

School of Economics and Finance

Fat-tails in VAR Models

Ching-Wai (Jeremy) Chiu, Haroon Mumtaz and Gabor Pinter

Working Paper No. 714

March 2014

ISSN 1473-0278



Queen Mary
University of London

Fat-tails in VAR models*

Ching-Wai (Jeremy) Chiu[†]

Haroon Mumtaz[‡]

Gabor Pinter[§]

February 28, 2014

Abstract

We confirm that standard time-series models for US output growth, inflation, interest rates and stock market returns feature non-Gaussian error structure. We build a 4-variable VAR model where the orthogonalised shocks have a Student t-distribution with a time-varying variance. We find that in terms of in-sample fit, the VAR model that features both stochastic volatility and Student-t disturbances outperforms restricted alternatives that feature either attributes. The VAR model with Student -t disturbances results in density forecasts for industrial production and stock returns that are superior to alternatives that assume Gaussianity. This difference appears to be especially stark over the recent financial crisis. JEL codes: C32, C53. Key words: Bayesian VAR, Fat Tails, Stochastic volatility.

1 Introduction

In policy circles, increasing attention has been given to fat tail events, particularly since the outbreak of the recent crisis and during the ensuing uncertainty surrounding the political and economic environment. Many argue that recent events can hardly be explained by models that are based on a Gaussian shock structure (Mishkin, 2011). This has been recognised by recent efforts of the DSGE literature including Curdia et al. (2013) and Chib and Ramamurthy (2014) who found evidence that models with a multivariate t-distributed shock structure are strongly favoured by the data over standard Gaussian models. This project seeks to complement these efforts by focusing on VAR models with Student-t distributed shocks (Student, 1908). We build on previous work on univariate models with Student-t distributed shocks by Geweke (1992, 1993, 1994, 2005); Koop (2003), and the seminar paper of Zellner (1976) on the Bayesian treatment of multivariate regression models. In addition, we draw on the DSGE literature (Fernandez-Villaverde and Rubio-Ramirez, 2007; Justiniano and Primiceri, 2008; Liu et al., 2011) by incorporating stochastic volatility of the error structure, because by solely focusing on fat-tails and ignoring lower-frequency changes in the volatility of the shocks (as in Ascari et al., 2012) tends to bias the results towards finding evidence in favour of fat tails, as pointed out by Curdia et al. (2013). Our work is therefore closely related to Curdia et al. (2013). However the focus of our analysis is very different from Curdia et al. (2013). Our main aim is to investigate if allowing for fat tails and stochastic volatility can help improve the empirical performance of Bayesian VARs, both in terms of model fit and forecasting performance.

We find that in terms of in-sample fit, the VAR model that features both stochastic volatility and Student-t disturbances outperforms restricted alternatives that include either feature individually. The VAR model with Student -t disturbances results in density forecasts for industrial production and stock returns that are superior to alternatives that assume Gaussianity. The Student-t assumption appears especially important over the 2008 and 2009 period. Forecast densities for industrial production generated from VARs with Gaussian disturbances assign a zero probability to the low levels of industrial production actually realised in

*The authors would like to thank Christopher Sims and participants at the seminar at Birkbeck College for useful comments. The views in this paper do not reflect the views of the Bank of England.

[†]Bank of England

[‡]Queen Mary College. Corresponding author. h.mumtaz@qmul.ac.uk

[§]Bank of England

late 2008. In contrast, when Student-t shocks are incorporated, the left tail of the forecast density includes the actual outcome.

The structure of the paper is as follows. Section 2 provides a description of the VAR model with stochastic volatility and Student-t distributed shocks (TVARSVOL) model together with the priors and the conditional posteriors and the computation of the marginal likelihood. This section also describes the restricted models considered in our study. Section 3 presents the posterior estimates and compares the models based on in-sample fit and forecasting performance. Section 4 concludes.

2 BVAR model with fat tails and stochastic volatility

The model presented in this section is a multivariate time series model with both time varying variance covariance matrix and Student-t distributed shocks in each of the equations (denoted by TVAR-SVOL). As in Primiceri (2005), the stochastic volatility is meant to capture possible heteroscedasticity of the shocks and potential nonlinearities in the dynamic relationships of the model variables, which are related to the low-frequency changes in the volatility. Introducing t-distribution in the shock structure is meant to capture high-frequency changes of in volatility that are often of extreme magnitudes, hence potentially providing an effective treatment of outliers and extreme events.¹ By allowing for stochastic volatility and t-distributed shocks, we let the data determine whether time variation in the model structure derives from rare but potentially transient events, or from persistent shifts in the volatility regime.

Consider a simple VAR model:

$$Y_t = c + B_1 y_{t-1} + \dots + B_p y_{t-p} + u_t \quad t = 1, \dots, T. \quad (1)$$

where y_t is an $n \times 1$ vector of observed endogenous variables, and c is an $n \times 1$ vector of constants; B_i , $i = 1, \dots, p$ are $n \times n$ matrices of coefficients; u_t are heteroscedastic shocks associated with the VAR equations. In particular, we assume that the covariance matrix of u_t is defined as

$$\text{cov}(u_t) = \Sigma_t = A^{-1} H_t A^{-1'} \quad (2)$$

where A is a lower triangular matrix and $H_t = \text{diag} \left(\sigma_{1,t}^2 \times \frac{1}{\lambda_{1,t}}, \sigma_{2,t}^2 \times \frac{1}{\lambda_{2,t}}, \dots, \sigma_{n,t}^2 \times \frac{1}{\lambda_{n,t}} \right)$ with

$$\ln \sigma_{k,t} = \ln \sigma_{k,t-1} + s_{kt}, \text{var}(s_k) = g_k \quad (3)$$

for $k = 1, 2, \dots, n$.

As shown by Geweke (1993) and Koop (2003), assuming a Gamma prior for $\lambda_{k,t}$ of the form $p(\lambda_{k,t}) = \prod_{t=1}^T \Gamma(1, v_{\lambda,k})$ leads to a scale mixture of normals for the orthogonal residuals $\tilde{\varepsilon}_t = Au_t$ where $\tilde{\varepsilon}_t = \{\tilde{\varepsilon}_{1,t}, \tilde{\varepsilon}_{2,t}, \dots, \tilde{\varepsilon}_{n,t}\}$

² Geweke (1993) proves that this formulation is equivalent to a specification that assumes a Student t-distribution for $\tilde{\varepsilon}_{k,t}$ with $v_{\lambda,k}$ degrees of freedom. Our specification allows the variance of this density to change over time via equation 3.

There are two noteworthy things about the BVAR model. First, it allows for both low and high frequency movements in volatility through the stochastic volatility $\sigma_{k,t}$ and the weights $\lambda_{k,t}$ respectively. Second, note that these features apply to the *orthogonal* residuals Au_t . This assumption allows the degrees of freedom for the Student t-distribution to be independent across equations and simplifies the estimation algorithm.³ However, the assumption also implies dependence on the structure of the A matrix. We show in the sensitivity analysis that the ordering of the key variables does not have an impact on the main results.

¹In an important paper, Jacquier et al. (2004) provides a detailed analysis of this issue in a univariate framework.

²Note that $\Gamma(a, b)$ denotes a Gamma density with mean a and degrees of freedom b .

³Chahad and Ferroni (2014) present a VAR model that incorporates a multivariate t-density for the error term.

2.1 Estimation and model selection

In this section, we describe the prior distributions and provide details of the MCMC algorithm used to estimate the model described above. We also introduce the alternative models considered in this study and discuss the computation of the marginal likelihood for model comparison.

