Industrial Production, Inflation rate, Unemployment rate, Chicago board options exchange's volatility index (VIX).
- The data is collected at monthly frequency from 01/1970 - 12/2019 from the Federal Reserve Bank of St. Louis, see McCracken and Ng (2016).
- The growth changes are calculated using the first difference of the logarithm of Industrial Production Index and CPI. The VIX is calculated in the log scale.
- 14 VAR models (p=4 lags) w/wo SV: Gaussian, Student-t, hyperbolic skew Student-t (Skew-t), Orthogonal Student-t (oStudent-t), Orthogonal Hyperbolic skew Student-t (OST), multi-Student-t (mStudent-t), multi-hyperbolic skew Student-t (MST).
- 146 recursive forecast during 2007-2019.

Data

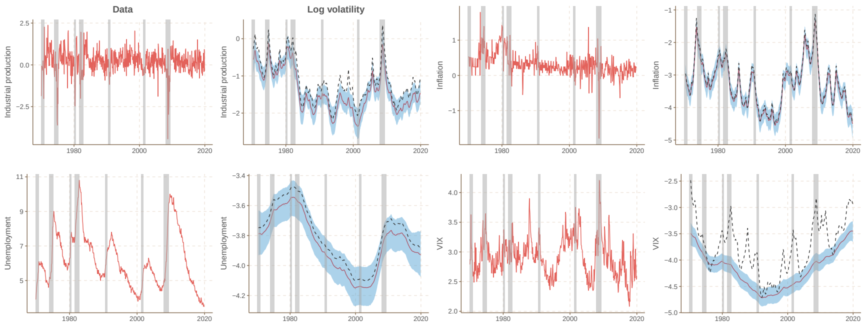


Figure: The plots show the growth of industrial production, inflation rate, unemployment rate, VIX in comparison to their estimated stochastic volatility from the OST.SV model.

The figures on the left-hand side show the growth changes of the variables while the figures on the right-hand side draw the estimated mean log volatility of the OST.SV model in red color with their 50% credible interval. The dash line shows the estimated mean log volatility from the Gaussian.SV model. The shaded areas highlight the recession periods based on the NBER indicators.

Numerical Performance and Convergence

Table: 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).

Table: 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	0.017	0.016	1.034	0.043	0.042	1.065
σ_2	0.060	0.023	1.001	0.061	0.024	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	0.128	0.160	1.049	0.464	0.480	1.065
γ_2	0.028	0.091	1.001	0.027	0.090	1.001
γ_3	0.011	0.072	1.003	-0.003	0.074	1.001
γ_4	0.121	0.036	1.006	0.122	0.037	1.002
ν_1	11.312	6.041	1.054	21.480	14.264	1.055
ν_2	28.721	14.287	1.001	30.416	15.279	1.004
ν_3	38.397	15.827	1.006	39.466	16.119	1.003
ν_4	10.926	2.935	1.010	10.953	2.967	1.005

The table shows the estimation of the parameters σ, γ, ν, and the Gelman and Rubin's convergence statistics ($\hat{R}$ statistics, **?**). 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

Posterior comparison of MST models w/wo SV

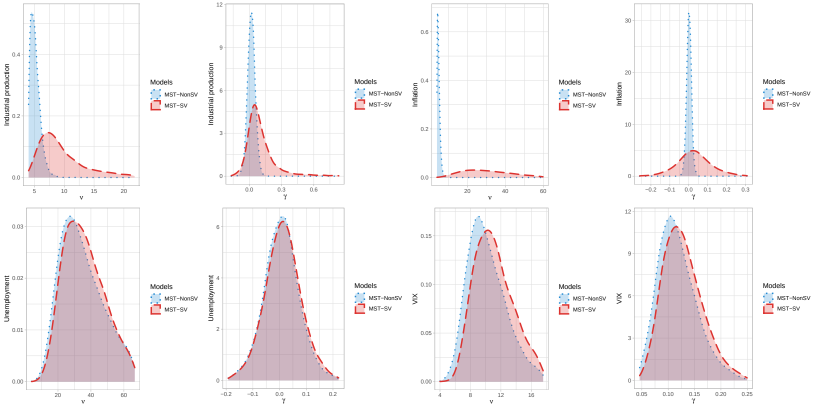


Figure: The plots show the posterior samples of the fat tail parameters and skewness parameters of the MST model with/without SV.