2.1.1 Priors

To define priors for the VAR dynamic coefficients, we follow the dummy observation approach of Banbura et al. (2010). We assume Normal priors, $p(B) \sim N(B_0, S_0)$, where $B = \text{vec}([c; b_j])$, $B_0 = (x_d' x_d)^{-1} (x_d' y_d)$ and $S_0 = (Y_D - X_D b_0)' (Y_D - X_D b_0) \otimes (x_d' x_d)^{-1}$. The prior is implemented by the dummy observations y_D and x_D that are defined as:

$$y_D = \begin{bmatrix} \frac{\text{diag}(\gamma_1 s_1 \dots \gamma_n s_n)}{\tau} \\ 0_{n \times (p-1) \times n} \\ \dots \\ \text{diag}(s_1 \dots s_n) \\ \dots \\ 0_{1 \times n} \end{bmatrix}, \quad x_D = \begin{bmatrix} \frac{J_p \otimes \text{diag}(s_1 \dots s_n)}{\tau} & 0_{np \times 1} \\ 0_{n \times np} & 0_{n \times 1} \\ \dots \\ 0_{1 \times np} & c \end{bmatrix} \quad (4)$$

where γ_1 to γ_n denote the prior mean for the parameters on the first lag obtained by estimating individual AR(1) regressions, τ measures the tightness of the prior on the VAR coefficients, and c is the tightness of the prior on the constant term. We use relatively loose priors and set $\tau = 1$. The scaling factor s_i are set using the standard deviation of the residuals from the individual AR(1) equations. We set $c = 1/1000$, implying a relatively flat prior on the constant. In addition, we introduce priors on the sum of lagged coefficients by defining the following dummy observations:

$$y_S = \frac{\text{diag}(\gamma_1 \mu_1 \dots \gamma_n \mu_n)}{\lambda}, \quad x_S = \begin{bmatrix} \frac{(1_{1 \times p}) \otimes \text{diag}(\gamma_1 \mu_1 \dots \gamma_n \mu_n)}{\lambda} & 0_{n \times 1} \end{bmatrix} \quad (5)$$

where μ_1 to μ_n denote the sample means of the endogenous variables using a training sample, and the tightness of period on this sum of coefficients is set to $\lambda = 10\tau$.

We follow Geweke (1993); Koop (2003) in setting a hierarchical prior on the parameter controlling the degree of freedom of the Student-t distributions $v_{\lambda, n}$ and the weighting vector $\lambda_{k, t}$,

$$p(v_{\lambda, n}) \sim \Gamma(v_0, 2) \quad (6)$$

$$p(\lambda_{k, t}) \sim \Gamma(1, v_{\lambda, n}) \quad (7)$$

In the benchmark case, the prior mean v_0 is assumed to equal 20. This allocates a substantial prior weight to fat-tailed distributions as well as distributions that are approximately Normal. We show in the sensitivity analysis below that a higher value for v_0 produces similar results for key parameters. The rest of the priors are relatively standard. We follow Cogley and Sargent (2005) in setting the prior on the variance of the shocks to the volatility transition equation 3, and propose an inverse-gamma distribution, $p(g_k) \sim IG(D_0, T_0)$, where $T_0 = 1$ and $D_0 = 0.001$ are the degrees of freedom and scale parameter, respectively. The prior for the off-diagonal elements A is $P(A) \sim N(0, 1000)$

2.1.2 The Gibbs sampling algorithm

The Gibbs algorithm for the TVAR-SVOL model cycles through the following six conditional posterior distributions:

1. $G(\lambda_{k, t} \setminus \Psi)$ where Ψ denotes the remaining parameters of the model.
2. $G(v_{\lambda, k} \setminus \lambda_{k, t})$

3. $G(g_{k,t} \setminus \Psi)$
4. $G(\sigma_{k,t}^2 \setminus \Psi)$
5. $G(A \setminus \Psi)$
6. $G(B \setminus \Psi)$

The details of each conditional posterior density is provided below:

Drawing $G(\lambda_{k,t} \setminus \Psi)$ The conditional posterior distributions related to the t-distributed shock structure of the model are described in Koop (2003). Note that conditional on B and A , the orthogonalized residuals can be obtained as $\tilde{\varepsilon}_t = Au_t$. The conditional posterior distribution for $\lambda_{k,t}$ derived in Geweke (1993) applies to each column of $\tilde{\varepsilon}_t$. As shown in Koop (2003) this posterior density is a gamma distribution with mean $(v_{\lambda,k} + 1) / \frac{1}{\sigma_{k,t}} \tilde{\varepsilon}_{k,t}^2 + v_{\lambda,k}$ and degrees of freedom $v_{\lambda,k} + 1$. Note that $\tilde{\varepsilon}_{k,t}$ is the k th column of the matrix $\tilde{\varepsilon}_t$.

Drawing $G(v_{\lambda,k} \setminus \lambda_n)$ The conditional distribution for the degree of freedom parameter capturing the fatness of tails is non-standard and given by:

$$G(v_{\lambda,k} \setminus \lambda_k) \propto \left(\frac{v_{\lambda,k}}{2}\right)^{\frac{T v_{\lambda,n}}{2}} \Gamma\left(\frac{v_{\lambda,n}}{2}\right)^{-N} \exp\left(-\left(\frac{1}{v_0} + 0.5 \sum_{t=1}^T [\ln(\lambda_{t,n}^{-1}) + \lambda_{t,n}]\right) v_{\lambda,n}\right) \quad (8)$$

As in Geweke (1993) we use the Random Walking Metropolis Hastings Algorithm to draw from this conditional distribution. More specifically, for each of the n equations of the VAR, we draw $v_{\lambda,n}^{new} = v_{\lambda,n}^{old} + c^{1/2} \epsilon$ with $\epsilon \sim N(0, 1)$. The draw is accepted with probability $\frac{G(v_{\lambda,n}^{new} \setminus \lambda_n)}{G(v_{\lambda,n}^{old} \setminus \lambda_n)}$ with c chosen to keep the acceptance rate around 40%.

Drawing $G(g_k \setminus \Psi)$ The conditional posterior of $G(g_k \setminus \Psi)$ is inverse Gamma as in Cogley and Sargent (2005). The posterior scale parameter is $D_0 + (\ln \sigma_{k,t}^2 - \ln \sigma_{k,t-1}^2)'$ $(\ln \sigma_{k,t}^2 - \ln \sigma_{k,t-1}^2)$ with degrees of freedom $T + T_0$.

Drawing $G(\sigma_{k,t}^2 \setminus \Psi)$ The conditional posterior $G(\sigma_{k,t}^2 \setminus \Psi)$ is sampled using the Metropolis Hastings algorithm in Jacquier et al. (1994). Given a draw for β the VAR model can be written as $A'(\tilde{Y}_t) \times \bar{H}_t^{1/2} = u_t$ where $\tilde{Y}_t = Y_t - c - \sum_{l=1}^L B_l Y_{t-l} = v_t$ and $VAR(u_t) = \tilde{H}_t$. Here $\bar{H}_t = \text{diag}(\lambda_1, \lambda_2, \dots)$ and $\tilde{H}_t = \text{diag}(\sigma_{1,t}^2, \sigma_{2,t}^2, \dots)$. Conditional on other VAR parameters, the distribution $\sigma_{k,t}^2$ is then given by:

$$\begin{aligned} f(\sigma_{k,t}^2 \setminus \sigma_{k,t-1}^2, \sigma_{k,t+1}^2, u_{n,t}) &= f(u_{n,t} \setminus \sigma_{n,t}^2) \times f(\sigma_{n,t}^2 \setminus \sigma_{n,t-1}^2) \times f(\sigma_{k,t+1}^2 \setminus \sigma_{k,t}^2) \\ &= \frac{1}{\sigma_{k,t}} \exp\left(\frac{-u_{n,t}^2}{2\sigma_{k,t}^2}\right) \times \frac{1}{\sigma_{k,t}^2} \exp\left(\frac{-(\ln \sigma_{k,t}^2 - \mu)^2}{2\sigma_{h_k}}\right), \end{aligned}$$

where μ and σ_{h_k} denote the mean and the variance of the log-normal density $\frac{1}{\sigma_{k,t}^2} \exp\left(\frac{-(\ln \sigma_{k,t}^2 - \mu)^2}{2\sigma_{h_k}}\right)$. Jacquier et al. (1994) suggest using $\frac{1}{\sigma_{k,t}^2} \exp\left(\frac{-(\ln \sigma_{k,t}^2 - \mu)^2}{2\sigma_{h_k}}\right)$ as the candidate generating density with the acceptance probability defined as the ratio of the conditional likelihood $\frac{1}{\sigma_{k,t}^2} \exp\left(\frac{-u_{k,t}^2}{2\sigma_{k,t}^2}\right)$ at the old and the new draw. This algorithm is applied at each period in the sample.

Drawing $G(A\backslash\Psi)$ The conditional posterior $G(A\backslash\Psi)$ for the off-diagonal elements of matrix A is standard. Consider the representation of the system as in Cogley and Sargent (2005), adopted for our 4-variable VAR below:

$$\begin{pmatrix} v_t \\ v_{2t} + v_{1t}\alpha_{21,t} \\ v_{3t} + v_{2t}\alpha_{32,t} + v_{1t}\alpha_{31,t} \\ v_{4t} + v_{3t}\alpha_{43,t} + v_{2t}\alpha_{42,t} + v_{1t}\alpha_{41,t} \end{pmatrix} = \begin{bmatrix} \left(\sigma_{1,t}\frac{1}{\lambda_{1t}}\right)^{1/2} \varepsilon_{1t} \\ \left(\sigma_{2,t}\frac{1}{\lambda_{2t}}\right)^{1/2} \varepsilon_{2t} \\ \left(\sigma_{3,t}\frac{1}{\lambda_{3t}}\right)^{1/2} \varepsilon_{3t} \\ \left(\sigma_{4,t}\frac{1}{\lambda_{4t}}\right)^{1/2} \varepsilon_{4t} \end{bmatrix} \quad (9)$$

The second, third and fourth lines gives the following system linear regressions:

$$\begin{aligned} v_{2t} &= -v_{1t}\alpha_{21,t} + (\sigma_{2,t}\varpi_{2,t})^{1/2} e_{2t} \\ v_{3t} &= -v_{2t}\alpha_{32,t} - v_{1t}\alpha_{31,t} + (\sigma_{3,t}\varpi_{3,t})^{1/2} e_{3t} \\ v_{4t} &= -v_{3t}\alpha_{43,t} - v_{2t}\alpha_{42,t} - v_{1t}\alpha_{41,t} + (\sigma_{4,t}\varpi_{4,t})^{1/2} e_{4t} \end{aligned} \quad (10)$$

where, conditional on $\lambda_{k,t}$ and $\sigma_{k,t}$, the parameters α 's have a Normal posterior and formulas for Bayesian linear regressions apply.

Drawing $G(B\backslash\Psi)$ Finally, the the posterior distribution of the VAR coefficients is linear and Gaussian, $G(B\backslash\Psi) \sim N(B_{T\backslash T}, P_{T\backslash T})$. We use the Kalman filter to estimate $B_{T\backslash T}$ and $P_{T\backslash T}$ where we account for the fact that the covariance matrix of the VAR residuals changes through time. The final iteration of the filter delivers $B_{T\backslash T}$ and $P_{T\backslash T}$.

2.1.3 Marginal Likelihood

For convenience, re-consider the main equations of the estimated model given by:

$$Y_t = c + \sum_{j=1}^P b_j Y_{t-j} + \Sigma_t^{1/2} e_t, \quad (11)$$

$$\Sigma_t = A^{-1} H_t A^{-1'} \quad (12)$$

$$H_t = \text{diag}(\sigma_{1t}^2 (1/\lambda_1) \dots) \quad (13)$$

Then Chib (1995)'s estimate of the marginal likelihood is based on the following identity:

$$\ln G(Y_t) = \ln F(Y_t \backslash \hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi) + \ln P(\hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi) - \ln H(\hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi) \quad (14)$$

where the subscript $\hat{\cdot}$ denotes the posterior mean, $F(\cdot)$ denotes the likelihood function, $P(\cdot)$ is the joint prior density, $H(\cdot)$ is the posterior distribution and Ξ denotes the state variables in the model. Equation 14 is simply the Bayes equation in logs re-arranged with the marginal likelihood $G(Y_t)$ on the LHS. Note that this equation holds at any value of the parameters, but is usually evaluated at high density points like the posterior mean. The joint prior density is straight forward to evaluate. The likelihood and posterior are more involved and described in appendix A.

2.1.4 Data

We use the dataset of Stock and Watson (2012) and focus on three key macroeconomic variables for the US: industrial production, inflation and the interest rate. In addition, we add the SP500 stock market index. The data is available at monthly frequency, spanning the period from January 1959 to September

2011. As a measure of output we use industrial production (Total Index). Inflation is calculated based on the personal consumption expenditure (chain-type) price index. Interest rate is measured as the 3-month Treasury Bill (secondary market) rate. Output growth, inflation and stock returns are calculated by taking the first difference of the logarithm of the series. The primary data source for all the four variables is the St. Louis Fed.

2.2 Alternative models

We consider three restricted versions of the benchmark BVAR model with stochastic volatility and fat tails. First, we assume that the orthogonalised shocks are Gaussian and consider a VAR model with stochastic volatility only. This model (VARSVOL) is defined as

$$Y_t = c + B_1 y_{t-1} + \dots + B_p y_{t-p} + u_t \quad (15)$$

where

$$\Sigma_t = A^{-1} H_t A^{-1'} \quad (16)$$

$$H_t = \text{diag}(\sigma_{1t}^2, \sigma_{2t}^2, \dots, \sigma_{nt}^2) \quad (17)$$

where $\ln \sigma_{k,t}^2$ follows the process defined in equation 3. In contrast, the second restricted model does not incorporate stochastic volatility but only assumes that the orthogonalised residuals follow an independent t distribution (TVAR). This model, therefore, is defined as

$$Y_t = c + B_1 y_{t-1} + \dots + B_p y_{t-p} + u_t \quad (18)$$

$$\Sigma_t = A^{-1} H_t A^{-1'} \quad (19)$$

$$H_t = \text{diag}\left(\sigma_1^2 \frac{1}{\lambda_{1t}}, \sigma_2^2 \frac{1}{\lambda_{2t}}, \dots, \sigma_n^2 \frac{1}{\lambda_{nt}}\right) \quad (20)$$

The final model considered is a standard BVAR. The estimation of these restricted models is carried out via Gibbs sampling using a simplification of the algorithm presented in section 2.1.2. The marginal likelihood for each of these alternative models is computed via the Chib (1995) algorithm.

3 Empirical Results

In this section we present results on the relative performance of each of the empirical models, both in terms of in-sample fit and recursive forecast performance. Before moving to model comparison, however, we present some of the key parameter estimates of the benchmark model over the full sample and compare them with some of the restricted models.

3.1 A summary of the posterior

3.1.1 Degrees of freedom

Figure 1 plots the estimated marginal posterior density of the degrees of freedom (DOF) from the TVAR-SVOL. Consider the estimates for the industrial production index. There is strong evidence that the orthogonalised shock associated with this variable is characterised by fat tails with the posterior density centered around 4 or 5 DOF. Similarly, the estimated posterior for the DOF associated with the orthogonalised residuals of the SP500 equation points towards non-normality. In contrast, the estimated posteriors for inflation and the T-Bill rate equations indicate DOF that are substantially higher. This suggests that the usual normality assumption is appropriate for the residuals associated with these equations. We show in the sensitivity analysis below that these results are robust to changing the ordering of the variables in the VAR.

The dotted red lines in figure 1 show the posterior density of the DOF from the TVAR model. It is interesting to note that when the VAR model does not incorporate stochastic volatility, the estimated