Posterior comparison of OST models w/wo SV

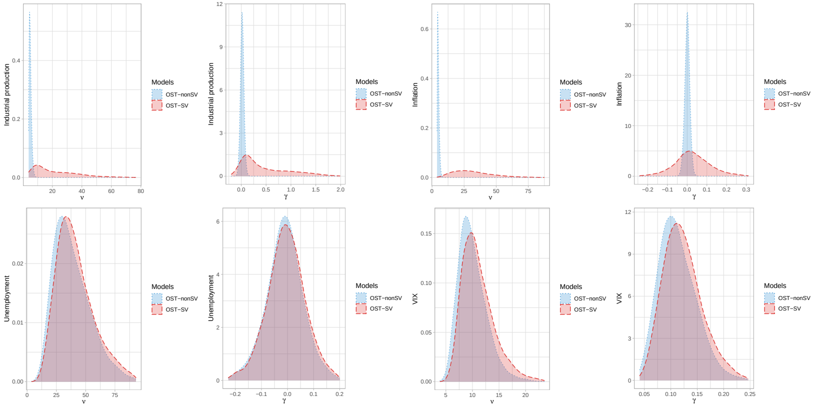


Figure: The plots show the posterior samples of the fat tail parameters and skewness parameters of the OST model with/without SV.

Model comparison and Forecast metrics

We compare 14 VAR models with different assumptions of tail distribution and stochastic volatility using the in-sample goodness of fit and out-of-sample forecast.

- The model marginal likelihood is calculated based on the cross entropy methods, see Chan and Eisenstat (2018).
- The out-of-sample forecast among models can also be measured by the mean square forecast error (MSFE), the log predictive density (LP), and the scale continuous rank probability scores (SCRPS) of the posterior predictive distribution.

Marginal likelihood

The marginal likelihood of the Skew-*t*-SV-VAR model requires high-dimensional integration

$$p(\mathrm{y}_{1:T}) = \int p(\mathrm{y}_{1:T}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \quad (6)$$

- Approximate using the important sampling,

$$\hat{p}_{IS}(\mathrm{y}_{1:T}) = \frac{1}{N}\sum_{n=1}^{N}\frac{\hat{p}(\mathrm{y}_{1:T}|\boldsymbol{\theta}^{(n)})p(\boldsymbol{\theta}^{(n)})}{f(\boldsymbol{\theta}^{(n)}|\lambda^*)}.$$

- Divide θ into two parameter groups $\boldsymbol{\theta}_1 = \{vec(\mathrm{B})', \mathrm{a}', \boldsymbol{\gamma}', \boldsymbol{\nu}', \boldsymbol{\sigma}^{2'}, \mathrm{h}_0'\}'$ and $\boldsymbol{\theta}_2 = \{\mathrm{W}_{1:T}', \mathrm{H}_{1:T}'\}'$. Then, calculate the integrate likelihood $\hat{p}(\mathrm{y}_{1:T}|\boldsymbol{\theta}_1^{(n)})$ first.

$$\hat{p}_{IS}(\mathrm{y}_{1:T}) = \frac{1}{N}\sum_{n=1}^{N}\frac{\hat{p}(\mathrm{y}_{1:T}|\boldsymbol{\theta}_1^{(n)})p(\boldsymbol{\theta}_1^{(n)})}{f(\boldsymbol{\theta}_1^{(n)}|\lambda^*)}.$$

Marginal likelihood

The marginal likelihood of the Skew-t-SV-VAR model requires high-dimensional integration

$$p(\mathrm{y}_{1:T}) = \int p(\mathrm{y}_{1:T}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \quad (7)$$

Algorithm 1. (Marginal likelihood estimation via the cross-entropy method)