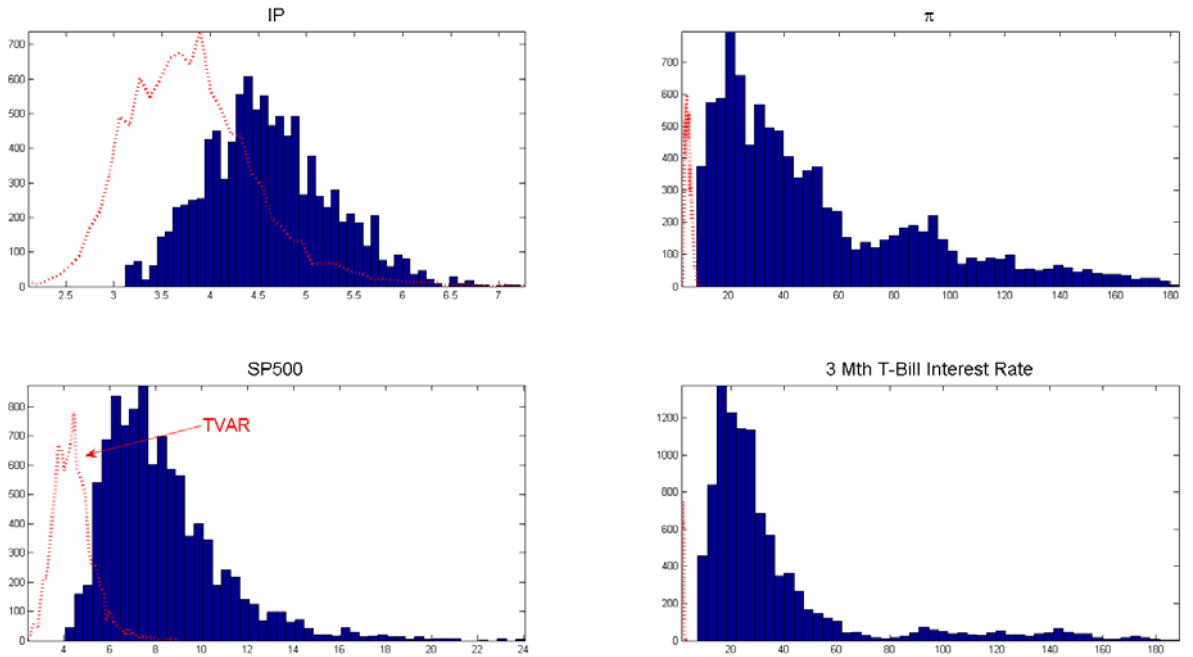


Figure 1: The posterior density of degrees of freedom from the benchmark model (TVAR-SVOL) and the TVAR.

posterior densities indicate stronger evidence in favour of fat tails for all four residuals. This confirms the argument in Curdia et al. (2013) that ignoring low frequency movements in volatility may bias the estimates of DOF downwards.

3.1.2 Stochastic volatility

Figure 2 plots the estimated stochastic volatility from the benchmark model and compares it with the estimate obtained from the VARSVOL model. Consider the top left panel of the figure. The estimated volatility of the IP shock from the benchmark model is estimated to be high until the early 1980s. It then declines smoothly and by 1985 is substantially lower than its pre-1985 average. There is some evidence of an increase in this volatility towards the end of the sample period. It is interesting to note that the estimated volatility of this shock from the model that does not account for the possibility of fat tails behaves very differently. The dotted black line shows that this estimate is more volatile indicating large fluctuations over the 1970s and the 1980s. While the decline in volatility in the early 1980s coincides across the two models, the VARSVOL model indicates a substantial increase in shock volatility that is missing from the benchmark estimate. Given that the shock to the IP equation displays fat tails (see figure 1), this difference highlights the fact that ignoring the possibility of non-normal disturbances can lead to very different interpretation of historical movements in volatility.

3.2 Model comparison

3.2.1 Marginal Likelihood

Table 1 lists the estimated log marginal likelihood for each model using the full sample. The marginal likelihood is estimated via the Chib (1995) method using 10,000 additional Gibbs iterations to estimate the components of the posterior density. It is clear from table 1 that the benchmark model displays the best in-

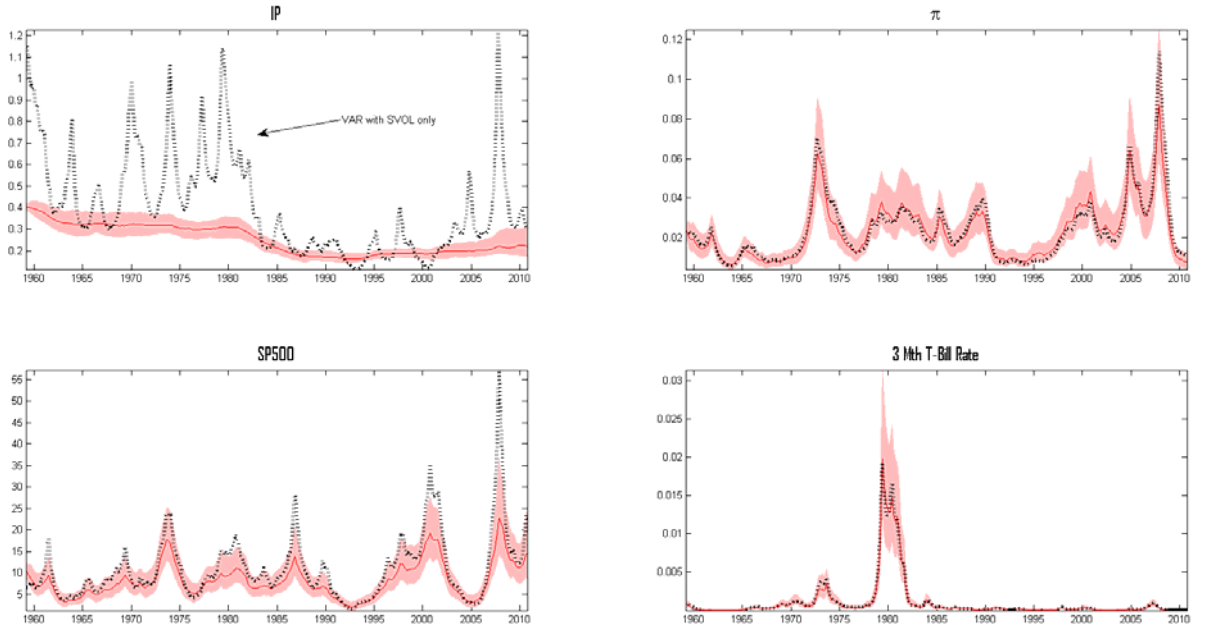


Figure 2: Stochastic Volatility from the TVARSVOL and the VARSVOL model.

Model	Log Marginal Likelihood
TVAR-SVOL	-1725.3503
VAR-SVOL	-1757.9607
TVAR	-2444.6197
BVAR	-2852.222

Table 1: Marginal Likelihood

sample fit while the BVAR has the lowest estimated marginal likelihood. Allowing for fat tails or stochastic volatility improves the fit. However, it is the combination of fat tails and stochastic volatility that delivers the best fitting specification. This indicates that both these features are crucial for the data we study.

3.2.2 Forecast performance

We compare the forecast performance of the four models considered above via a pseudo out of sample forecasting exercise. The four models are estimated recursively from January 1970 to September 2010. At each iteration, we construct the forecast density for the models:

$$P\left(\hat{Y}_{t+k} \setminus Y_t\right) = \int P\left(\hat{Y}_{t+k} \setminus Y_t, \Psi_{t+k}\right) P\left(\Psi_{t+k} \setminus \Psi_t, Y_t\right) P\left(\Psi_t \setminus Y_t\right) d\Psi \quad (21)$$

where $k=1,2,\dots,12$ and Ψ denotes the model parameters. The last term in equation 21 represents the posterior density of the parameters that is obtained via the MCMC simulation. The preceding two terms denote the forecast of the (time-varying) parameters and the data that can be obtained by simulation. The point forecast is obtained as the mean of the the forecast density. The recursive estimation delivers 490 forecast densities.

Table 2 presents the average root mean squared error (RMSE) for each model relative to that obtained using the BVAR. The table shows that it is difficult to distinguish between the models in terms of point

	1M	3M	6M	12M
IP				
TVAR-SVOL	0.904	0.914	0.927	0.945
TVAR	0.899	0.906	0.921	0.950
VARSVOL	0.903	0.909	0.925	0.946
π				
TVAR-SVOL	1.011	1.027	1.048	1.065
TVAR	0.994	1.011	1.031	1.050
VARSVOL	0.993	1.011	1.027	1.041
SP500				
TVAR-SVOL	0.951	0.953	0.959	0.974
TVAR	0.950	0.951	0.956	0.970
VARSVOL	0.956	0.955	0.958	0.971
R				
TVAR-SVOL	0.935	0.885	0.879	0.912
TVAR	0.930	0.886	0.888	0.928
VARSVOL	0.940	0.896	0.881	0.908

Table 2: RMSE relative to the Bayesian VAR model

forecasts. For variables such as industrial production, the interest rate and the stock price index, each of the three models produce forecasts that lead to a 5% to 10% reduction in RMSE relative to the BVAR. For inflation, the point forecast performance of the models under consideration is very similar to that of the BVAR.

In the section below we focus on *density* forecast comparison. The density forecasts are evaluated using probability integral transforms (PIT) and log scores (LS). The former are calculated as

$$PIT_t = \Phi(Y_{t+k}) \quad (22)$$

where $\Phi(Y_{t+k})$ denotes the CDF associated with the forecast density evaluated at the realised data. Note that if the forecast density equals the true density then PIT_t are distributed uniformly over $(0, 1)$. At the one step horizon, PIT_t are also independently distributed, while independence may be violated at longer horizons due to serial dependence in multi step forecasts.

In addition to the PITs we consider the log scores. These are defined as

$$LS_t = \ln P(Y_{t+k})$$

where $P(Y_{t+k})$ denotes the forecast density evaluated at the realised data. A higher value for LS_t suggests a more accurate density forecast. Note that we employ kernel methods to estimate the density and distribution function of the forecasts. This enables us to account for any potential non-linearities in the forecast distribution.

PIT comparison

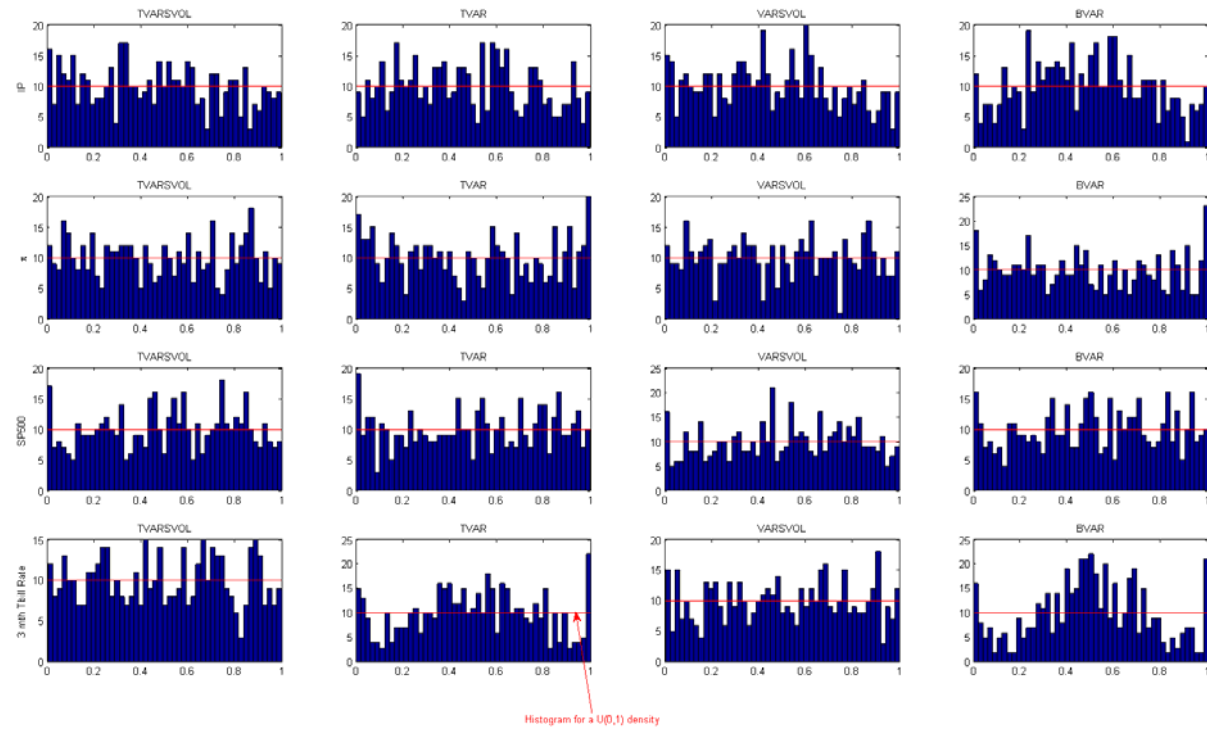


Figure 3: PIT histograms at the one month forecast horizon.

Figure 3 plots the histogram of the estimated PITs and the implied histogram from a uniform distribution and provides a visual assessment of density calibration. In figure 3 we consider the PITs for the one month ahead forecast. The results are very similar at other horizons and available on request.

Consider the estimates for industrial production. The PIT histogram produced by the BVAR model appears hump shaped with mass concentrated over the 0.4 to 0.6 interval indicating departures from uniformity. In contrast, the PIT histogram from the TVARSVOL model is closer to the uniform distribution. Similarly, the distribution of the PITs from the TVAR model appears to approximate a uniform distribution. The histogram from VARSVOL model, on the other hand, shows mass on the left tail and appears to differ from a uniform distribution. The forecast density of inflation from the benchmark model and the VARSVOL model appear to be better calibrated than the estimates from the TVAR and the BVAR, with the histograms from the latter displaying mass at the tails. For the SP500, it is difficult to distinguish between the PIT histograms across models, with the estimates displaying mass at the tails. Finally, for the 3mth tbill yield, the TVARSVOL and the VARSVOL model appear to perform better, the the PIT distributions from the BVAR and the TVAR model displaying a hump shape.

Overall, the PIT distributions provide some visual evidence that both fat tails and stochastic volatility are important for obtaining a well calibrated forecast density for variables such as industrial production.

	1M	3M	6M	12M
IP				
TVAR-SVOL	25.328	28.941	29.131	11.427
TVAR	27.356	32.937	35.005	30.937
VARSVOL	25.154	-13.997	-23.766	4.700
π				
TVAR-SVOL	36.352	59.407	62.374	61.425
TVAR	34.845	52.010	46.987	51.887
VARSVOL	24.869	74.385	36.016	15.308
SP500				
TVAR-SVOL	28.294	31.653	8.082	-21.550
TVAR	23.196	23.189	18.207	19.572
VARSVOL	32.420	31.181	17.647	-9.359
R				
TVAR-SVOL	155.022	177.246	33.412	6.605
TVAR	85.714	44.122	16.813	7.378
VARSVOL	153.714	175.536	34.522	10.032

Table 3: Percentage improvement in log scores over the Bayesian VAR model

Log score comparison Table 3 considers the log score for each model relative to that obtained via the BVAR. The table presents the average estimates across the forecasting sample, with a positive number indicating an improvement over the BVAR model. Consider the results for industrial production. At the 1 month horizon, allowing for fat tails or stochastic volatility leads to a similar improvement over the BVAR density forecasts. This is not the case at longer horizons where fat tails are clearly important. At the 6 month horizon, the TVAR offers a 35% improvement over the BVAR log score. In contrast, the VARSVOL model performs worse than the BVAR. Therefore, it appears that allowing for t-distributed shocks is crucial for industrial production at policy relevant forecasting horizons. The results for SP500 are similar. At the 6 month and the 1 year horizon, the TVAR model outperforms the other models, highlighting the role of fat tails.

For inflation and interest rates, both stochastic volatility and fat tails appear to be important. The TVAR-SVOL model produces the largest improvement over the BVAR for inflation at the 6 and the 12

month horizon. At the 1 and the 3 month horizon, the benchmark model produces the best performance, with the VARSVOL model delivering the largest improvement over the BVAR at longer horizons.