1. Two parameter groups $\boldsymbol{\theta}_1 = \{vec(\mathrm{B})', \mathrm{a}', \boldsymbol{\gamma}', \boldsymbol{\nu}', \boldsymbol{\sigma}^{2\prime}, \mathrm{h}_0'\}'$ and $\boldsymbol{\theta}_2 = \{\mathrm{W}_{1:T}', \mathrm{H}_{1:T}'\}'$.
   Obtain the posterior samples $\boldsymbol{\theta}_1^{(1)}, \ldots, \boldsymbol{\theta}_1^{(R)}$ from the posterior density $\pi(\boldsymbol{\theta}|\mathrm{y}_{1:T})$.
2. Consider the parametric family $f(\boldsymbol{\theta}_1; \lambda)$ parameterized by parameter λ such that

$$\lambda^* = \arg\max_{\lambda} \frac{1}{R} \sum_{r=1}^{R} \log f(\boldsymbol{\theta}_1^{(r)}|\lambda)$$

3. Obtain new samples $\boldsymbol{\theta}_1^{(1)}, \ldots, \boldsymbol{\theta}_1^{(N)}$ from $f(\boldsymbol{\theta}_1; \lambda^*)$. Then the marginal likelihood is calculated via important sampling

$$\hat{p}_{IS}(\mathrm{y}_{1:T}) = \frac{1}{N} \sum_{n=1}^{N} \frac{\hat{p}(\mathrm{y}_{1:T}|\boldsymbol{\theta}_1^{(n)})p(\boldsymbol{\theta}_1^{(n)})}{f(\boldsymbol{\theta}_1^{(n)}|\lambda^*)}.$$

Marginal likelihood

The integrated likelihood $p(y_{1:T}|\theta_1)$ require a high dimensional integral over the latent states $\theta_2 = \{W'_{1:T}, H'_{1:T}\}'$,

$$p(y_{1:T}|\theta_1) = \int\int p(y_{1:T}|\theta_1, W_{1:T}, H_{1:T}) p(W_{1:T}, H_{1:T}|\theta_1) dW_{1:T} dH_{1:T}$$

$$p(y_{1:T}|\theta_1) = \int p(y_{1:T}|\theta_1, H_{1:T}) p(H_{1:T}|\theta_1) dH_{1:T}$$

$$\approx \sum_{l=1}^{L} \frac{1}{L} \frac{p(y_{1:T}|\theta_1, H^{(l)}_{1:T}) p(H^{(l)}_{1:T}|\theta_1)}{f(H^{(l)}_{1:T}|\lambda_H)}$$

We proposes an important sampling distribution $f(H_{1:T}|\lambda_H)$ and simulate $H^{(l)}_{1:T} \sim f(H_{1:T}|\lambda_H)$ for $l = 1, \ldots, L$. Then, $p(y_{1:T}|\theta_1, H^{(l)}_{1:T})$ is the conditional likelihood which can be derived in a closed form multivariate distribution.

Model comparison

Table: Log marginal likelihood for VAR models with and without stochastic volatility

		Gaussian	Student-t	Skew-t	OT	MT	OST	MST
Non SV	LML	-220.868	-131.325	-139.614	-131.819	-129.095	-131.257	-128.307
	sd	(0.002)	(0.004)	(0.009)	(0.011)	(0.015)	(0.023)	(0.028)
SV	LML	-52.111	-34.413	-32.811	-36.060	-33.513	-26.706	-24.072
	sd	(0.069)	(0.037)	(0.078)	(0.251)	(0.032)	(0.238)	(0.235)

We compare the LMLs of 14 VAR models with/without SV. We use the cross entropy methods by Chan and Eisenstat (2018) to calculate the LMLs. We first sample 100,000 draws from the conditional posterior distributions with 10,000 draws as burn-in. Then, all LMLs estimated using 100,000 draws from the proposal distributions, see details in Section **??**. The standard errors of the estimation using the batch means method (10 batches) are reported in the brackets. Estimates of the log marginal likelihoods using the **?** method are reported in Table **??** in the Appendix.