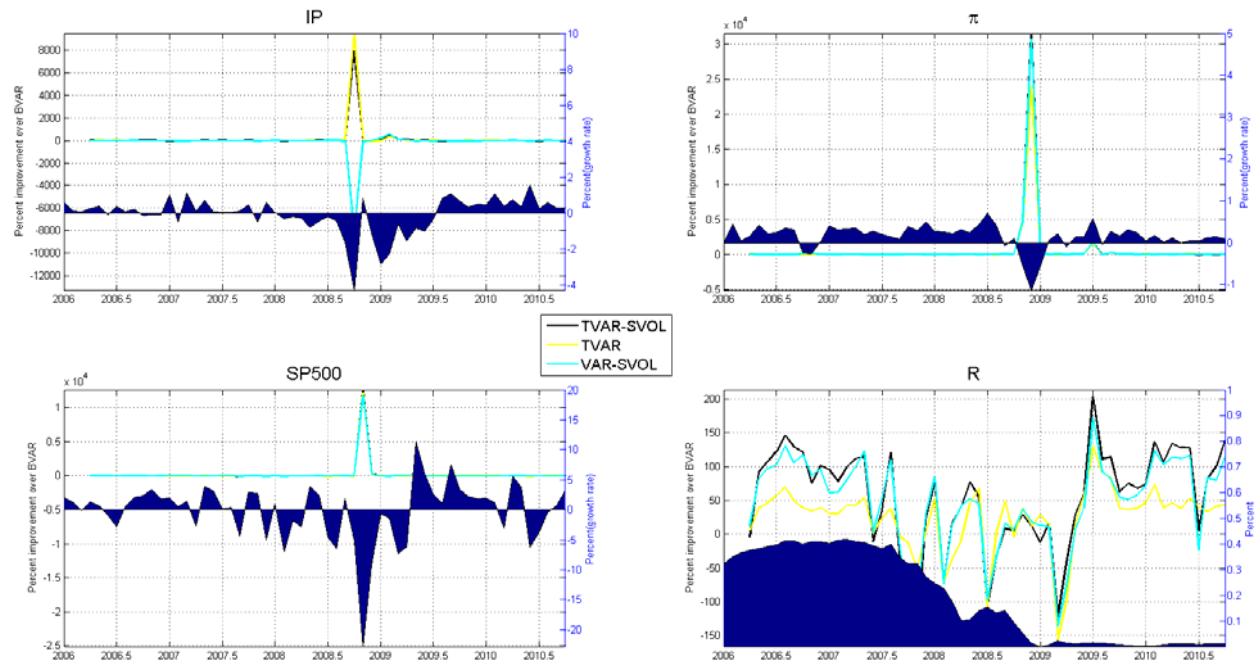


Figure 4: Log scores (3 month horizon) relative to those from the BVAR model over the recent financial crisis.

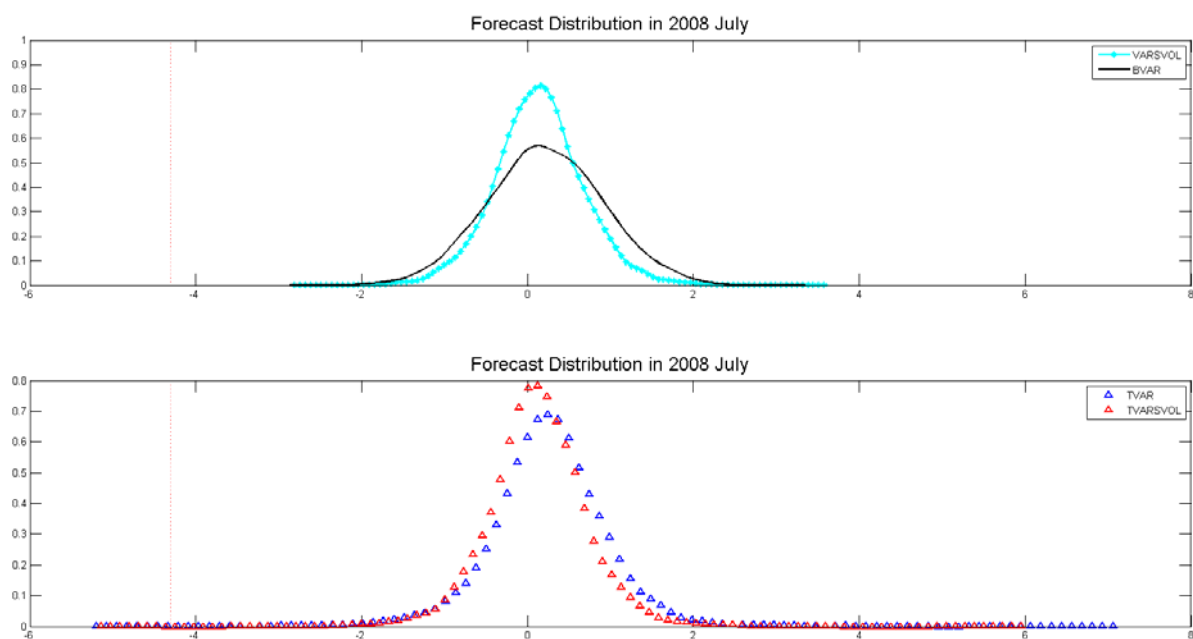


Figure 5: 3 month ahead forecast distribution for industrial production and actual out-turn in September 2008 (red vertical line).

In figure 4 we consider the evolution of log scores over the recent financial crisis. The left axis in each panel shows the percentage improvement in log scores over the BVAR model. In this figure we consider the 3 month forecasting horizon but the results are similar at other horizons. The right axis shows the actual data for the variable under consideration which is plotted as an area chart.

The top left panel shows the results for industrial production. The performance of the three models is similar before the onset of the deep recession at the end of 2008. The large decline in industrial production coincided with a very large divergence in the performance of models with and without fat tails. The TVAR and the TVARSVOL model show a huge improvement in the log score. In contrast, the accuracy of the VARSVOL model deteriorates substantially relative to the BVAR model. The reason for this divergence in forecast performance is immediately clear from figure 5 which shows the 3 month ahead forecast density of industrial production for September 2008 from the four models together with the actual out-turn in that month. The left tails of the densities from the BVAR and the VARSVOL model do not include the actual industrial production out-turn of -4.23%. In contrast, the densities from the models with fat tails cover this eventuality. This highlights the fact that the assumption of normality may lead to one to ignore the possibility of large movements in the data as seen in the recent financial crisis. It is interesting to note that the performance of the three models was similar during the second dip in industrial production seen in December 2008 and January 2009. This is because the 2% fall during this episode was accounted for by the forecast densities from all models.

For the stock price index and inflation, both stochastic volatility and t-disturbances appear to be important, with the TVARSVOL model showing a large improvement during late 2008 and early 2009. The performance of these models was more mixed for the interest rate over the initial cutting phase of 2007 and 2008. However, stochastic volatility appears to be important over the post 2009 period that was characterised by persistently low interest rates.

	1M	3M	6M	12M
IP	23.184	20.395	30.253	30.544
π	38.261	62.279	40.970	60.381
SP500	32.287	31.482	16.459	-4.231
R	154.945	177.268	33.664	7.591

Table 4: Percentage improvement in log scores from the TVARSVOL over the Bayesian VAR model (Sensitivity analysis)

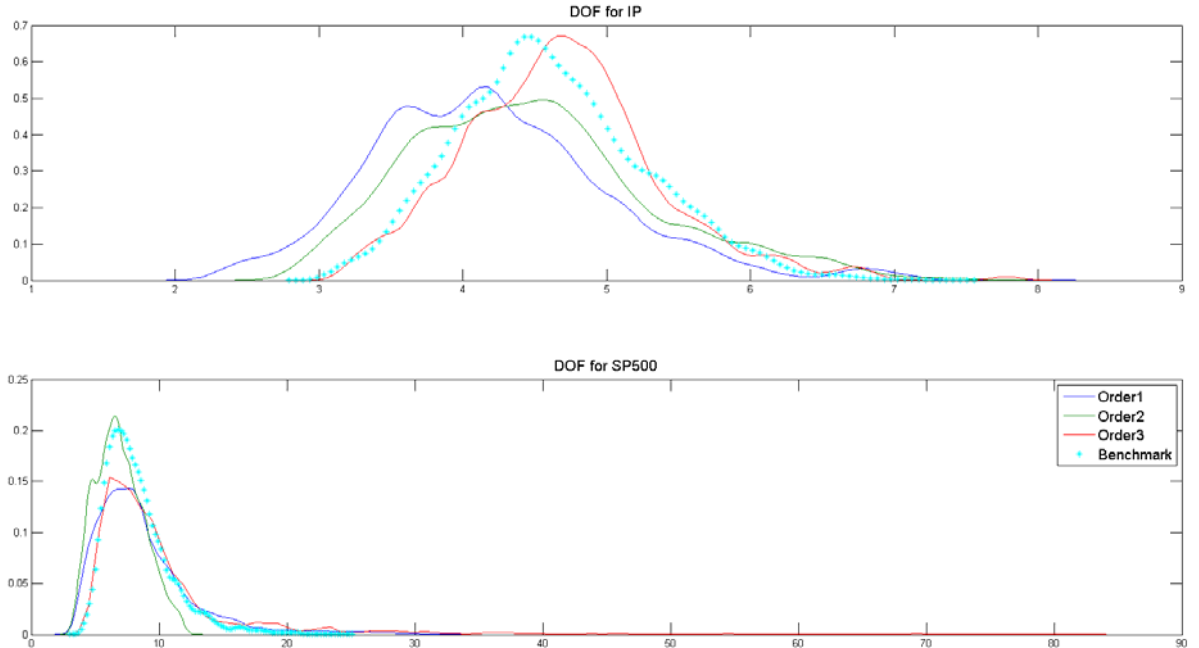


Figure 6: Sensitivity of the DOF posterior to alternative orderings

3.2.3 Sensitivity Analysis

In table 4 we present the log scores (relative to the BVAR) from a version of the benchmark model that uses an alternative prior. This version of the model the prior for the degrees of freedom parameter is chosen so that a higher weight is given to the possibility of normality. In particular, we use the prior $p(v_{\lambda,n}) \sim \Gamma(v_0, 2)$ where $v_0 = 50$. A comparison of table 4 with the results presented in table 3 indicates that for industrial production and stock market returns (the variables for which the orthogonalised errors displayed the most evidence for non-gaussianity), the average relative log scores are fairly similar to the benchmark case. This provides some evidence that the key results do not depend on the benchmark prior.