Probability Integral Transform (PIT) histogram plots $u_t = F(y_t^{obs}|y_{1:t-1})$

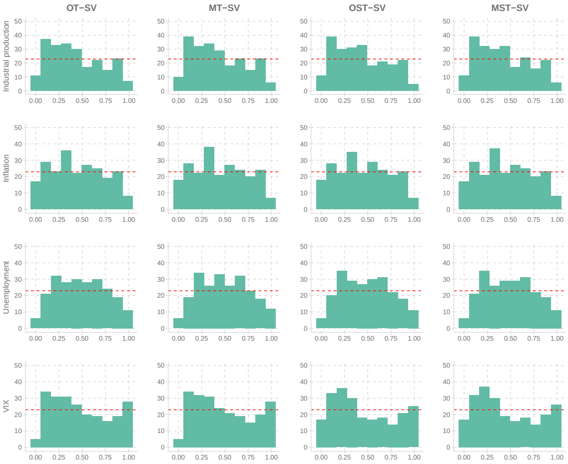


Figure: The plots show the PITs of one month forecast horizon of 146 recursive estimations (2007-2019).

Out-of-samples density forecast - LP

Table: Improvement in LP over the Gaussian VAR model (2007-2019)

	1M	3M	6M	12M	1M	3M	6M	12M
	(a) Industrial Production				(b) Inflation			
Gaussian	-1.005	-1.078	-1.153	-1.187	-0.378	-0.562	-0.626	-0.673
Student-t	0.044*	0.040*	0.040*	0.036*	0.026*	0.004	-0.016*	-0.029*
Skew-t	0.031†	0.032*†	0.036*	0.040*	0.011*†	-0.006†	-0.023*	-0.033*
OT	0.052*	0.041*	0.036*	0.031*	0.038*	0.007	-0.028*	-0.061*
MT	0.049*	0.034*	0.031	0.026*	0.038*	0.008	-0.027*	-0.057*
OST	0.051*	0.043*	0.043*†	0.043*†	0.038*	0.013†	-0.019*†	-0.050*†
MST	0.034*	0.033*	0.033*	0.036*†	0.032*†	0.019	-0.020*†	-0.046*†
Gaussian-SV	-0.850*	-0.883*	-0.984*	-1.024*	-0.031*	-0.242*	-0.243*	-0.247*
Student-t-SV	0.002	0.012	0.025*	0.031*	-0.001	0.007	0.016*	0.019*
Skew-t-SV	0.002	0.006	0.017*	0.030*	-0.007	-0.001	0.003†	0.006†
OT-SV	0.006	0.016	0.035*	0.035*	-0.001	-0.000	0.005	0.004
MT-SV	0.004	0.016	0.038*	0.041*	-0.002	0.001	0.005	0.003
OST-SV	0.006	0.018	0.021*	0.027*	-0.005†	0.001	0.005	0.007
MST-SV	0.001	0.012	0.028*	0.038*	-0.005	0.000	0.004	0.003

The first line in each panel reports the LP of the benchmark Gaussian VAR model without stochastic volatility. The relative improvement over the benchmark is computed as the average of the LP obtained from 146 recursive estimations for different specifications of VAR models during 2007-2019. The entries greater than 0 indicate that the given model is better.

Out-of-samples density forecast - LP

Table: Improvement in LP over the Gaussian VAR model (2007-2019)

	1M	3M	6M	12M	1M	3M	6M	12M
	(c) Unemployment rate				(d) VIX			
Gaussian	-0.066	-0.600	-1.079	-1.599	-0.112	-0.484	-0.646	-0.774
Student-*t*	-0.010*	-0.047*	-0.049*	-0.049*	0.020*	-0.034*	-0.047*	-0.055*
Skew-*t*	-0.011*	-0.037*†	-0.030*†	-0.017*†	0.019*	-0.011*†	-0.017*†	-0.027*
OT	-0.002*	-0.015*	-0.011*	-0.021*	0.026*	-0.023*	-0.037*	-0.047*
MT	0.001*	-0.018*	-0.017*	-0.026*	0.027*	-0.022*	-0.035*	-0.043*
OST	-0.001†	-0.009*†	-0.002†	-0.001†	0.053*†	0.041*†	0.039*†	0.029*†
MST	0.002*†	-0.012*†	-0.007*†	-0.007†	0.039*†	0.030*†	0.027*†	0.020*†
Gaussian-SV	0.522*	-0.016*	-0.528*	-1.257	0.327*	-0.135*	-0.309*	-0.472*
Student-*t*-SV	-0.006*	-0.022*	-0.040	-0.037	0.023*	0.002	0.000	-0.003
Skew-*t*-SV	-0.005*	-0.018	-0.031	-0.043	0.056*†	0.024	0.014	-0.010
OT-SV	-0.002	-0.018	-0.031	-0.021	0.028*	0.000	-0.003	-0.006
MT-SV	-0.004	-0.025	-0.037	-0.008	0.028*	0.001	-0.003	-0.003
OST-SV	-0.002	-0.012	-0.022	-0.029	0.071*†	0.035†	0.017	-0.013
MST-SV	-0.003	-0.025	-0.043	-0.004	0.065*†	0.034†	0.012	-0.018

The first line in each panel reports the LP of the benchmark Gaussian VAR model without stochastic volatility. The relative improvement over the benchmark is computed as the average of the LP obtained from 146 recursive estimations for different specifications of VAR models during 2007-2019. The entries greater than 0 indicate that the given model is better.

Cumulative log Bayes factors

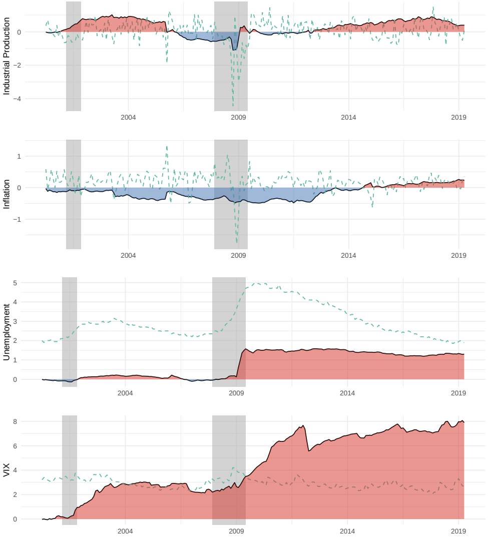


Figure: Cumulative log Bayes factors of the predictive density for the 3-month ahead forecast between the OT SV and OST SV models.

Cumulative log Bayes factors

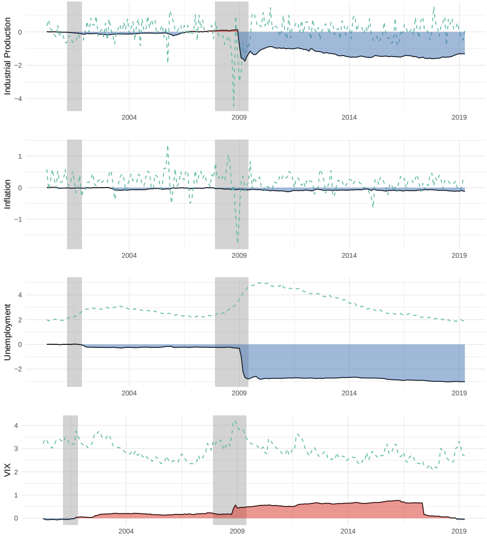


Figure: Cumulative log Bayes factors of the predictive density for the 3-month ahead forecast between the MST-SV and OST-SV models

Conclusion and Discussion

Contributions:

- VAR models with heavy tails and skewness + an R package.

Findings:

- Stochastic volatility improves point and density forecast.
- Ignoring the stochastic volatility of the shocks will overestimate the fatness of the tail distribution **and underestimate the skewness**.
- In general, fat tail improves in-sample goodness of fit and out-of-sample forecast.
- Slight evidence of skewness.

Thank you

Model comparison and Forecast metrics

Following Andersson and Karlsson (2008), the LP of the posterior predictive distribution is computed using the Rao-Blackwellization.

$$\mathrm{LP}_{i,h} = \log p(y^o_{i,t+h}|\mathrm{y}_{1:t}) = \log \int_{\boldsymbol{\theta}} p(y^o_{i,t+h}|\boldsymbol{\theta}, \mathrm{y}_{1:t}) p(\boldsymbol{\theta}|\mathrm{y}_{1:t}) d\boldsymbol{\theta}$$
$$= \log \sum_{r=1}^{R} \frac{1}{R} p(y^o_{i,t+h}|\boldsymbol{\theta}^{(r)}, \mathrm{y}_{1:t})$$

where $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(R)}$ are the posterior samples of the VAR models.