Figure 6 presents the marginal posterior for the DOF for the industrial production and SP500 returns using alternative orderings for these variables in the TVARSVOL model. For example, while in the benchmark case IP is ordered first, ‘order1’, ‘order2’ and ‘order3’ in figure 6 refer to versions of the model where IP is ordered second, third and fourth respectively. Similarly, SP500 is ordered first, second and fourth in these alternative models. It is clear from the top panel of the figure that the strong evidence for non-normality of the orthogonal residuals of the IP equation is not influenced by the recursive structure of the A matrix in equation 2. The bottom panel of the figure suggests a similar conclusion for SP500. While there is a rightward skew in the marginal density when SP500 is ordered last, the posterior is centered around a value of DOF less than 10 in all cases.

4 Conclusions

This paper introduces a BVAR model that incorporates stochastic volatility and fat tailed disturbances. We show that this model fits a monthly US dataset better than alternatives that do not include these features. The estimates of the model suggest strong evidence that disturbances to industrial production and stock market returns are non-normal. Incorporating this non-normality in the model leads substantial improvements in the accuracy of forecast densities. In particular BVARs with Gaussian disturbances fail to attach any probability to low values of industrial production seen in late 2008. These results highlight the importance of incorporating the possibility of fat tails in forecasting models.

A Computation of the Marginal Density

A.1 Likelihood

The likelihood function of the model is calculated using a particle filter using 10,000 particles. We re-write the model in state space form

$$\begin{aligned} X_t &= H\Gamma_t \\ \Gamma_t &= \mu + F\Gamma_{t-1} + Q_t^{1/2}\varepsilon_t \\ \ln q_{Kt} &= \ln q_{Kt-1} + v_t \end{aligned}$$

where $\varepsilon_t = \{\varepsilon_{1t}, \dots, \varepsilon_{Nt}\}$ with ε_{Kt} follows a Student-t density with v_K degrees of freedom and q_{Kt} denotes the diagonal elements of Q . X_t is observed data, while $\Phi_t = (\Gamma_t, q_{Kt})$ are the state-variables. Given the non-normal disturbances, the Kalman filter cannot be employed.

Consider the distribution of the state variables in the model denoted Φ_t conditional on information up to time t (denoted by z_t)

$$f(\Phi_t \setminus z_t) = \frac{f(X_t, \Phi_t \setminus z_{t-1})}{f(X_t \setminus z_{t-1})} = \frac{f(X_t \setminus \Phi_t, z_{t-1}) \times f(\Phi_t \setminus z_{t-1})}{f(X_t \setminus z_{t-1})} \quad (23)$$

Equation 23 says that this density can be written as the ratio of the joint density of the data and the states $f(X_t, \Phi_t \setminus z_{t-1}) = f(X_t \setminus \Phi_t, z_{t-1}) \times f(\Phi_t \setminus z_{t-1})$ and the likelihood function $f(X_t \setminus z_{t-1})$ where the latter is defined as

$$f(X_t \setminus z_{t-1}) = \int f(X_t \setminus \Phi_t, z_{t-1}) \times f(\Phi_t \setminus z_{t-1}) d\Phi_t \quad (24)$$

Note also that the conditional density $f(\Phi_t \setminus z_{t-1})$ can be written as

$$f(\Phi_t \setminus z_{t-1}) = \int f(\Phi_t \setminus \Phi_{t-1}) \times f(\Phi_{t-1} \setminus z_{t-1}) d\Phi_{t-1} \quad (25)$$

These equations suggest the following filtering algorithm to compute the likelihood function:

1. Given a starting value $f(\Phi_0 \setminus z_0)$ calculate the predicted value of the state

$$f(\Phi_1 \setminus z_0) = \int f(\Phi_1 \setminus \Phi_0) \times f(\Phi_0 \setminus z_0) d\Phi_0$$

2. Update the value of the state variables based on information contained in the data

$$f(\Phi_1 \setminus z_1) = \frac{f(X_1 \setminus \Phi_1, z_0) \times f(\Phi_1 \setminus z_0)}{f(X_1 \setminus z_0)}$$

where $f(X_1 \setminus z_0) = \int f(X_1 \setminus \Phi_1, z_0) \times f(\Phi_1 \setminus z_0) d\Phi_1$ is the likelihood for observation 1. By repeating these two steps for observations $t = 1 \dots T$ the likelihood function of the model can be calculated as $\ln lik = \ln f(X_1 \setminus z_0) + \ln f(X_2 \setminus z_1) + \dots + \ln f(X_T \setminus z_{t-1})$

In general, this algorithm is inoperable because the integrals in the equations above are difficult to evaluate. The particle filter makes the algorithm feasible by using a Monte-Carlo method to evaluate these integrals. In particular, the particle filter approximates the conditional distribution $f(\Phi_1 \setminus z_0)$ via M draws or particles from the Student-t density using the transition equation of the model. For each draw for the state variables the conditional likelihood $W^m = f(X_1 \setminus z_0)$ is evaluated. Conditional on the draw for the state variables, the predicted value for the variables \hat{X}_{i1}^M can be computed using the observation equation and the prediction error decomposition is used to evaluate the likelihood W^m . Note that as the predictive density is degenerate in this model, we need to add measurement error. The update step involves a draw from the density $f(\Phi_1 \setminus z_1)$. This is done by sampling with replacement from the sequence of particles with the re-sampling probability given by $\frac{W^m}{\sum_{m=1}^M W^m}$. This re-sampling step updates the draws for Φ based on information contained in the data for that time period. By the law of large numbers the likelihood function for the observation can be approximated as $\ln lik_t = \ln \frac{\sum_{m=1}^M W^m}{M}$.

A.2 Evaluation of the posterior density $H(\cdot)$

Consider the following decomposition

$$H\left(\hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, v_\lambda, \Xi\right) = H\left(\hat{B} \setminus \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) \times H\left(\hat{A} \setminus \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) \times H\left(\hat{g} \setminus \hat{\lambda}, \hat{v}_\lambda, \Xi\right) \quad (26)$$

$$\times H\left(\hat{\lambda} \setminus \hat{v}_\lambda, \Xi\right) \times H\left(\hat{v}_\lambda, \Xi\right) \quad (27)$$

Each term can be evaluated directly or by using a further MCMC run.

1. $H\left(\hat{B} \setminus \hat{A}, \hat{\sigma}^2, \hat{\lambda}, v_\lambda, \Xi\right)$. This a complete conditional density with a known form. This density is Normal with mean and variance that can be calculated via the Kalman filter. The evaluation is done via an additional Gibbs sampler that draws from (a) $H\left(B_i \setminus \hat{A}, \hat{\sigma}^2, \hat{\lambda}, v_\lambda, \Xi_j\right)$ and (b) $H\left(\Xi_i \setminus \hat{A}, \hat{\sigma}^2, \hat{\lambda}, v_\lambda, B_i\right)$.