Bolin and Wallin (2019) proposed a scaled version of the CRPS

$$\mathrm{SCRPS}_{i,h} = -\frac{E_f \left| y_{i,t+h|t} - y^o_{i,t+h} \right|}{E_f \left| y_{i,t+h|t} - y'_{i,t+h|t} \right|} - \frac{1}{2} \log E_f \left| y_{i,t+h|t} - y'_{i,t+h|t} \right|,$$

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. SCRPS is locally scale invariant and robust to extreme observations.

D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi. Microeconomic origins of macroeconomic tail risks. *American Economic Review*, 107(1):54–108, 2017.

M. K. Andersson and S. Karlsson. Bayesian forecast combination for var models. *Bayesian econometrics*, 23:501–524, 2008.

G. Ascari, G. Fagiolo, and A. Roventini. Fat-tail distributions and business-cycle models. *Macroeconomic Dynamics*, 19(2):465–476, 2015.

D. Bolin and J. Wallin. Scale invariant proper scoring rules scale dependence: Why the average crps often is inappropriate for ranking probabilistic forecasts. *arXiv preprint arXiv:1912.05642*, 2019.

J. C. Chan and E. Eisenstat. Bayesian model comparison for time-varying parameter VARs with stochastic volatility. *Journal of Applied Econometrics*, 33(4):509–532, 2018.

S. Chib and S. Ramamurthy. Dsge models with student-t errors. *Econometric Reviews*, 33(1-4):152–171, 2014.

C.-W. J. Chiu, H. Mumtaz, and G. Pinter. Forecasting with VAR models: Fat tails and stochastic volatility. *International Journal of Forecasting*, 33(4):1124–1143, 2017.

L. J. Christiano. Comment [On the Fit of New Keynesian Models]. *Journal of Business & Economic Statistics*, 25(2):143–151, 2007. ISSN 07350015.

T. E. Clark. Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, 29(3):327–341, 2011.

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

T. Cogley and T. J. Sargent. Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302, 2005.

J. Cross and A. Poon. Forecasting structural change and fat-tailed events in australian macroeconomic variables. *Economic Modelling*, 58:34–51, 2016.

V. Cúrdia, M. Del Negro, and D. L. Greenwald. Rare shocks, great recessions. *Journal of Applied Econometrics*, 29(7):1031–1052, 2014.

G. Fagiolo, M. Napoletano, and A. Roventini. Are output growth-rate distributions fat-tailed? some evidence from OECD countries. *Journal of Applied Econometrics*, 23(5):639–669, 2008.

S. Karlsson. Forecasting with Bayesian vector autoregression. In *Handbook of economic forecasting*, volume 2, pages 791–897. Elsevier, 2013.

S. Kim, N. Shephard, and S. Chib. Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3):361–393, 1998.

G. Koop and D. Korobilis. *Bayesian multivariate time series methods for empirical macroeconomics*. Now Publishers Inc, 2010.

X. Liu. On tail fatness of macroeconomic dynamics. *Journal of Macroeconomics*, 62: 103154, 2019.

M. W. McCracken and S. Ng. FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589, 2016.

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: Concepts, Techniques and Tools-revised edition*. Princeton university press, 2015.

F. S. Mishkin. Monetary policy strategy: lessons from the crisis. Technical report, National Bureau of Economic Research, 2011.

S. Ni and D. Sun. Bayesian estimates for vector autoregressive models. *Journal of Business & Economic Statistics*, 23(1):105–117, 2005.

A. Panagiotelis and M. Smith. Bayesian density forecasting of intraday electricity prices using multivariate skew-t distributions. *International Journal of Forecasting*, 24(4):710–727, 2008.

G. E. Primiceri. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852, 2005.

H. Uhlig. Bayesian vector autoregressions with stochastic volatility. *Econometrica: Journal of the Econometric Society*, pages 59–73, 1997.