After a burn in period $H\left(\hat{B} \setminus \hat{A}, \hat{\sigma}^2, \hat{\lambda}, v_\lambda, \Xi\right)$

2. $H\left(\hat{A} \setminus \hat{\sigma}^2, \hat{\lambda}, v_\lambda, \Xi\right) = \int H\left(\hat{A} \setminus \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, B, \Xi\right) \times H\left(B \setminus \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) dB$. This term can be approximated using an additional Gibbs run that samples from the following conditionals with the current and previous draws indexed by i and j (a) $H\left(A_i \setminus \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, B_j, \Xi_j\right)$ and (b) $H\left(B_i \setminus \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, A_j, \Xi_j\right)$ and (c) $H\left(\Xi_i \setminus B_j, \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, A_j\right)$. After a burn in period $H\left(\hat{A} \setminus \hat{\sigma}^2, \hat{\lambda}, v_\lambda, \Xi\right) \approx \frac{1}{J} \sum_{i=1}^J H\left(\hat{A} \setminus \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, B_i, \Xi_j\right)$ where $H\left(\hat{A} \setminus \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, B_i, \Xi_j\right)$ is the Normal density described above.

3. $H\left(\hat{g} \setminus \hat{\lambda}, \hat{v}_\lambda, \Xi\right) = \int \left[\int H\left(\hat{g} \setminus \hat{\lambda}, \hat{v}_\lambda, \hat{B}, \hat{A}, \Xi\right) \times H\left(\hat{A} \setminus \hat{\lambda}, \hat{v}_\lambda, \hat{B}, \Xi\right) dA \right] \times H\left(\hat{B} \setminus \hat{\lambda}, \hat{v}_\lambda, \Xi\right) d\hat{B}$. This term

can be approximated using an additional Gibbs run that samples from the following conditionals (a) $H\left(g_i \setminus \hat{\lambda}, \hat{v}_\lambda, B_j, A_j, \Xi_j\right)$ (b) $H\left(B_i \setminus g_j, \hat{\lambda}, \hat{v}_\lambda, A_j, \Xi_j\right)$ (c) $H\left(A_i \setminus B_j, g_j, \hat{\lambda}, \hat{v}_\lambda, \Xi_j\right)$ and (d) $H\left(\Xi_i \setminus A_j, B_j, g_j, \hat{\lambda}, \hat{v}_\lambda\right)$.

After a burn-in period $H\left(\hat{g} \setminus \hat{\lambda}, \hat{v}_\lambda, \Xi\right) \approx \frac{1}{J} \sum_{i=1}^J H\left(g_i \setminus \hat{\lambda}, \hat{v}_\lambda, B_j, A_j, \Xi_j\right)$ where this is an inverse Gamma pdf.

4. $H\left(\hat{\lambda} \setminus \hat{v}_\lambda, \Xi\right)$. As in 3 above, this term can be approximated by a Gibbs run that draws from the following densities (a) $H\left(\lambda_i \setminus g_j, \hat{v}_\lambda, B_j, A_j, \Xi_j\right)$ (b) $H\left(g_i \setminus \lambda_j, \hat{v}_\lambda, B_j, A_j, \Xi_j\right)$ (c) $H\left(B_i \setminus g_j, \lambda_j, \hat{v}_\lambda, A_j, \Xi_j\right)$ (d) $H\left(A_i \setminus B_j, g_j, \lambda_j, \hat{v}_\lambda, \Xi_j\right)$ and (e) $H\left(\Xi_i \setminus A_i, B_j, g_j, \lambda_j, \hat{v}_\lambda\right)$ After a burn-in period $H\left(\hat{\lambda} \setminus \hat{v}_\lambda\right) \approx \frac{1}{J} H\left(\hat{\lambda} \setminus \sigma_j^2, \hat{v}_\lambda, B_j, A_j\right)$, where this is a Gamma pdf.

5. The final term $H\left(\hat{v}_\lambda\right)$ is an unknown density. Therefore the algorithm Chib and Jeliazkov (2001) of is required. They show that this density can be approximated as

$$H\left(\hat{v}_\lambda\right) = \frac{E_1\left(\alpha\left(v_\lambda, \hat{v}_\lambda \setminus B, A, \sigma^2, \lambda, \Xi\right) q\left(v_\lambda, \hat{v}_\lambda \setminus B, A, \sigma^2, \lambda, \Xi\right)\right)}{E_2\left(\alpha\left(\hat{v}_\lambda, v_\lambda^j \setminus B, A, \sigma^2, \lambda, \Xi\right)\right)}$$

where $\alpha\left(v_\lambda^{old}, v_\lambda^{new}\right)$ denotes the acceptance probability of Metropolis move from v_λ^{old} to v_λ^{new} and $q\left(v_\lambda^{old}, v_\lambda^{new}\right)$ is the candidate density. The numerator term can be approximated by averaging the quantity from the main MCMC run: $\alpha\left(v_\lambda^j, \hat{v}_\lambda \setminus B, A, \sigma^2, \lambda, \Xi\right) q\left(v_\lambda^j, \hat{v}_\lambda \setminus B, A, \sigma^2, \lambda, \Xi\right)$ where j indexes the MCMC draws. The denominator term requires an additional Gibbs sampler as $\alpha\left(\hat{v}_\lambda, v_\lambda^j \setminus B, A, \sigma^2, \lambda, \Xi\right)$ is conditioned on the posterior mean \hat{v}_λ . This sampler draws from each posterior density conditioned on \hat{v}_λ and then draws from the candidate density $v_\lambda^j \sim q\left(\hat{v}_\lambda, v_\lambda \setminus B, A, \sigma^2, \lambda\right)$. The average acceptance probability from produces an estimate of the denominator.

References

- Ascari, G., G. Fagiolo, and A. Roventini (2012, January). Fat-tail distributions and business-cycle models. LEM Papers Series 2012/02, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy.
- Banbura, M., D. Giannone, and L. Reichlin (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics* 25(1), 71–92.
- Chahad, M. and F. Ferroni (2014, August). Structural var, rare events and the transmission of credit risk in the euro area. Mimeo, Bank of France.
- Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association* 90(423).
- Chib, S. and I. Jeliazkov (2001, March). Marginal likelihood from the metropolis-hastings output. *Journal of the American Statistical Association* 96, 270–281.
- Chib, S. and S. Ramamurthy (2014). Dsge models with student-t errors. *Econometric Reviews* 33(1-4), 152–171.
- Cogley, T. and T. J. Sargent (2005, April). Drift and volatilities: Monetary policies and outcomes in the post wwii u.s. *Review of Economic Dynamics* 8(2), 262–302.
- Curdia, V., M. D. Negro, and D. L. Greenwald (2013). Rare shocks, great recessions. Technical report.
- Fernandez-Villaverde, J. and J. F. Rubio-Ramirez (2007). Estimating macroeconomic models: A likelihood approach. *Review of Economic Studies* 74(4), 1059–1087.
- Geweke, J. (1992). Priors for macroeconomic time series and their application. Technical report.
- Geweke, J. (1993, Suppl. De). Bayesian treatment of the independent student- t linear model. *Journal of Applied Econometrics* 8(S), S19–40.
- Geweke, J. (1994, August). Priors for macroeconomic time series and their application. *Econometric Theory* 10(3-4), 609–632.
- Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. Wiley.
- Jacquier, E., N. G. Polson, and P. E. Rossi (1994, October). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics* 12(4), 371–89.
- Jacquier, E., N. G. Polson, and P. E. Rossi (2004, September). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122(1), 185–212.
- Justiniano, A. and G. E. Primiceri (2008, June). The time-varying volatility of macroeconomic fluctuations. *American Economic Review* 98(3), 604–41.
- Koop, G. (2003). *Bayesian Econometrics*. Wiley.
- Liu, Z., D. F. Waggoner, and T. Zha (2011). Sources of macroeconomic fluctuations: A regime-switching dsge approach. *Quantitative Economics* 2(2), 251–301.
- Mishkin, F. S. (2011, February). Monetary policy strategy: Lessons from the crisis. NBER Working Papers 16755, National Bureau of Economic Research, Inc.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies* 72(3), 821–852.

Stock, J. H. and M. W. Watson (2012, May). Disentangling the channels of the 2007-2009 recession. NBER Working Papers 18094, National Bureau of Economic Research, Inc.

Student (1908). The probable error of a mean. *Biometrika* 6(1), pp. 1–25.

Zellner, A. (1976). Bayesian and non-bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association* 71(354), pp. 400–405.

**This working paper has been produced by
the School of Economics and Finance at
Queen Mary, University of London**

**Copyright © 2014 Ching-Wai (Jeremy) Chiu, Haroon Mumtaz
and Gabor Pinter. All rights reserved**

**School of Economics and Finance
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 7356
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm**