Forecasting with VAR Models:
Fat Tails and Stochastic Volatility

Ching-Wai (Jeremy) Chiu   Haroon Mumtaz   Gabor Pinter[*]

January 6, 2017

Abstract

We provide evidence that modelling both fat tails and stochastic volatility are important in improving in-sample fit and out-of-sample forecasting performance. To show this, we construct a VAR model where the orthogonalised shocks feature Student's t distribution as well as time-varying variance. We estimate the model using US data on industrial production growth, inflation, interest rates and stock returns. In terms of in-sample fit, the VAR model featuring both stochastic volatility and t-distributed disturbances outperforms restricted alternatives that feature either attributes. The VAR model with t disturbances results in density forecasts for industrial production and stock returns that are superior to alternatives that assume Gaussianity, and this difference is especially stark over the recent Great Recession. Further international evidence confirms that accounting for both stochastic volatility and Student's t-distributed disturbances may lead to improved forecast accuracy.

Key words: Bayesian VAR, Fat-tails, Stochastic volatility, Great Recession
(*JEL:* C11, C32, C52)

1   Introduction

Could empirical macroeconomic models with a more realistic shock distribution be able to better predict economic downturns? Since the Great Recession and during the ensuing uncertainty surrounding the political and economic environment, both academic and policy circles have paid increasing attention to fat tail events. Many argue that recent events could hardly be explained or predicted by models that are based on a Gaussian shock structure, mainly because these models assign virtually zero probability to the macroeconomic outcomes that we have recently observed.[1] This has been recognised by recent efforts of the DSGE literature including Curdia, del Negro, and Greenwald (2014) 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 paper contributes to the literature by empirically investigating the in-sample fit and out-of-sample forecasting performance of a VAR model incorporated with Student's t errors (Student, 1908) and stochastic volatility (TVARSVOL). Building on the previous work on univariate (Geweke, 1992, 1993, 1994) and multivariate (Ni and Sun, 2005) models with Student's t-distributed shocks, as well as work on the DSGE literature (Fernandez-Villaverde and Rubio-Ramirez, 2007; Justiniano and Primiceri, 2008; Liu, Waggoner, and Zha, 2011) on stochastic volatility of the error structure, we provide a Gibbs sampling algorithm to estimate the TVARSVOL model. Moreover, we apply the particle filter to compute the marginal likelihood, and compare the in-sample fit and the out-of-sample forecasting performance of this model against three other models, namely, a linear Gaussian BVAR model (BVAR), a VAR model with Student's t error (TVAR) and a VAR model with stochastic volatility (VARSVOL).

We show that incorporating both fat tails and stochastic volatility can be important in improving in-sample fit and out-of-sample forecasting performance. Using monthly data on industrial production growth, inflation rate, short-term interest rate and the SP500 return for the US, the TVARSVOL model outperforms the other three models in terms of in-sample fit. When it comes to out-of-sample forecasting, we present international evidence that VAR models with Student's t-distributed shocks result in density forecasts for industrial production and stock returns being superior to alternatives that assume Normality.

Our results have at least two important implications when interpreting historical data. First, the structural shift in industrial production volatility in the early 1980s, often referred to as the Great Moderation, may be overestimated when the VAR model does not account for Student's t-disturbances. Second, the Student's t assumption appears especially important over the 2008 and 2009 period. Forecast densities for industrial production generated from VARs with Gaussian disturbances assign a negligible probability to the collapse of industrial production observed in late 2008. In contrast, when Student's t shocks are incorporated, the left tail of the forecast density includes the actual outcome.

Our paper is related to the DSGE analysis of Curdia, del Negro, and Greenwald (2014) who

---

[1]These issues have been discussed in more detail by Mishkin (2011); Elliott and Timmermann (2013); Ng and Wright (2013) amongst many others.

show that by solely focusing on fat-tails and ignoring lower-frequency changes in the volatility of shocks (as in Ascari, Fagiolo, and Roventini (2015)) tends to bias the results towards finding evidence in favour of fat tails. Our work is also related to Clark and Ravazzolo (2015) who work with (V)AR models using quarterly real-time data (GDP growth rate, inflation rate, unemployment rate and short-term interest rate) of the US. They find empirical evidence that models with stochastic volatility increase the accuracy of both point and density forecasts relative to models assuming homoscedasticity. Our paper considers monthly data-sets incorporating with both real and financial variables from the US and three other developed countries, and provide evidence that modelling fat-tailed errors on top of stochastic volatility is important in improving forecasting performance.

The structure of the paper is as follows. Section 2 provides a description of the 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, compares the models based on in-sample fit and forecasting performance, and provides sensitivity analysis. Section 4 provides further international evidence on the forecasting performance of the different models estimated on data from Canada, Germany and the UK. Section 5 concludes.

2 The Model

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 TVARSVOL). 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.[2] Introducing Student's t-distribution in the shock structure is meant to capture high-frequency changes in volatility that are often of extreme magnitudes, hence potentially providing an effective treatment of outliers and extreme events.[3] 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:

$$Y_t = c + B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t \qquad t = 1, \ldots, T, \tag{2.1}$$

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, $i = 1, \ldots, p$ 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

---

[2]See Uhlig (1997) and Primiceri (2005) amongst many others.

[3]In an important paper, Jacquier, Polson, and Rossi (2004) provides a detailed analysis of this issue in a univariate framework. Also see Gerlach, Carter, and Kohn (2000) for a rich discussion of outliers in a Bayesian context, and Bauwens, Koop, Korobilis, and Rombouts (2015) for exploring the role of structural changes in affecting Bayesian forecasts.

is defined as:

$$cov(u_t) = \Sigma_t = A^{-1} H_t A^{-1'}, \tag{2.2}$$

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

$$\ln \sigma_{k,t} = \ln \sigma_{k,t-1} + s_{k,t}, var(s_{k,t}) = g_k, \tag{2.3}$$

for $k = 1, 2, ..., n$. In line with Geweke (1993), the weights $[\lambda_{1,t}, \lambda_{2,t}, ..., \lambda_{n,t}]$ are indexed by time t because they are to capture any high-frequency movements in volatility over time, as opposed to the low-frequency movements in volatility which are in turn captured by $[\sigma_{1,t}, \sigma_{2,t}, ..., \sigma_{n,t}]$. As shown by Geweke (1993), assuming a Gamma prior for $\lambda_{k,t}$ of the form $p(\lambda_k) = \prod\limits_{t=1}^{T} p(\lambda_{k,t}) = \prod\limits_{t=1}^{T} \tilde{\Gamma}(1, v_{\lambda,k})$ leads to a scale mixture of normals for the orthogonal residuals $\tilde{\varepsilon}_t = Au_t$ where $\tilde{\varepsilon}_t = \{\tilde{\varepsilon}_{1,t}, \tilde{\varepsilon}_{2,t}, ..\tilde{\varepsilon}_{n,t}\}$ and $cov(\tilde{\varepsilon}_t) = H_t$. Note that $\tilde{\Gamma}(a, b)$ denotes a Gamma density with mean a and degrees of freedom b.[4] The above formulation is equivalent to a specification that assumes Student's 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 2.3.

There are two noteworthy things about the BVAR model. First, as discussed earlier, 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's t-distribution to be independent across equations and simplifies the estimation algorithm.[5] 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.

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 Banbura, Giannone, and Reichlin (2010) in implementing the dummy observation approach (Doan, Litterman, and Sims, 1983;

---

[4]The probability density function for the Gamma distribution is:

$$\tilde{\Gamma}(\mu, v) = \begin{cases} \left[\left(\frac{2\mu}{\nu}\right)^{\nu/2} \Gamma\left(\frac{\nu}{2}\right)\right]^{-1} y^{\frac{\nu-2}{2}} \exp\left(-\frac{y\nu}{2\mu}\right) & if \ 0 < y < \infty \\ 0 & otherwise \end{cases}$$

Under this parametrisation, the $\tilde{\Gamma}(v_0, 2)$ distribution is an exponential distribution with mean v_0.

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

Sims and Zha, 1998).[6] We follow Geweke (1993) in setting a hierarchical prior on the parameter controlling the degree of freedom of the Student's t distributions $v_{\lambda,k}$ and the weighting vector $\lambda_{k,t}$, for $k = 1, ..., n$:

$$p(v_{\lambda,k}) \sim \tilde{\Gamma}(v_0, 2) \tag{2.4}$$

$$p(\lambda_{k,t}) \sim \tilde{\Gamma}(1, v_{\lambda,k}), \tag{2.5}$$

In our TVARSVOL model, 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 2.3, and propose an inverse-gamma distribution, $p(g_k) \sim IG(D_0, T_0)$, where $T_0 = 1$ and $D_0 = 0.0001$ 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 Sampler and Computing the Marginal Likelihood

The Gibbs algorithm for the TVARSVOL model cycles through six conditional posterior distributions. The details of each conditional posterior density are found Section A of the Appendix.

To compute the marginal likelihood, re-consider the main equations of the estimated model given by:

$$Y_t = c + \sum_{j=1}^{p} B_j Y_{t-j} + \Sigma_t^{1/2} \varepsilon_t, \tag{2.6}$$

$$\Sigma_t = A^{-1} H_t A^{-1'} \tag{2.7}$$

$$H_t = diag\left(\sigma_{1,t}^2 \frac{1}{\lambda_{1,t}}, \sigma_{2,t}^2 \frac{1}{\lambda_{2,t}}, ..., \sigma_{n,t}^2 \frac{1}{\lambda_{n,t}}\right). \tag{2.8}$$

where $\varepsilon_t \sim N(0, I_n)$ and I_n is an identity matrix.

Following Chib (1995), the estimate of the marginal likelihood is based on the following identity:

$$\ln \Lambda(Y_t) = \ln F\left(Y_t | \hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) + \ln \Upsilon\left(\hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) - \ln \Theta\left(\hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right), \tag{2.9}$$

where g, λ and v_λ respectively represent vectors of g_k, $\lambda_{k,t}$ and $v_{\lambda,k}$, $k = 1, ..., n$, $t = 1, ..., T$, the superscript ˆ denotes the posterior mean, $F(\cdot)$ denotes the likelihood function, $\Upsilon(\cdot)$ is the joint prior density, $\Theta(\cdot)$ is the posterior distribution and Ξ denotes the state variables in the model. Equation 2.9 is simply the Bayes equation in logs re-arranged with the marginal likelihood

---

[6]More specifically, we assume Normal priors on the parameters controlling the dynamics of the VAR. One hyper parameter controls the overall tightness of the prior on the VAR coefficients which we set $\tau = 0.1$ following Banbura, Giannone, and Reichlin (2010). Further details can found in our working paper Chiu, Mumtaz, and Pinter (2015).

$\Lambda(Y_t)$ on the left hand side. 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 straightforward to evaluate. The computation of the likelihood and posterior density is more involved and described in Section B of the Appendix.

2.1.3 Data

We use the data-set of Stock and Watson (2012) and focus on three key macroeconomic variables for the US: industrial production, inflation and the interest rate. The choice of these variables is relatively standard and follows the literature (Primiceri (2005); Banbura, Giannone, and Reichlin (2010); D'Agostino, Gambetti, and Giannone (2013)) that used similar variables when estimating small-scale forecasting models. In addition, we add the SP500 stock market return index. This is motivated by the increasing evidence on the relevance of financial variables for improving macroeconomic forecasts.[7]

The data is 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. Industrial production 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. We use 13 lags in our estimation (Banbura, Giannone, and Reichlin (2010)), but we also explore shorter lag length specification as a robustness check.

2.2 Alternative Models

We consider three restricted versions of the VAR 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

$$Y_t = c + B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t, \tag{2.10}$$

where

$$cov(u_t) = \Sigma_t = A^{-1} H_t A^{-1'} \tag{2.11}$$

$$H_t = diag\left(\sigma_{1,t}^2, \sigma_{2,t}^2, ..\sigma_{n,t}^2\right), \tag{2.12}$$

where $\ln \sigma_{k,t}^2$ follows the process defined in equation 2.3. 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:

$$Y_t = c + B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t \tag{2.13}$$

---

[7]See Stock and Watson (2003) and numerous subsequent papers including Koop and Korobilis (2014) and the references herein.

Figure 1: The posterior density of degrees of freedom in the TVARSVOL and in the TVAR models
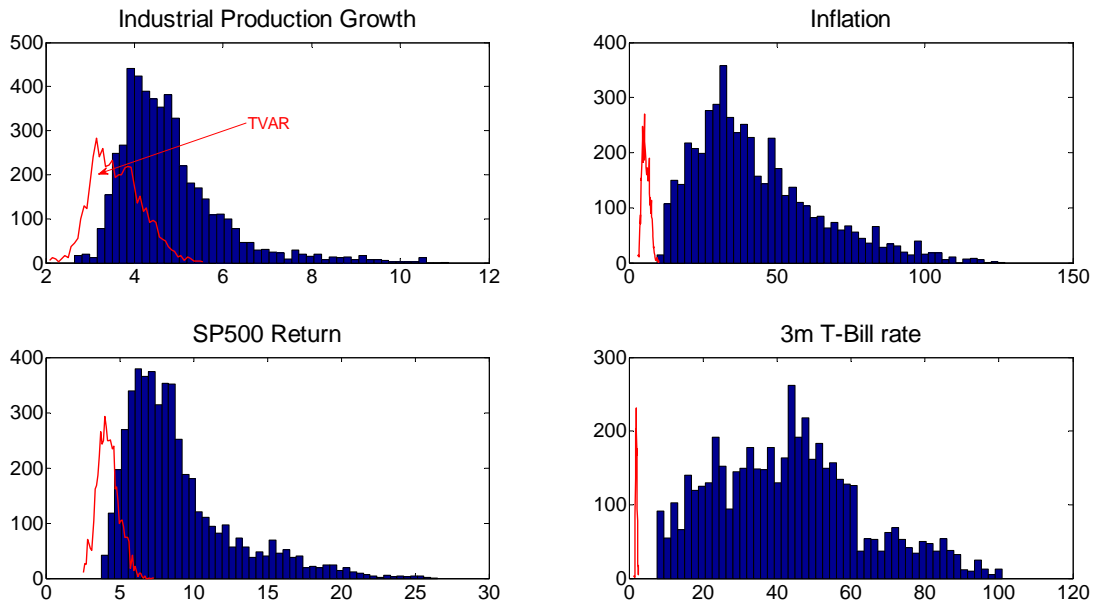
Notes: The red lines (blue bars) represent the frequency distribution of the DOF parameters from the TVAR (TVARSVOL) model.

where

$$cov(u_t) = \Sigma_t = A^{-1} H_t A^{-1'} \tag{2.14}$$

$$H_t = diag\left(\sigma_1^2 \frac{1}{\lambda_{1,t}}, \sigma_2^2 \frac{1}{\lambda_{2,t}}, ..\sigma_n^2 \frac{1}{\lambda_{n,t}}\right). \tag{2.15}$$

The final model considered is a standard linear Gaussian VAR model (BVAR). The estimation of these restricted models is carried out via Gibbs sampling using a simplification of the algorithm used in our TVARSVOL model. 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. We adopt a 'recursive estimation scheme': our first estimation covers data up to January 1970, with subsequent estimations adding one more month's data each time. In total, there are 490 recursive estimation for each of the four models under study. Before moving to model comparison, we present some of the key parameter estimates of the TVARSVOL 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 TVARSVOL. Consider the estimates for the industrial production growth. There is strong evidence that the orthogonalised shock associated with this variable is characterised by fat tails with the posterior density centred 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 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 posterior densities indicate stronger evidence in favour of fat tails for all four residuals. This confirms the argument in Curdia, del Negro, and Greenwald (2014) that ignoring low frequency movements in volatility may bias the estimates of DOF downwards, thereby overestimating the fatness of tails in the shock distributions.
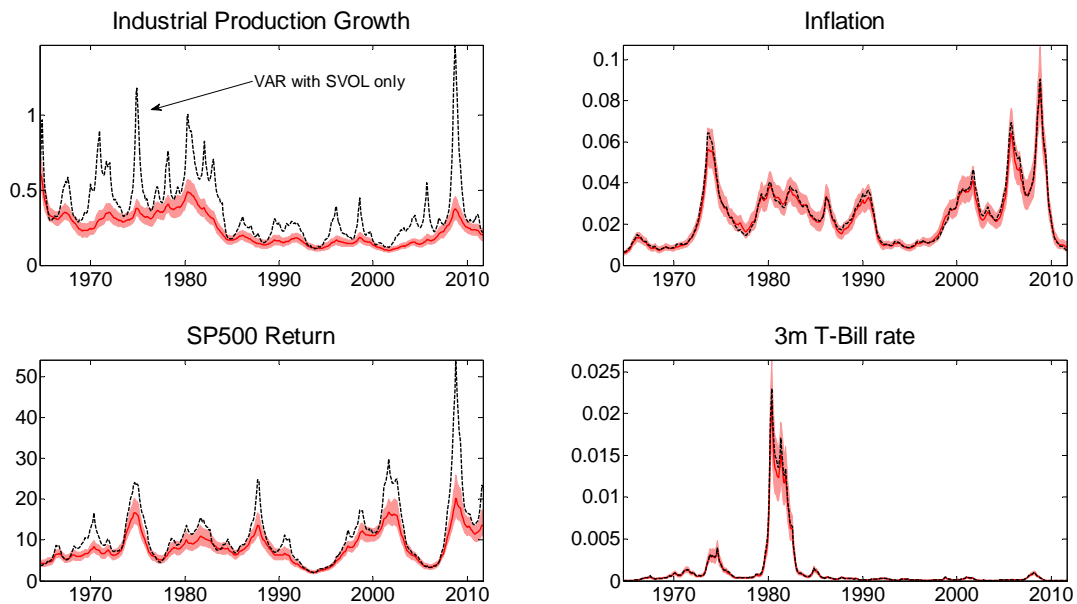
3.1.2 Stochastic Volatility

Figure 2 plots the estimated stochastic volatility from the TVARSVOL 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 TVARSVOL 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 TVARSVOL 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.

Similar conclusions have been reached by Jacquier, Polson, and Rossi (2004) in a univariate context and by Curdia, del Negro, and Greenwald (2014) using an estimated DSGE framework. These results suggest that the size of the structural break in the volatility of of US output in the early 1980s studied by the Great Moderation literature (McConnell and Perez-Quiros, 2000) may be overestimated.

These results imply that adopting the unit root assumption for stochastic volatility models,

Figure 2: Stochastic Volatility in the TVARSVOL and in the VARSVOL model

widely used since Primiceri (2005), while ignoring high-frequency movements of volatility may cause problems for interpreting time-varying dynamics in the shock process. To that end, we perform a Monte Carlo analysis of model misspecification for a bivariate VAR model where the true data generating process features both stochastic volatility and Student's t errors; see Section C of the Appendix for the full details of this analysis. We estimate the VARSVOL and TVARSVOL models and compare how well the estimated volatility paths of each model, conditional on the simulated data, are *relative to* the true paths.

For simplification, we consider a bivariate VAR model of output and prices (with two lags) whose dynamic processes are estimated on real data between 1950 and 2013. We assume both features of stochastic volatility and Student's *t* errors in the true data generating process and simulate 300 such datasets, which are then separately estimated by VARSVOL and TVARSVOL models. We present our results in percentage differences; positive percentage points indicates over-estimation of the true volatility.

Our simulations provide strong evidence that the VARSVOL model, which ignores Student's *t* errors, consistently significantly overestimates the volatility whereas the TVARSVOL model can come close to recovering the true data generating process. In particular, conditional on our simulation set-up, the VARSVOL model produces median volatility estimates up to 70 percentage points larger than the true values. Moreover, this problem of misspecification becomes more severe as the data feature more fatness in the shock distribution. Overall, our simulations suggests that the unit root assumption for stochastic volatility models may not be favourable if the actual data generating process contains fat tails.

3.2 Model Comparison

3.2.1 Marginal Likelihood

We begin the model comparison by calculating the marginal likelihood for each of the four models.

Table 1: Marginal Likelihood

Model	Log Marginal Likelihood
TVARSVOL	-1725.3
VARSVOL	-1757.9
TVAR	-2444.6
BVAR	-2852.2

Notes: Each model includes 13 lags and is estimated using 50,000 iterations (with 40,000 burns). The computation of the marginal likelihood uses 20,000 iterations (with 10,000 burns).

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 as described in Section 2.1.2 earlier. We use 10,000 additional Gibbs iterations to estimate the components of the posterior density. It is clear from Table 1 that the TVARSVOL model displays the best in-sample fit while the BVAR has the lowest estimated marginal likelihood. Allowing for fat tails or stochastic volatility improves the fit relative to the BVAR model. 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 proceed by comparing 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:

$$P\left(\hat{Y}_{t+k}|Y_t\right) = \int P(\hat{Y}_{t+k}|Y_t, \Psi_{t+k}) P\left(\Psi_{t+k}|\Psi_t, Y_t\right) P\left(\Psi_t|Y_t\right) d\Psi, \tag{3.1}$$

where $k = 1, 2, \ldots, 12$ and Ψ denotes the model parameters. The last term in equation 3.1 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. In the section below, we focus on *density* forecast comparison as described in detail by Amisano and Geweke (2011, 2013) amongst other recent papers. The density forecasts are evaluated using log scores (LS) are defined as:

$$LS_t = \ln P\left(Y_{t+k}\right), \tag{3.2}$$

where $P\left(Y_{t+k}\right)$ 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.

In addition, we also compute probability integral transforms (PITs) defined as:

$$PIT_t = \Phi\left(Y_{t+k}\right), \tag{3.3}$$

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

3.2.2.1 Log Score Comparisons Table 2 considers the log score for each model relative to that of BVAR for one-month, three-month, six-month and twelve-month ahead forecasts. The table presents the average log scores across the forecasting sample, with a positive number indicating an improvement over the BVAR model.

Consider the results for industrial production. At the one-month horizon, allowing for fat tails or stochastic volatility leads to a similar improvement over the BVAR density forecasts. This is is not the case at longer horizons where fat tails are clearly important. At the six-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 six-month and the one-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 TVARSVOL model produces the largest improvement over the BVAR for inflation at the six-month and twelve-month horizon. As for the short-term interest rate, the TVARSVOL model produces the best performance At the one-month and three- month horizon, whereas the VARSVOL model delivering the largest improvement over the BVAR at longer horizons.

Furthermore, we ask how the different models perform in jointly forecasting financial variables (SP500 and the interest rate) relative to macro-variables (industrial production and inflation). To answer this question we use the adaptive kernel density estimator of Botev, Grotowski, and Kroese (2010) to construct bivariate log scores for the IP growth – inflation and SP500 – interest rate pairs. Table 3 shows the relative improvements of the three volatility models relative to the BVAR. The results suggest that the relative improvement of each model tends to be larger for the SP500 – interest rate pair than for the IP growth – inflation pair. The results also confirm that the modelling of both high- and low-frequency movements in volatility increases forecast accuracy.[8]

3.2.2.2 PIT Comparisons Figure 3 plots the histograms of PITs estimated by one-month ahead forecasts, and histograms from a uniform distribution in order to provide visual assess-

---

[8]Section D reports the evaluation of the *point* forecasts by means of root-mean-squared errors (RMSE).

Table 2: Percentage Improvement in Log Scores over the BVAR Model

	1M	3M	6M	12M
IP Growth				
BVAR (benchmark)	*-1.236*	*-1.345*	*-1.424*	*-1.401*
TVARSVOL	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				
BVAR (benchmark)	*-0.062*	*-0.625*	*-0.694*	*-0.909*
TVARSVOL	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 Return				
BVAR (benchmark)	*-2.984*	*-3.040*	*-2.890*	*-2.942*
TVARSVOL	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				
BVAR (benchmark)	*0.842*	*-0.250*	*0.604*	*0.318*
TVARSVOL	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

Notes: The numbers are relative to the BVAR model and are computed from the average log scores obtained from the 490 recursive estimations for each of the four models (1960 estimated models in total). Each recursive estimation uses 13 lags and 11,000 iterations. The first line of each panel reports the average log scores of the benchmark BVAR model.

ment of density calibration. The results are very similar at other forecast 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 interval between 0.4 and 0.6, indicating departures from uniformity. In contrast, the PITs for 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.

As for inflation, both the TVARSVOL and the VARSVOL models appear to be better calibrated in terms of PITs than the the TVAR and the BVAR do, as the the histograms for the latter models display mass at the tails. Although it is difficult to distinguish between the PIT histograms across models for SP500, the TVARSVOL and the VARSVOL models do appear to perform better for the short-term interest rate, because their PIT distributions do not display hump shapes but the BVAR and TVAR models do.

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

Table 3: Percentage Improvement in Bivariate Log Scores over the BVAR Model

	1M	3M	6M	12M
IP-π pair				
BVAR (benchmark)	*-1.098*	*-1.298*	*-1.350*	*-1.495*
TVARSVOL	40.108	35.511	24.799	19.269
TVAR	38.675	32.284	23.423	20.850
VARSVOL	29.790	23.577	12.630	3.494
SP500-R pair				
BVAR (benchmark)	*-1.202*	*-1.997*	*-2.225*	*-2.509*
TVARSVOL	89.296	82.362	44.618	6.758
TVAR	50.027	41.581	23.906	12.628
VARSVOL	85.950	77.289	43.168	9.217

Note: The numbers are relative to the BVAR model and are computed from the average bivariate log score obtained from the 490 recursive estimations for each of the four models. The first line of each panel reports the average bivariate log scores of the benchmark BVAR model.

Figure 3: PIT histograms at the one month forecast horizon
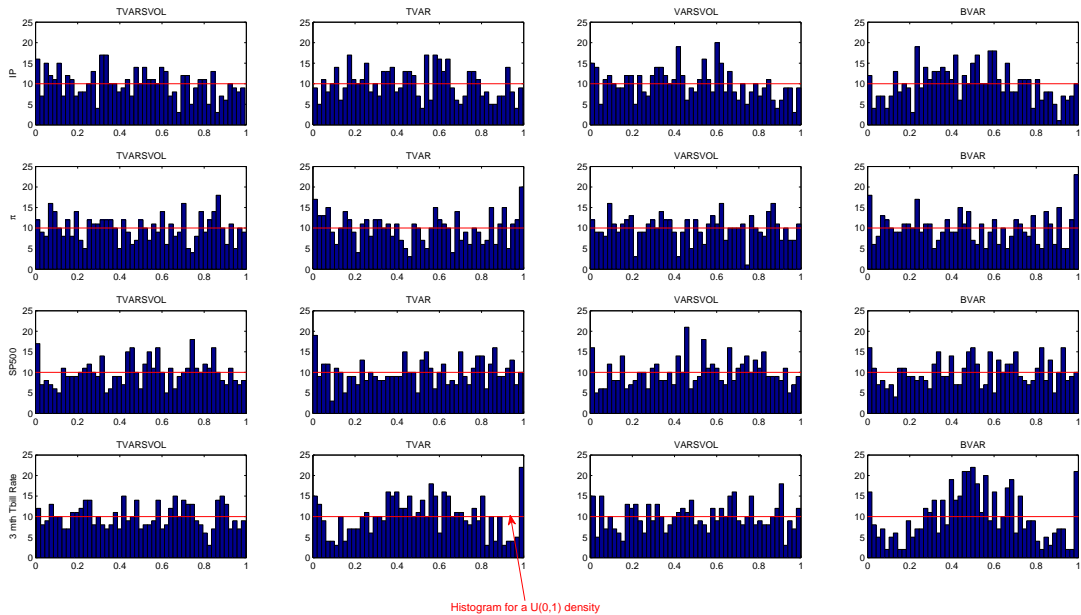
3.2.3 Forecasting the Great Recession

One of the main criticisms macroeconometric models have received is related to their inability to forecast the recent the Great Recession. This leads econometricians and macroeconomists to question the adequacy of their analysis (Ng and Wright, 2013). To see whether accounting for fat-tails would have changed this result, we consider the evolution of log scores over the recent financial crisis as shown in Figure 4.

The left axis in each panel shows the percentage improvement in log scores over the benchmark BVAR model. In this figure we consider the three-month forecasting horizon but the results are similar at other horizons. The right axis shows the actual data, plotted as an area chart, for each variable.

The top left panel shows the results for industrial production. The performance of the three

13

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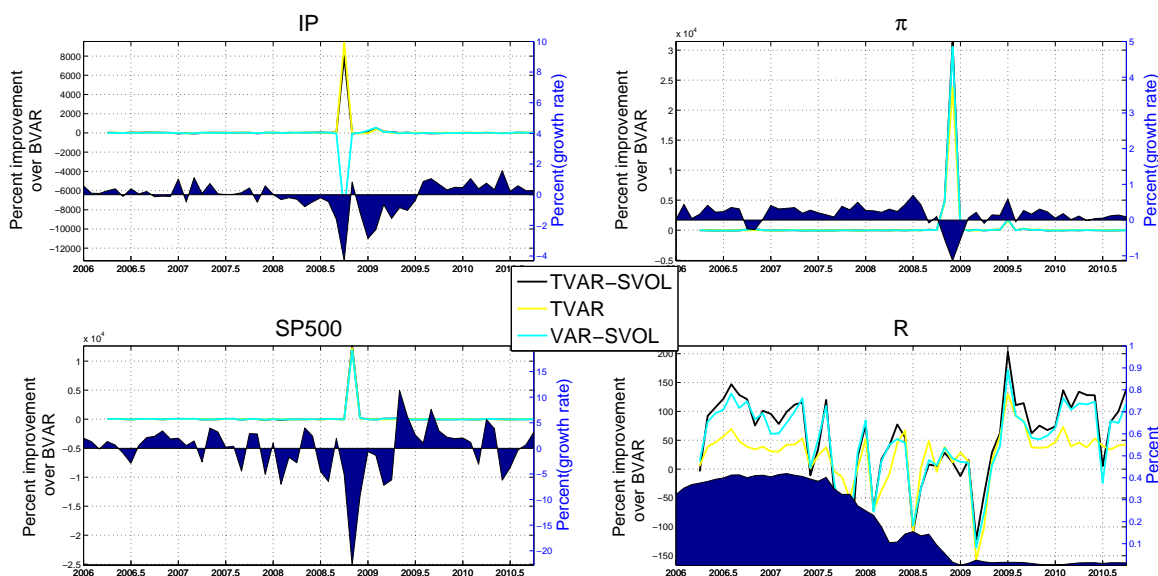

models is similar before the onset of the deep recession at the end of 2008. The large decline in industrial production coincided with a very large divergence in the performance of models with and without fat tails. The TVAR and the TVARSVOL model show a huge improvement in the log score. In contrast, the accuracy of the VARSVOL model deteriorates substantially relative to the BVAR model.

To get a better understanding of this divergence in the forecast performance, we take a closer look at the outcome in September 2008 when industrial production fell by 4.3%. Given that the mean and the standard deviation of industrial production growth in sample are 0.23% and 0.83%, respectively, a forecasting model with normally distributed shocks would assign virtually zero probability to such an outcome.

To illustrate this, Figure 5 shows the three-month ahead forecast density of industrial production for September 2008 from the four models together with the actual out-turn in that month depicted by the vertical red line. The left tails of the densities from the BVAR and the VARSVOL model do not include the actual industrial production out-turn of −4.3%. In contrast, the densities from the models with fat tails cover this eventuality. This highlights the fact that the assumption of normality may lead 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.

As for the stock returns 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 is mixed for the short-term interest rate over the initial cutting phase of 2007 and 2008. However, stochastic volatility appears to be

14

Figure 5: Forecasting the Great Recession
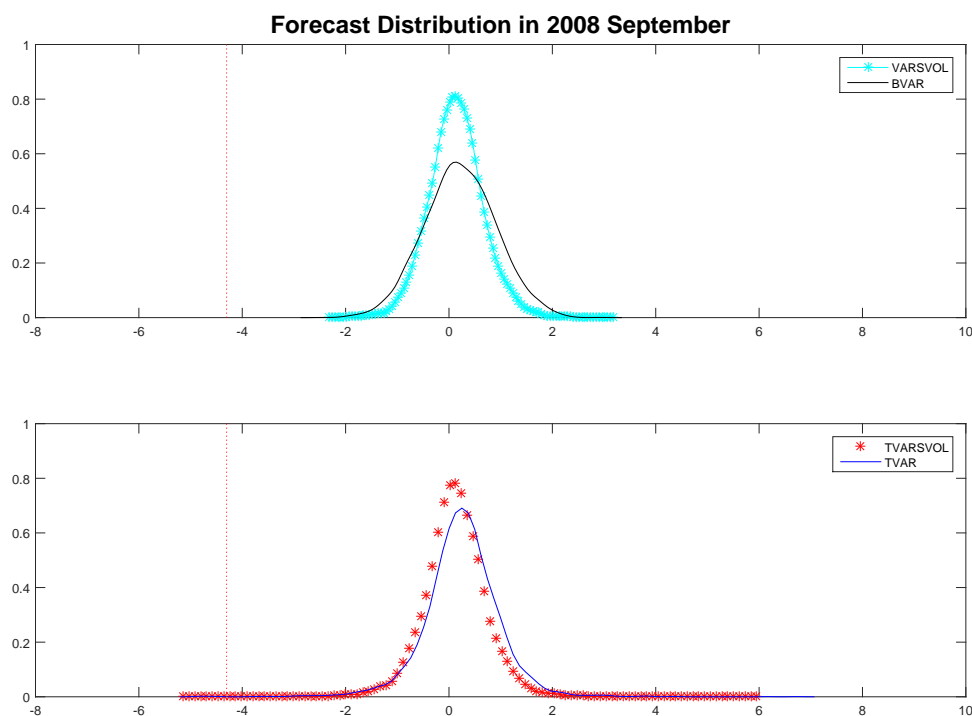
Note: The panels show the 3 month-ahead forecast distributions for industrial production in September 2008. The actual out-turn (−4.3%) is depicted by the dashed vertical red line.

important over the post 2009 period that was characterised by persistently low interest rates.[9]

3.2.4 Forecasting Performance in Various Sub-sample Periods

To further investigate how introducing fat tails and stochastic volatility to a VAR model can improve out-of-sample forecasting performance, we focus on *three-month* ahead forecasts and break down the improvement of log scores into three sub-sample periods: the pre-Great Moderation period (January 1970 - December 1983), the Great Moderation period (January 1984 - December 2006), and the Great Recession (January 2007 - September 2011). Results are reported in Table 4.

With the exception of inflation forecasts in the 1970s and early 1980s, the TVARSVOL model always dominates BVAR, and in many cases, out-competes VARSVOL. This is true during the Great Recession period where the US economy experienced an increase in volatility and big outliers in IP growth and inflation rates, as well as during Great Moderation which was characterized by lower economic volatility and lack of occurrence of outliers. In other words, evidence seems to suggest that while including stochastic volatilities can be beneficial in improving forecast accuracy most of the time, introducing fat tails can deliver additional significant improvement when the economy experiences extreme events.

---

[9]This is of course subject to the caveat that the federal funds rate are subject to a zero lower bound during this period.

Table 4: Percentage Improvement in Log Scores for three-month ahead forecasts over the BVAR Model for three sub-sample periods

	1970M1-1983M12	1984M1-2006M12	2007M1-2011M9
	IP Growth		
BVAR (benchmark)	*-1.483*	*-0.894*	*-3.592*
TVARSVOL	13.402	10.561	200.03
TVAR	20.823	8.036	227.74
VARSVOL	10.066	9.165	-431.45
	π		
BVAR (benchmark)	*-0.042*	*0.330*	*-8.666*
TVARSVOL	-56.882	2.619	844.46
TVAR	-6.039	0.364	681.15
VARSVOL	-30.589	0.494	803.46
	SP500 Return		
BVAR (benchmark)	*-2.862*	*-2.674*	*-5.955*
TVARSVOL	8.617	6.029	275.33
TVAR	8.087	6.082	264.93
VARSVOL	5.765	6.453	278.73
	R		
BVAR (benchmark)	*-3.443*	*1.461*	*1.245*
TVARSVOL	425.87	46.553	45.046
TVAR	84.535	23.354	20.382
VARSVOL	427.42	46.251	36.280

Notes: The numbers are relative to the BVAR model and are computed from the average log scores obtained from the recursive estimations for each of the four models within each sub-sample period. The first line of each panel reports the average log scores of the benchmark BVAR model.

3.2.5 Sensitivity Analysis

In this subsection, we ask whether our results are robust to choosing alternative priors for the degree of fatness of shock distribution. In the TVARSVOL model, we used the prior $v_0 = 20$ for $p(v_{\lambda,k}) \sim \tilde{\Gamma}(v_0, 2)$, $k = 1, 2, ..., n$ which assigns a reasonable probability to extreme events. As an alternative prior, we re-estimate the models using the value $v_0 = 50$, implying a higher prior weight on the possibility of normality.

Table 5: Percentage Improvement in Log Scores of the TVARSVOL over the BVAR Model when $v_0 = 50$ (sensitivity analysis)

	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

Notes: The numbers are relative to BVAR model and computed from the average log scores obtained from the 490 recursive estimations. Each recursive estimation uses 13 lags and 11,000 iterations. Average log scores of the benchmark BVAR model are reported in Table 2 and are not reproduced here for brevity.

Table 5 presents the estimated log scores (relative to the BVAR) from TVARSVOL using this alternative prior. The results indicate that for industrial production and stock market returns (the variables for which the orthogonalised errors displayed the most evidence for non-

Table 6: Forecast Comparison with the TVP-VARSVOL Model

	(a) Relative RMSE values					(b) Relative Log-scores values			
	1M	3M	6M	12M		1M	3M	6M	12M
	IP Growth					IP Growth			
BVAR (benchmark)	*0.510*	*0.604*	*0.648*	*0.684*		*-1.183*	*-1.296*	*-1.342*	*-1.279*
TVARSVOL	0.979	0.984	0.985	0.985		16.164	-4.756	26.016	17.005
TVP-VARSVOL	0.972	0.973	0.977	0.981		8.780	8.641	9.864	13.629
	π					π			
BVAR (benchmark)	*0.143*	*0.173*	*0.188*	*0.206*		0.074	-0.152	-0.348	-0.605
TVARSVOL	1.006	1.028	1.042	1.042		15.719	22.085	35.892	43.107
TVP-VARSVOL	0.980	0.975	0.974	0.958		20.530	21.051	46.934	75.820
	R					R			
BVAR (benchmark)	*0.023*	*0.043*	*0.064*	*0.094*		*0.269*	*-Inf*	*0.274*	*0.229*
TVARSVOL	0.991	0.951	0.949	0.963		223.713	Inf	67.003	6.647
TVP-VARSVOL	0.993	0.990	0.984	0.975		192.674	Inf	58.181	17.086

Note: The numbers are relative to the BVAR model and are computed from the average RMSE and log scores obtained from the 490 recursive estimations. Each recursive estimation uses 2 lags and 11,000 iterations. The values are relative to the BVAR model. The first line of each panel reports the average RMSE and log scores of the benchmark BVAR model. The –Inf value for the log-score of three-month ahead interest rate is caused by huge forecast errors in 1980, a period characterised by highly volatile monetary policy. The BVAR model, but not our other models, struggles to produce appropriate density forecast during such volatile periods in this three-variable model, hence leading to an infinite percent relative improvements.

Gaussianity), the average relative log scores are fairly similar. This provides evidence that the key results do not depend on the prior for the degree of fatness of shocks distributions.

In Section E of the Appendix, we provide further evidence that our results are robust against alternative orderings of the variables and different choices of hyperparameters governing the prior distribution of the variance of the shocks g_k in equation 2.3.

3.3 Comparison with Time-varying Coefficient VARs

As a further robustness check, we explore how the TVARSVOL model performs relative to a VAR with time-varying parameters and stochastic volatility (TVP-VARSVOL) as in Primiceri (2005) and D'Agostino, Gambetti, and Giannone (2013), amongst others. Given the computational intensity of performing recursive forecasts with this model, we estimate a three-variable system by omitting the SP500 return series and by considering two lags only.[10] We adopt the same set-up for the TVARSVOL and the benchmark BVAR models so that we can perform out-of-sample forecasting model comparison. This analysis also serves as a robustness check to investigate whether *reducing* the information set and the lag length from 13 to 2 will change the relative forecast performance of the TVARSVOL model. A brief description of the TVP-VARSVOL model can be found in Section F of the Appendix.

---

[10]The addition of an extra variable in a TVP-VARSVOL model leads to an exponential increase in computational time (recall we need to compute 490 recursive forecasts each of which should include a sufficiently large number of posterior draws.) We argue that the three-variable model is sufficient to illustrate that a more general modelling of the covariance matrix can be as important as modelling time-variation in the dynamic parameters. Exploring whether time-varying parameters or time-varying covariance matrix is more important for forecast accuracy is an important avenue for future research, which is beyond the scope of the current paper.

Table 6 presents the results for the point forecasts (evaluated by Root-mean-squared errors, or RMSE) and density forecasts from the TVP-VARSVOL and TVARSVOL models relative to the BVAR. They suggest that the TVARSVOL model continues to perform well especially in terms of density forecast accuracy. The TVP-VARSVOL model also performs well, especially on the point forecasts of inflation. This is consistent with the results of D'Agostino, Gambetti, and Giannone (2013) who find that accounting for time-varying coefficients significantly improves the accuracy of point forecasts for inflation. Nevertheless, TVARSVOL performs well in terms of density forecasts relative to the TVP-VARSVOL model. This suggests that achieving more superior VAR forecasts can be done not only by accounting for time-variation in the dynamic coefficients but also by modelling appropriately the variance-covariance matrix of VAR models.

4   Results for Canada, Germany and the UK

In this section, we estimate the four model specifications for other countries including the United Kingdom (UK), Germany and Canada. From the Global Financial Database, we retrieve (i) monthly data for the UK covering the period January 1959- September 2011, including the FTSE all-share Index, CPI, industrial production index, and the three-month Treasury Bill yield; and (ii) monthly data for Germany covering the period January 1961 - September 2011, including the CDAX composite index, CPI and industrial production. Monthly data for Canada covering the period January 1961 - September 2011 and for the short-term German interest rate are obtained from the OECD Economic Outlook. Monthly growth rates are calculated for all the variables except the interest rate. As in the case of the US, we recursively estimate the forecast density 3.1 for of each of the models and for each of the countries.[11] We present the results for both point forecasts and density forecasts.

Table 7 presents the results for Canada. With the exception of inflation, point forecasts of all variables are more accurate relative to the BVAR model which performs the worst in predicting the interest rate.

Results for log scores suggest that, relative to the VARSVOL model, the TVAR tends to outperform in terms of forecasting inflation and stock returns and tends to under-perform in terms of forecasting industrial production growth and interest rates. Over the one-month horizon, the TVARSVOL model comes first in predicting stock returns and the interest rate.

Results for the UK, presented in Table 8, highlight the dangers of ignoring the slow-moving component of volatility when using fat-tailed VAR models for forecasting purposes. Outliers such as the a 400 basis point jump in short rate in July 1973 (discussed in the note of Table 8) leads to an over-estimation of the fatness of the shock distribution in case of the TVAR model, which in this case substantially worsens the point forecasts for all the variables. The reason why this problem is particularly severe in the case of the UK is related to the high level of inflation volatility that characterised the 1970s and early 1980s (Liu and Mumtaz 2011),

---

[11]The recursive algorithm starts with the estimation of each of the models up to January 1970 in the case of the UK and up to January 1972 in the case of Canada and Germany. This delivers 490 forecast densities for the UK and 466 forecast densities for Germany and Canada.

Table 7: Forecast Comparison Results for Canada (with BVAR as the benchmark)

(a) Relative RMSE values					(b) Relative Log-scores values			
	1M	3M	6M	12M	1M	3M	6M	12M
IP Growth					IP Growth			
BVAR (benchmark)	*0.879*	*0.999*	*1.043*	*1.067*	*-1.539*	*-1.541*	*-1.554*	*-1.546*
TVARSVOL	0.962	0.957	0.963	0.971	6.134	5.870	4.299	1.484
TVAR	0.964	0.959	0.966	0.977	5.240	4.348	3.445	0.910
VARSVOL	0.960	0.958	0.964	0.973	6.549	-1.560	4.097	1.293
π					π			
BVAR (benchmark)	*0.244*	*0.284*	*1.043*	*1.067*	*-0.399*	*-0.405*	*-0.400*	*-0.433*
TVARSVOL	1.006	1.012	1.010	1.003	-35.920	-0.637	-24.371	-5.239
TVAR	1.003	1.008	1.008	1.003	10.962	10.753	6.609	4.270
VARSVOL	1.000	1.005	1.004	0.995	-45.191	-68.749	-63.174	-26.838
Stock Return					Stock Return			
BVAR (benchmark)	*3.243*	*3.800*	*3.977*	*4.067*	*-3.160*	*-3.164*	*-3.319*	*-3.243*
TVARSVOL	0.972	0.967	0.972	0.982	17.056	8.471	14.831	21.412
TVAR	0.972	0.963	0.969	0.979	18.477	17.775	29.877	34.501
VARSVOL	0.976	0.965	0.970	0.980	1.253	1.506	-45.327	-18.320
R					R			
BVAR (benchmark)	*0.026*	*0.052*	*0.081*	*0.123*	*0.478*	*0.518*	*0.340*	*-0.074*
TVARSVOL	0.971	0.946	0.933	0.911	164.370	69.064	31.375	25.806
TVAR	0.961	0.929	0.916	0.904	145.807	48.796	17.867	22.081
VARSVOL	0.968	0.940	0.926	0.902	161.846	69.760	34.593	30.046

Note: The numbers are relative to the BVAR model and are computed from average RMSE and log scores obtained from the 466 recursive estimations for each of the four models (1864 estimated models in total). Each recursive estimation uses 13 lags and 11,000 iterations. The values are relative to the BVAR model. The first line of each panel reports the average RMSE and log scores of the benchmark BVAR model.

Table 8: Forecast Comparison Results for the UK (with BVAR as the benchmark)

(a) Relative RMSE values					(b) Relative Log-scores values			
	1M	3M	6M	12M	1M	3M	6M	12M
IP Growth					IP Growth			
BVAR (benchmark)	0.904	1.026	1.096	1.157	-1.839	-1.888	-1.885	-1.885
TVARSVOL	0.955	0.973	0.986	1.000	28.354	32.823	35.220	36.373
TVAR	0.960	1.394	1.302	1.234	26.456	29.045	27.649	26.557
VARSVOL	0.970	0.980	0.989	1.131	31.983	30.677	30.949	28.351
π					π			
BVAR (benchmark)	0.247	0.308	0.343	0.385	-0.834	-1.066	-0.987	-1.183
TVARSVOL	0.969	0.952	0.944	0.922	63.984	78.263	68.334	79.877
TVAR	0.951	1.916	2.010	1.862	59.271	67.140	51.538	61.716
VARSVOL	0.931	0.923	0.926	1.083	63.364	80.726	65.190	72.388
Stock Return					Stock Return			
BVAR (benchmark)	3.999	4.631	4.835	4.997	-4.160	-4.149	-4.066	-4.134
TVARSVOL	0.985	0.989	0.995	1.001	109.816	109.552	99.838	105.163
TVAR	1.050	1.085	1.102	3.783	101.013	95.805	99.265	106.761
VARSVOL	0.990	0.991	0.999	2.108	112.098	108.108	96.489	98.820
R					R			
BVAR (benchmark)	0.339	0.613	0.917	1.377	-Inf	-2.042	-2.356	-2.603
TVARSVOL	0.986	0.965	0.968	0.969	Inf	87.846	64.693	39.441
TVAR	1.380	1.520	1.508	5.918	Inf	67.655	45.342	23.607
VARSVOL	0.965	0.943	0.964	3.029	Inf	83.237	57.088	25.683

Note: The numbers are relative to the BVAR model and are computed from the average RMSE and log scores values obtained from the 490 recursive estimations for each of the four models (1960 estimated models in total). Each recursive estimation uses 13 lags and 11,000 iterations. The values are relative to the BVAR model. The first line of each panel reports the average RMSE and log scores of the benchmark BVAR model. The calculated log score value for one-month-ahead interest rates density forecasts is –Inf. The Bank of England engineered a 400 bp jump in the short rate between the end of June and the end of July of 1973. The one-period-ahead forecast density produced in 1973M6 produced by the BVAR has no probability mass to cover such a jump, leading to a –Inf in log score. Our models, which are able to generate a tail long enough to cover this event (though the probability is very low), produce an infinite percent improvement in the log score.

Table 9: Forecast Comparison Results for Germany (with BVAR as the benchmark)

	(a) Relative RMSE values				(b) Relative Log-scores values			
	1M	3M	6M	12M	1M	3M	6M	12M
	IP Growth				IP Growth			
BVAR (benchmark)	1.253	1.495	1.595	1.702	-2.083	-2.107	-2.089	-2.200
TVARSVOL	0.946	0.955	0.971	0.983	23.165	16.330	13.709	20.284
TVAR	0.948	0.957	0.973	0.984	21.639	14.465	11.694	18.903
VARSVOL	0.952	0.956	0.970	0.982	14.637	12.250	10.145	21.281
	π				π			
BVAR (benchmark)	0.167	0.199	0.213	0.224	-1.298	-Inf	-1.121	-1.212
TVARSVOL	0.989	0.993	0.997	1.001	76.055	Inf	79.753	5.820
TVAR	0.982	0.985	0.990	0.994	111.258	Inf	121.558	104.502
VARSVOL	0.998	0.996	1.000	1.003	128.675	Inf	113.001	117.387
	Stock Return				Stock Return			
BVAR (benchmark)	4.108	4.714	4.899	5.049	-3.580	-3.409	-3.449	-3.512
TVARSVOL	0.973	0.978	0.983	0.986	53.356	35.559	37.985	41.023
TVAR	0.978	0.982	0.986	0.989	40.707	33.865	38.310	37.958
VARSVOL	0.971	0.980	0.983	0.987	51.325	34.981	37.943	40.467
	R				R			
BVAR (benchmark)	0.019	0.038	0.061	0.099	2.026	1.266	0.740	-0.153
TVARSVOL	0.925	0.847	0.848	0.855	50.090	31.095	31.269	44.232
TVAR	0.924	0.864	0.868	0.878	28.100	19.721	17.103	57.359
VARSVOL	0.900	0.836	0.840	0.853	38.159	30.262	31.774	52.103

Note: The numbers are relative to the BVAR model and are computed from the average RMSE and log scores values obtained from the 466 recursive estimations for each of the four models (1864 estimated models in total). Each recursive estimation uses 13 lags and 11,000 iterations. The first line of each panel reports the average RMSE and log scores of the benchmark BVAR model. The presence of –Inf log score value for three-month-ahead inflation forecasts is due to extreme values: an extremely large negative observation occurred in January 1991 (−2.07%) in the midst of the economic recession following the German reunification. Given that the mean and the standard deviation of the monthly German inflation rate in the sample are 0.23% and 0.27%, respectively, the BVAR model assigns virtually zero probability to this event.

which was also substantially higher than in the US, Canada and Germany during the same time (Mumtaz and Surico 2012).

To sum up, the log score results for the UK suggest that the TVARSVOL model outperforms the other competing volatility models, with the exception of the one-month ahead forecast of stock returns and industrial production growth, in which case the VARSVOL delivers more accurate density forecasts.

Results for Germany are broadly in line with the previous findings. With the exception of inflation, the point forecasts of all variables improve by either volatility models, and the TVARSVOL model tends to dominate other volatility models, at least in the short-run.

Finally, Table 10 shows the estimated values of the marginal likelihood for the three countries. The results are in line with the evidence for the US: The BVAR and the TVAR have worse in-sample fit than the TVARSVOL and VARSVOL models. The TVARSVOL has the best in-sample fit in the case of Canada, whereas the VARSVOL model comes first in the case of Germany and the UK. This may be explained by the fact that the nature of shocks differ across countries. For example, as in the case of the UK, the extremely volatile period before the

1980s and the relatively calm period after may appear to be better explained by changes in low frequency volatility rather than by high frequency volatility manifested by fat tails. In other words, the UK is less subject to high frequency volatility shocks, implying that VARSVOL model provides a better in-sample fit.

Table 10: Marginal Likelihood

Model	Canada	Germany	UK
TVARSVOL	-1857.2	-2428.7	-3201.6
VARSVOL	-1980.0	-2406.6	-3058.6
TVAR	-2479.6	-3136.2	-4375.8
BVAR	-2133.4	-4445.8	-3615.1

Notes: Each model includes 13 lags and is estimated using 50,000 iterations (with 40,000 burns). The computation of the marginal likelihood uses 20,000 iterations (with 10,000 burns).

Overall, our findings suggest that models which account for heteroscedasticity in the error structure can exhibit considerably improved forecast accuracy relative to the benchmark linear Gaussian BVAR model. These results are consistent with those recently presented by Clark and Ravazzolo (2015). An additional result is that explicitly modelling both the low- and high-frequency movements in volatility could provide further improvements in the forecast accuracy as well as in the in-sample fit, as shown in the case of the US and Canada.

5   Conclusions

This paper introduces a VAR model that incorporates stochastic volatility and fat tailed disturbances. We show that this model fits a monthly US data-set better than alternatives that do not include these features. The estimates of the model present strong evidence that disturbances to industrial production growth and stock market returns are non-normal. Incorporating this non-normality in the model leads to substantial improvement in the accuracy of forecast densities. In particular, linear BVAR models with Gaussian disturbances fail to attach any probability to low values of industrial production growth seen in late 2008 in the US. Our results, that are also consistent with findings for a further set of countries, highlight the importance of incorporating the possibility of fat tails in forecasting models.

References

AMISANO, G., AND J. GEWEKE (2011): "Optimal prediction pools," *Journal of Econometrics*, 164(1), 130–141. 3.2.2

———— (2013): "Prediction using several macroeconomic models," Working Paper Series 1537, European Central Bank. 3.2.2

ASCARI, G., G. FAGIOLO, AND A. ROVENTINI (2015): "Fat-tail Distributions and Business-cycle Models," *Macroeconomic Dynamics*, Forthcoming. 1

BANBURA, M., D. GIANNONE, AND L. REICHLIN (2010): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25(1), 71–92. 2.1.1, 6, 2.1.3

BAUWENS, L., G. KOOP, D. KOROBILIS, AND J. V. ROMBOUTS (2015): "The Contribution of Structural Break Models to Forecasting Macroeconomic Series," *Journal of Applied Econometrics*, 30(4), 596–620. 3

BLAKE, A. P., AND H. MUMTAZ (2012): *Applied Bayesian econometrics for central bankers*, no. 4 in Technical Books. Centre for Central Banking Studies, Bank of England. F

BOTEV, Z. I., J. F. GROTOWSKI, AND D. P. KROESE (2010): "Kernel density estimation via diffusion," *Ann. Statist.*, 38(5), 2916–2957. 3.2.2.1

CHAHAD, M., AND F. FERRONI (2014): "Structural VAR, Rare Events and the Transmission of Credit Risk in the Euro Area," Discussion paper. 5

CHIB, S. (1995): "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association*, 90(423). 2.1.2, 2.2, 3.2.1

CHIB, S., AND I. JELIAZKOV (2001): "Marginal Likelihood From the Metropolis-Hastings Output," *Journal of the American Statistical Association*, 96, 270–281. 5

CHIB, S., AND S. RAMAMURTHY (2014): "DSGE Models with Student-t Errors," *Econometric Reviews*, 33(1-4), 152–171. 1

CHIU, C.-W. J., H. MUMTAZ, AND G. PINTER (2015): "Forecasting with VAR Models: Fat Tails and Stochastic Volatility," CReMFi Discussion Papers 2, CReMFi, School of Economics and Finance, QMUL. 6

CLARK, T. E., AND F. RAVAZZOLO (2015): "Macroeconomic Forecasting Performance Under Alternative Specifications of Time-varying Volatility," *Journal of Applied Econometrics*, Forthcoming. 1, 4

COGLEY, T., AND T. J. SARGENT (2005): "Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S," *Review of Economic Dynamics*, 8(2), 262–302. 2.1.1, A.3, A.5, E, 9

CURDIA, V., M. DEL NEGRO, AND D. L. GREENWALD (2014): "Rare Shocks, Great Recessions," *Journal of Applied Econometrics*, 29(7), 1031–1052. 1, 3.1.1, 3.1.2

D'AGOSTINO, A., L. GAMBETTI, AND D. GIANNONE (2013): "Macroeconomic forecasting and structural change," *Journal of Applied Econometrics*, 28(1), 82–101. 2.1.3, 3.3, 3.3

DOAN, T., R. B. LITTERMAN, AND C. A. SIMS (1983): "Forecasting and Conditional Projection Using Realistic Prior Distributions," NBER Working Papers 1202, National Bureau of Economic Research, Inc. 2.1.1

ELLIOTT, G., AND A. TIMMERMANN (2013): "Handbook of Economic Forecasting," vol. 2, Part A. Elsevier. 1

FERNANDEZ-VILLAVERDE, J., AND J. F. RUBIO-RAMIREZ (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74(4), 1059–1087. 1

GERLACH, R., C. CARTER, AND R. KOHN (2000): "Efficient Bayesian Inference for Dynamic Mixture Models," *Journal of the American Statistical Association*, 95(451), 819–828. 3

GEWEKE, J. (1992): "Priors for macroeconomic time series and their application," Discussion paper. 1

——— (1993): "Bayesian Treatment of the Independent Student- t Linear Model," *Journal of Applied Econometrics*, 8(S), S19–40. 1, 2, 2.1.1, A.1, A.2

——— (1994): "Priors for Macroeconomic Time Series and Their Application," *Econometric Theory*, 10(3-4), 609–632. 1

JACQUIER, E., N. G. POLSON, AND P. E. ROSSI (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 12(4), 371–89. A.4

——— (2004): "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," *Journal of Econometrics*, 122(1), 185–212. 3, 3.1.2

JUSTINIANO, A., AND G. E. PRIMICERI (2008): "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98(3), 604–41. 1

KOOP, G. (2003): *Bayesian Econometrics*. Wiley. A.1

KOOP, G., AND D. KOROBILIS (2014): "A new index of financial conditions," *European Economic Review*, 71, 101 – 116. 7

LIU, P., AND H. MUMTAZ (2011): "Evolving Macroeconomic Dynamics in a Small Open Economy: An Estimated Markov Switching DSGE Model for the UK," *Journal of Money, Credit and Banking*, 43(7), 1443–1474. 4

LIU, Z., D. F. WAGGONER, AND T. ZHA (2011): "Sources of macroeconomic fluctuations: A regime-switching DSGE approach," *Quantitative Economics*, 2(2), 251–301. 1

MCCONNELL, M. M., AND G. PEREZ-QUIROS (2000): "Output Fluctuations in the United States: What Has Changed since the Early 1980's?," *American Economic Review*, 90(5), 1464–1476. 3.1.2

MISHKIN, F. S. (2011): "Monetary Policy Strategy: Lessons from the Crisis," NBER Working Papers 16755, National Bureau of Economic Research, Inc. 1

Mumtaz, H., and P. Surico (2012): "Evolving International Inflation Dynamics: World And Country-Specific Factors," *Journal of the European Economic Association*, 10(4), 716–734. 4

Ng, S., and J. H. Wright (2013): "Facts and Challenges from the Great Recession for Forecasting and Macroeconomic Modeling," *Journal of Economic Literature*, 51(4), 1120–54. 1, 3.2.3

Ni, S., and D. Sun (2005): "Bayesian Estimates for Vector Autoregressive Models," *Journal of Business and Economic Statistics*, 23(1), pp. 105–117. 1

Primiceri, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852. 2, 2.1.3, 3.1.2, 3.3, C

Sims, C. A., and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), 949–68. 2.1.1

Stock, J. H., and M. W. Watson (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41(3), 788–829. 7

——— (2012): "Disentangling the Channels of the 2007-2009 Recession," NBER Working Papers 18094, National Bureau of Economic Research, Inc. 2.1.3

Student (1908): "The Probable Error of a Mean," *Biometrika*, 6(1), pp. 1–25. 1

Uhlig, H. (1997): "Bayesian Vector Autoregressions with Stochastic Volatility," *Econometrica*, 65(1), 59–74. 2

Appendix

A   The Gibbs Sampling Algorithm for TVARSVOL

The Gibbs algorithm for the TVARSVOL model cycles through six conditional posterior distributions. The details of each conditional posterior density is provided in the following subsections. As a shorthand symbol, we use Ψ to denote all other parameters except the specific parameter(s) under discussion.

A.1   Drawing $G(\lambda_{k,t}|\Psi)$

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 orthogonalised 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 $\tilde{\varepsilon}_t$. 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 kth column of the matrix $\tilde{\varepsilon}_t$.

A.2   Drawing $G(v_{\lambda,k}|\Psi)$

The conditional distribution for the degree of freedom parameter capturing the fatness of tails is non-standard and given by:

$$G(v_{\lambda,k}|\Psi) \propto \left(\frac{v_{\lambda,k}}{2}\right)^{\frac{Tv_{\lambda,k}}{2}} \Gamma\left(\frac{v_{\lambda,k}}{2}\right)^{-T} \exp\left(-\left(\frac{1}{v_0} + 0.5 \sum_{t=1}^{T}\left[\ln\left(\lambda_{t,k}^{-1}\right) + \lambda_{t,k}\right]\right)v_{\lambda,k}\right). \quad \text{(A.1)}$$

As in Geweke (1993) we use the Random Walk Metropolis Hastings Algorithm to draw from this conditional distribution. More specifically, for each of the n equations of the VAR, we draw $v_{\lambda,k}^{new} = v_{\lambda,k}^{old} + \bar{c}^{1/2}\epsilon$ with $\epsilon \sim N(0,1)$. The draw is accepted with probability $\frac{G\left(v_{\lambda,k}^{new}|\bar{\lambda}_k\right)}{G\left(v_{\lambda,k}^{old}|\bar{\lambda}_k\right)}$, where $\bar{\lambda}_k = [\lambda_1, \lambda_2, ..., \lambda_T]$, with $\bar{c}$ chosen to keep the acceptance rate around 40%.

A.3   Drawing $G(g_k|\Psi)$

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

A.4   Drawing $G\left(\sigma_{k,t}^2|\Psi\right)$

The conditional posterior $G\left(\sigma_{k,t}^2|\Psi\right)$ is sampled using the Metropolis Hastings algorithm in Jacquier, Polson, and Rossi (1994). Given a draw for B and defining $u_t = Y_t - c - \sum_{l=1}^{L} B_l Y_{t-l}$, the VAR model can be written as $Au_t \times \bar{H}_t^{1/2} = \bar{u}_t$, where $cov\left(\bar{u}_t\right) = \tilde{H}_t$. Here $\bar{H}_t = diag(\lambda_{1,t}, \lambda_{2,t}..)$

and $\tilde{H}_t = diag\left(\sigma^2_{1,t}, \sigma^2_{2,t}, ..\right)$. Conditional on other VAR parameters, the distribution $\sigma^2_{k,t}$ is then given by:

$$f\left(\sigma^2_{k,t}|\sigma^2_{k,t-1}, \sigma^2_{k,t+1}, u_{k,t}\right) = f\left(u_{k,t}|\sigma^2_{k,t}\right) \times f\left(\sigma^2_{k,t}|\sigma^2_{k,t-1}\right) \times f\left(\sigma^2_{k,t+1}|\sigma^2_{k,t}\right)$$

$$= \frac{1}{\sigma_{k,t}} \exp\left(\frac{-u^2_{k,t}}{2\sigma^2_{k,t}}\right) \times \frac{1}{\sigma^2_{k,t}} \exp\left(\frac{-\left(\ln \sigma^2_{k,t} - \mu\right)^2}{2\sigma_{h_k}}\right),$$

where μ and σ_{h_k} denote the mean and the variance of the log-normal density $\frac{1}{\sigma^2_{k,t}} \exp\left(\frac{-\left(\ln \sigma^2_{k,t}-\mu\right)^2}{2\sigma_{h_k}}\right)$.

Jacquier, Polson, and Rossi (1994) suggest using $\frac{1}{\sigma^2_{k,t}} \exp\left(\frac{-\left(\ln \sigma^2_{k,t}-\mu\right)^2}{2\sigma_{h_k}}\right)$ as the candidate generating density with the acceptance probability defined as the ratio of the conditional likelihood $\frac{1}{\sigma_{k,t}} \exp\left(\frac{-u^2_{k,t}}{2\sigma^2_{k,t}}\right)$ at the old and the new draw. This algorithm is applied at each period in the sample.

A.5 Drawing $G\left(A|\Psi\right)$

The conditional posterior $G\left(A|\Psi\right)$ 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:

$$\begin{bmatrix} u_{1,t} \\ u_{2,t} + u_{1,t}\alpha_{21} \\ u_{3,t} + u_{2,t}\alpha_{32} + u_{1,t}\alpha_{31} \\ u_{4,t} + u_{3,t}\alpha_{43} + u_{2,t}\alpha_{42} + u_{1,t}\alpha_{41} \end{bmatrix} = \begin{bmatrix} \left(\sigma^2_{1,t}\frac{1}{\lambda_{1,t}}\right)^{1/2} \varepsilon_{1,t} \\ \left(\sigma^2_{2,t}\frac{1}{\lambda_{2,t}}\right)^{1/2} \varepsilon_{2,t} \\ \left(\sigma^2_{3,t}\frac{1}{\lambda_{3,t}}\right)^{1/2} \varepsilon_{3,t} \\ \left(\sigma^2_{4,t}\frac{1}{\lambda_{4,t}}\right)^{1/2} \varepsilon_{4,t} \end{bmatrix}. \tag{A.2}$$

where $\varepsilon_{k,t} \sim N(0,1), k = 1, .., n$. The second, third and fourth lines give the following system of linear regressions:

$$u_{2,t} = -u_{1,t}\alpha_{21} + \left(\sigma^2_{2,t}\frac{1}{\lambda_{2,t}}\right)^{1/2} \varepsilon_{2,t}$$

$$u_{3,t} = -u_{2,t}\alpha_{32} - u_{1,t}\alpha_{31} + \left(\sigma^2_{3,t}\frac{1}{\lambda_{3,t}}\right)^{1/2} \varepsilon_{3,t} \tag{A.3}$$

$$u_{4,t} = -u_{3,t}\alpha_{43} - u_{2,t}\alpha_{42} - u_{1,t}\alpha_{41} + \left(\sigma^2_{4,t}\frac{1}{\lambda_{4,t}}\right)^{1/2} \varepsilon_{4,t},$$

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

A.6 Drawing $G\left(B|\Psi\right)$

Finally, the posterior distribution of the VAR coefficients is linear and Gaussian, $G\left(B|\Psi\right) \sim N\left(B_{T|T}, P_{T|T}\right)$. We use the Kalman filter to estimate $B_{T|T}$ and $P_{T|T}$ where we account for the

fact that the covariance matrix of the VAR residuals changes through time. The final iteration of the filter delivers $B_{T|T}$ and $P_{T|T}$.

B   Computation of the Marginal Density

B.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:

$$
\begin{align}
X_t &= \bar{\Theta}\bar{\Gamma}_t \tag{B.1}\\
\bar{\Gamma}_t &= \mu + \bar{F}\bar{\Gamma}_{t-1} + \bar{Q}_t^{1/2}\bar{\varepsilon}_t \tag{B.2}\\
\ln \bar{q}_{K,t} &= \ln \bar{q}_{K,t-1} + \bar{v}_t, \tag{B.3}
\end{align}
$$

where $\bar{\varepsilon}_t = \{\bar{\varepsilon}_{1,t}, ..., \bar{\varepsilon}_{N,t}\}$ with $\bar{\varepsilon}_{K,t}$, $K = 1, ..., N$, follows a Student's t-density with v_K degrees of freedom and $\bar{q}_{K,t}$ denotes the diagonal elements of $\bar{Q}$. X_t is observed data, while $\Phi_t = \left(\bar{\Gamma}_t, \bar{q}_{K,t}\right)$ 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):

$$
f(\Phi_t|z_t) = \frac{f(X_t, \Phi_t|z_{t-1})}{f(X_t|z_{t-1})} = \frac{f(X_t|\Phi_t, z_{t-1}) \times f(\Phi_t|z_{t-1})}{f(X_t|z_{t-1})}. \tag{B.4}
$$

Equation B.4 implies that this density can be written as the ratio of the joint density of the data and the states $f(X_t, \Phi_t|z_{t-1}) = f(X_t|\Phi_t, z_{t-1}) \times f(\Phi_t|z_{t-1})$ and the likelihood function $f(X_t|z_{t-1})$ where the latter is defined as:

$$
f(X_t|z_{t-1}) = \int f(X_t|\Phi_t, z_{t-1}) \times f(\Phi_t|z_{t-1}) \, d\Phi_t. \tag{B.5}
$$

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

$$
f(\Phi_t|z_{t-1}) = \int f(\Phi_t|\Phi_{t-1}) \times f(\Phi_{t-1}|z_{t-1}) \, d\Phi_{t-1}. \tag{B.6}
$$

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

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

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

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

$$f\left(\Phi_1|z_1\right) = \frac{f\left(X_1|\Phi_1, z_0\right) \times f\left(\Phi_1|z_0\right)}{f\left(X_1|z_0\right)},$$

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

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 $f\left(\Phi_1|z_0\right)$ via M draws or particles from the Student's t-density using the transition equation of the model. For each draw for the state variables the conditional likelihood $W^m = f\left(X_1|z_0\right)$ 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 $f\left(\Phi_1|z_1\right)$. 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 $\ln lik_t = \ln \frac{\sum_{m=1}^{M} W^m}{M}$.

B.2   Evaluation of the Posterior Density $\Theta(\cdot)$

Consider the following decomposition:

$$\begin{aligned}\Theta\left(\hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) &= \Theta\left(\hat{B}|\hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) \times \Theta\left(\hat{A}|\hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) \times \Theta\left(\hat{g}|\hat{\lambda}, \hat{v}_\lambda, \Xi\right) \\ &\quad \times \Theta\left(\hat{\lambda}|\hat{v}_\lambda, \Xi\right) \times \Theta\left(\hat{v}_\lambda|\Xi\right).\end{aligned}$$

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

1. $\Theta\left(\hat{B}|\hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right)$. This is a complete conditional density with a known form: a Normal with mean and variance that can be calculated via the Kalman filter. The evaluation in done via an additional Gibbs sampler that draws from:

   (a) $\Theta\left(B_i|\hat{A}, \hat{\lambda}, \hat{v}_\lambda, \hat{g}, \Xi_j\right)$

   (b) $\Theta\left(\Xi_i|\hat{A}, \hat{\lambda}, \hat{v}_\lambda, \hat{g}, B_j\right)$

   where the subscripts i and j respectively indicate current and previous draws.

   After a burn-in period, $\Theta\left(\hat{B}|\hat{A}, \hat{\lambda}, \hat{v}_\lambda, \hat{g}, \Xi\right) \approx \frac{1}{I}\sum_{i=1}^{I} \Theta\left(\hat{B}_i|\hat{A}, \hat{\lambda}, \hat{v}_\lambda, \hat{g}, \Xi_j\right)$ where I is the number of the retained simulations.

2. $\Theta\left(\hat{A}|\hat{\lambda}, \hat{v}_\lambda, \hat{g}, \Xi\right) = \int \Theta\left(\hat{A}|\hat{\lambda}, \hat{v}_\lambda, B, \hat{g}, \Xi\right) \times \Theta\left(B|\hat{\lambda}, \hat{v}_\lambda, \hat{g}, \Xi\right) dB$. 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) $\Theta\left(A_i|\hat{g}, \hat{\lambda}, \hat{v}_\lambda, B_j, \Xi_j\right)$

   (b) $\Theta\left(B_i|\hat{g}, \hat{\lambda}, \hat{v}_\lambda, A_j, \Xi_j\right)$

   (c) $\Theta\left(\Xi_i|B_j, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, A_j\right)$

After a burn-in period, $\Theta\left(\hat{A}|\hat{\lambda}, \hat{v}_\lambda, \hat{g}, \Xi\right) \approx \frac{1}{I}\sum_{i=1}^{I} \Theta\left(\hat{A}_i|\hat{\lambda}, \hat{v}_\lambda, \hat{g}, B_j, \Xi_j\right)$ where $\Theta\left(A_i|\hat{\lambda}, \hat{v}_\lambda, \hat{g}, B_j, \Xi_j\right)$ has a Normal density and I is the number of the retained simulations.

3. $\Theta\left(\hat{g}|\hat{\lambda}, \hat{v}_\lambda, \Xi\right) = \int \left[\int \Theta\left(\hat{g}|\hat{\lambda}, \hat{v}_\lambda, \hat{B}, \hat{A}, \Xi\right) \times \Theta\left(\hat{A}|\hat{\lambda}, \hat{v}_\lambda, \hat{B}, \Xi\right) dA\right] \times \Theta\left(\hat{B}|\hat{\lambda}, \hat{v}_\lambda, \Xi\right) d\hat{B}$ where the bracket term $\int \Theta\left(\hat{g}|\hat{\lambda}, \hat{v}_\lambda, \hat{B}, \hat{A}, \Xi\right) \times \Theta\left(\hat{A}|\hat{\lambda}, \hat{v}_\lambda, \hat{B}, \Xi\right) dA = \Theta\left(\hat{g}|\hat{\lambda}, \hat{v}_\lambda, \hat{B}, \Xi\right)$. This term can be approximated using an additional Gibbs run that samples from the following conditionals:

   (a) $\Theta\left(g_i|\hat{\lambda}, \hat{v}_\lambda, B_j, A_j, \Xi_j\right)$

   (b) $\Theta\left(B_i|g_j, \hat{\lambda}, \hat{v}_\lambda, A_j, \Xi_j\right)$

   (c) $\Theta\left(A_i|B_j, g_j, \hat{\lambda}, \hat{v}_\lambda, \Xi_j\right)$

   (d) $\Theta\left(\Xi_i|A_j, B_j, g_j, \hat{\lambda}, \hat{v}_\lambda\right)$

After a burn-in period, $\Theta\left(\hat{g}|\hat{\lambda}, \hat{v}_\lambda, \Xi\right) \approx \frac{1}{I}\sum_{i=1}^{I} \Theta\left(\hat{g}_i|\hat{\lambda}, \hat{v}_\lambda, B_j, A_j, \Xi_j\right)$ where this is an inverse Gamma pdf and I is the number of the retained simulations.

4. $\Theta\left(\hat{\lambda}|\hat{v}_\lambda, \Xi\right)$. As in step 3 above, this term can be approximated by a Gibbs run that draws from the following densities:

   (a) $\Theta\left(\lambda_i|g_j, \hat{v}_\lambda, B_j, A_j, \Xi_j\right)$

   (b) $\Theta\left(g_i|\lambda_j, \hat{v}_\lambda, B_j, A_j, \Xi_j\right)$

   (c) $\Theta\left(B_i|g_j, \lambda_j, \hat{v}_\lambda, A_j, \Xi_j\right)$

   (d) $\Theta\left(A_i|B_j, g_j, \lambda_j, \hat{v}_\lambda, \Xi_j\right)$

   (e) $\Theta\left(\Xi_i|A_j, B_j, g_j, \lambda_j, \hat{v}_\lambda\right)$

After a burn-in period, $\Theta\left(\hat{\lambda}|\hat{v}_\lambda, \Xi\right) \approx \frac{1}{I}\sum_{i=1}^{I} \Theta\left(\hat{\lambda}_i|\hat{v}_\lambda, g_j, B_j, A_j, \Xi_j\right)$ which has a Gamma pdf and I is the number of the retained simulations.

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

$$\Theta\left(\hat{v}_\lambda|\Xi\right) = \frac{E_1\left(\alpha\left(v_\lambda, \hat{v}_\lambda|B, A, \lambda, \Xi\right) q\left(v_\lambda, \hat{v}_\lambda|B, A, \lambda, \Xi\right)\right)}{E_2\left(\alpha\left(\hat{v}_\lambda, v_\lambda|B, A, \lambda, \Xi\right)\right)},$$

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 $q(v_\lambda^{old}, v_\lambda^{new})$ is the candidate density. The numerator term can be approximated by averaging the quantity from the main MCMC run:

$$\alpha\left(v_\lambda^j, \hat{v}_\lambda | B, A, \lambda, \Xi\right) q\left(v_\lambda^j, \hat{v}_\lambda | B, A, \lambda, \Xi\right),$$

where j indexes the MCMC draws. The denominator term requires an additional Gibbs sampler as $\alpha\left(\hat{v}_\lambda, v_\lambda^j | B, A, \lambda, \Xi\right)$, the acceptance probability, is conditioned on the posterior mean $\hat{v}_\lambda$. This sampler draws from each posterior density conditioned on $\hat{v}_\lambda$, and then draws from the candidate density $v_\lambda^j \sim q\left(\hat{v}_\lambda, v_\lambda | B, A, \lambda, \Xi\right)$. The average acceptance probability produces an estimate of the denominator.

C   Monte Carlo Analysis of Stochastic Volatility Estimation

The model of Primiceri (2005) is by now the benchmark for estimating VAR models with stochastic volatility. This section presents results from a Monte Carlo exercise in order to illustrate the consequences of model misspecification, which may arise when the true data-generating process features both stochastic volatility and Student's t errors, while the estimation ignores Student's t errors. We simulate 300 data-sets using a bi-variate TVARSVOL model. For each data-set, we simulate 3000 observations and retain the last 250 to remove any effect caused by the initial values. The parameter values, as listed below, are taken from a bi-variate VAR featuring the quarterly growth rates of GDP (denoted as 'variable 1' below) and price level (denoted as 'variable 2' below) of the United States between 1950 and 2013. We intend our simulated data-sets to bear the features of the dynamics of the two major variables:

$$\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = \begin{bmatrix} 0.64 \\ 0.13 \end{bmatrix} + \begin{bmatrix} 0.32 & 0.05 \\ 0.06 & 0.64 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \pi_{t-1} \end{bmatrix} + \begin{bmatrix} 0.08 & -0.22 \\ 0.02 & 0.14 \end{bmatrix} \begin{bmatrix} y_{t-2} \\ \pi_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t^y \\ \varepsilon_t^\pi \end{bmatrix}, \quad \text{(C.1)}$$

with the following variance-covariance structure:

$$cov\begin{pmatrix} \varepsilon_t^y \\ \varepsilon_t^\pi \end{pmatrix} = \Sigma_t = \tilde{A}^{-1} \widehat{H}_t \tilde{A}^{-1\prime},$$

$$\widehat{H}_t = \begin{pmatrix} \sigma_{1,t}^2 & 0 \\ 0 & \sigma_{2,t}^2 \end{pmatrix},$$

where $\ln \sigma_{k,t} = \ln \sigma_{k,t-1} + s_{k,t}$, $var(s_k) = g_k$. Denote $u_t = [u_t^y \ u_t^\pi]'$ as the vector of orthogonalised shocks, which are iid and follow Student's t distribution with degrees of freedom m_1 and m_2, respectively. We assume the variance of the volatility shocks to be $g_k = 0.0001$ for both variables. The lower triangular matrix $\tilde{A}$ is assumed to be [1 0; 0.1 1]. As for the degrees of freedom, we assume $m_1 = 4$, $m_2 = 10$. For each of the 300 data-sets we estimate a VARSVOL

and a TVARSVOL model. Given the estimated volatility paths, we first compute the *median* of the posterior distribution, then calculate how different these medians are from the true volatility processes. We express the differences in percentage points.

Figure 6: Stochastic Volatility Estimates from the TVARSVOL Model
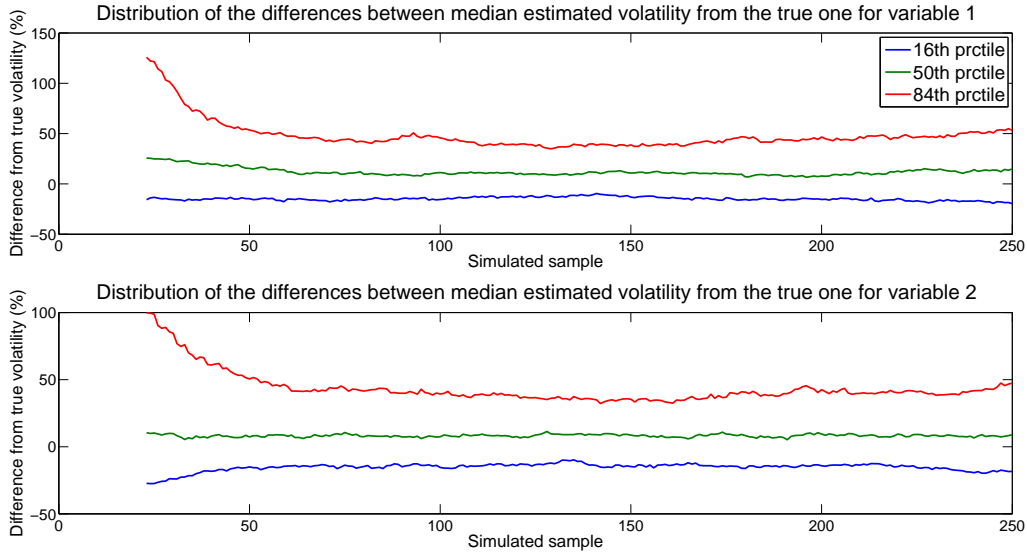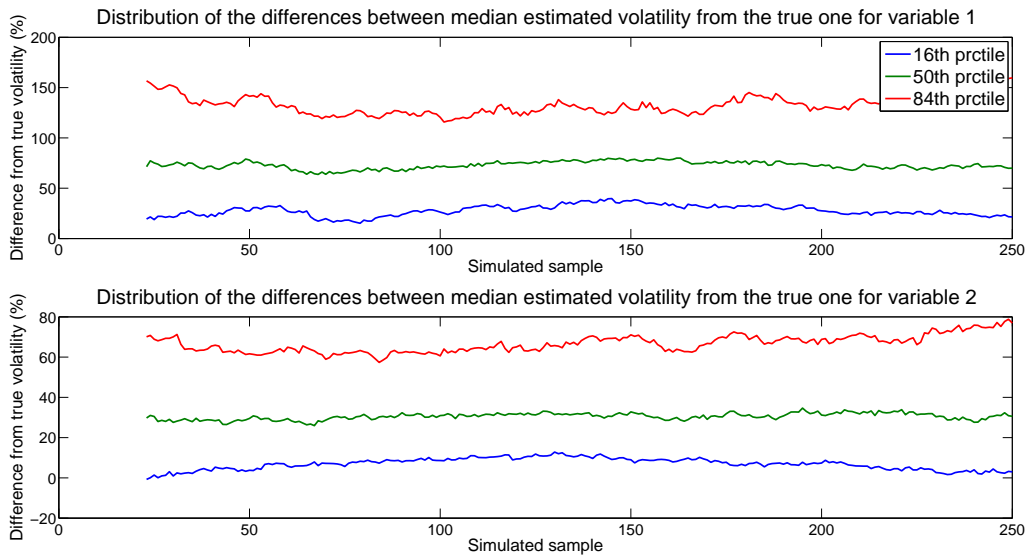

Figure 7: Stochastic Volatility Estimates from the VARSVOL Model

We then collect and sort these percentage differences across the 300 simulated data-sets and present them in Figures 6 and 7, as well as in Tables 11 and 12. A positive percentage points indicates over-estimation of the true volatility. At the 68% interval, we find strong evidence that the differences between the estimated median volatilities and the true ones are significantly positive for VARSVOL but not for TVARSVOL. At the 50th percentile, the VARSVOL model generates median volatility estimates that are 70 percentage points larger for variable 1, and 30 percentage points larger for variable 2. Overall, the simulation results suggest that the

32

VARSVOL model over-estimates the underlying true volatility, whereas TVARSVOL model provides more precise estimates of it.

Table 11: Stochastic Volatility Estimates : Variable 1

selected data point	TVARSVOL (% difference of the median estimated volatility from the true volatility)			VARSVOL (% difference of the median estimated volatility from the true volatility)		
	16th prctile	50th prctile	84th prctile	16th prctile	50th prctile	84th prctile
50	-12.03	17.82	59.20	30.34	74.21	136.78
75	-14.47	14.54	46.72	16.33	64.42	122.78
100	-14.13	13.84	47.86	28.58	71.40	126.67
125	-11.90	12.65	50.81	29.46	75.31	125.78
150	-11.65	15.85	48.28	35.22	79.73	130.59
175	-12.56	14.48	51.58	33.96	76.55	132.39
200	-13.75	14.10	46.32	30.28	72.55	126.55
225	-12.74	15.33	51.55	25.60	73.02	144.70

Note: The table shows the distribution of the *percentage differences* between the estimated median stochastic volatility and the true volatility path over the 300 simulated data-sets. The numbers correspond to the upper panels of Figures 6 and 7. Simulation details are reported in Section C of the Appendix.

Table 12: Stochastic Volatility Estimates : Variable 2

selected data point	TVARSVOL (% difference of the median estimated volatility from the true volatility)			VARSVOL (% difference of the median estimated volatility from the true volatility)		
	16th prctile	50th prctile	84th prctile	16th prctile	50th prctile	84th prctile
50	-14.25	8.46	57.94	3.26	28.83	63.47
75	-14.64	11.66	50.44	7.09	30.25	61.81
100	-15.38	10.96	46.32	10.15	30.17	62.51
125	-11.43	10.24	48.12	10.29	30.26	63.21
150	-11.35	10.39	41.79	9.96	32.35	68.12
175	-13.67	11.91	40.93	8.59	30.81	68.43
200	-12.77	11.20	47.51	6.99	31.91	68.27
225	-12.96	12.53	47.46	4.96	31.37	67.75

Note: The table shows the distribution of the *percentage differences* between the estimated median stochastic volatility and the true volatility path over the 300 simulated data-sets. The numbers correspond to the lower panels of Figures 6 and 7. Simulation details are reported in Section C of the Appendix.

D   Point Forecast Evaluation for the US data

The point forecast is obtained as the mean of the the forecast density. Table 13 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 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.

Table 13: RMSE relative to the BVAR model

|	1M	3M	6M	12M
IP Growth				
BVAR	*0.538*	*0.637*	*0.682*	*0.710*
TVARSVOL	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				
BVAR	*0.137*	*0.166*	*0.179*	*0.193*
TVARSVOL	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 Return				
BVAR	*2.796*	*3.360*	*3.529*	*3.584*
TVARSVOL	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				
BVAR	*0.024*	*0.046*	*0.069*	*0.097*
TVARSVOL	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

Notes: The numbers are relative to the BVAR and computed from the average RMSE values obtained from the 490 recursive estimations for each of the four models (1960 estimated models in total). Each recursive estimation uses 13 lags and 11,000 iterations. The first line of each panel reports the average RMSE of the benchmark BVAR model.

E Further sensitivity analysis

We consider alternative orderings of the variables in the TVARSVOL. Figure 8 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 main text IP is ordered first, 'order1', 'order2' and 'order3' 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 2.2. 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 centred around a value of DOF less than 10 in all cases.

We also consider different priors to the variance of the shocks g_k for TVARSVOL. In the main text, we follow Cogley and Sargent (2005) by proposing an inverse-gamma distribution $p(g_k) \sim IG(D_0, T_0)$, where $T_0 = 1$ and $D_0 = 0.0001$ are the degrees of freedom and scale parameter. As sensitivity checks, we loosen the prior scale by ten and hundred times such that the new scales are $D_0' = 0.001$ and $D_0'' = 0.01$ respectively. Figure 9 displays the posterior distribution of g_k under the three different prior scales. We find evidence that the posterior

Figure 8: Sensitivity of the DOF Posterior to Alternative Orderings
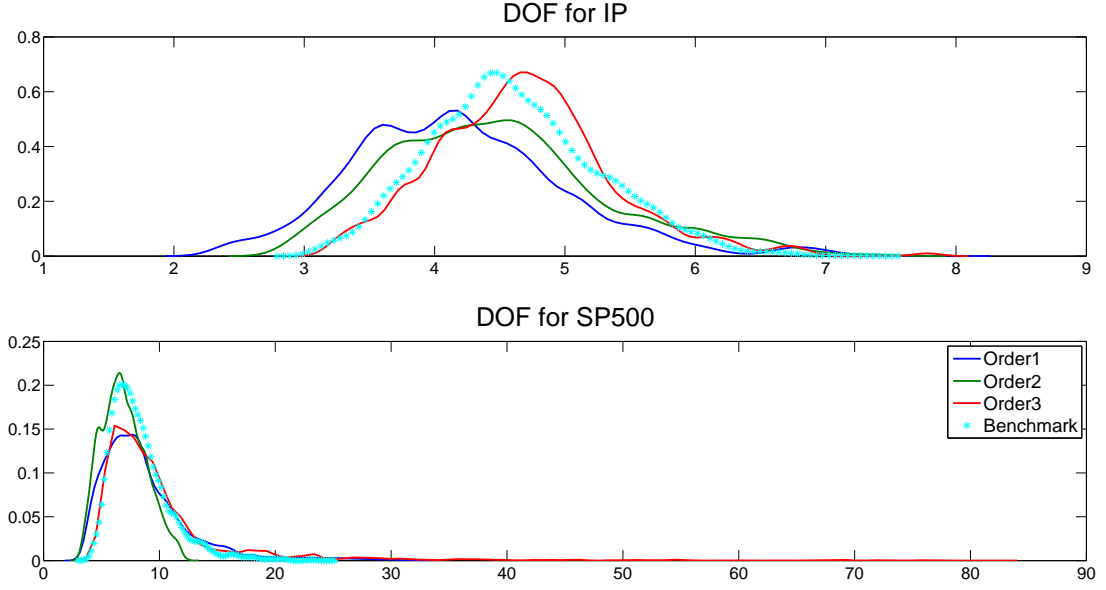
distributions for g_k of all variables are quite robust against the prior of the scale parameter.

F Time-Varying Parameter VAR with Stochastic Volatility

The TVP-VARSVOL model with three variables and two lags can be written as follows:

$$Y_t = c_t + B_{1,t}y_{t-1} + B_{p,t}y_{t-2} + u_t \qquad t = 1, \dots, T. \tag{F.1}$$

where u_t are heteroscedastic shocks with a variance-covariance matrix Ω_t defined as:

$$cov\,(u_t) = \Omega_t = A_t^{-1} H_t \left(A_t^{-1} \right)', \tag{F.2}$$

where the time-varying matrices H_t and A_t are defined as:

$$H_t = \begin{bmatrix} h_{1,t} & 0 & 0 \\ 0 & h_{2,t} & 0 \\ 0 & 0 & h_{3,t} \end{bmatrix} \qquad A_t = \begin{bmatrix} 1 & 0 & 0 \\ \alpha_{21,t} & 1 & 0 \\ \alpha_{31,t} & \alpha_{32,t} & 1 \end{bmatrix}. \tag{F.3}$$
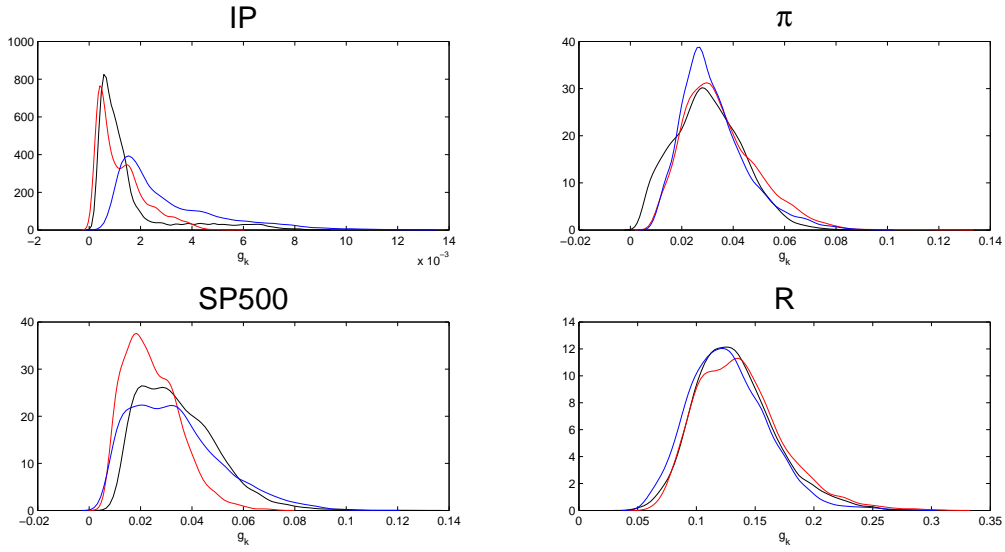
The log of $h_{i,t}$ and $\alpha_{ij,t}$ evolve as random walks:

$$\ln h_{i,t} = \ln h_{i,t-1} + \nu_{i,t} \qquad \alpha_{ij,t} = \alpha_{ij,t-1} + \tau_t, \tag{F.4}$$

where $var(\nu_{i,t}) = g_i$ and $var(\tau_t) = S$.

The system can be represented more compactly by stacking the right-hand-side coefficients into a vector $\hat{\Phi}_t$ as follows:

Figure 9: Posterior distributions of the variance of shocks

Notes: Black lines represent the posterior distributions of the g_k for each variable when the prior scale is $D_0 = 0.0001$, following Cogley and Sargent (2005). Red lines represent the posterior distributions when $D_0' = 0.001$ whereas blue lines represent $D_0'' = 0.01$.

$$Y_t = X_t' \hat{\Phi}_t + A_t^{-1} H_t \varepsilon_t, \tag{F.5}$$

where the data matrix is written as $X_t' = I_n \otimes \left[1, y_{t-1}', y_{t-2}' \right]$, and $\hat{\Phi}_t = vec\left([c_t, B_{1,t}, B_{2,t}]\right)$, and structural variance-covariance matrix $cov\left(\varepsilon_t\right) = I_n$ (I_n denotes an identity matrix). Following the literature, we model the dynamics of the parameter matrix as a random walk:

$$\hat{\Phi}_t = \hat{\Phi}_{t-1} + \eta_t, \tag{F.6}$$

where the innovation vector η_t follows a multivariate normal distribution with mean zero and variance Q.

The TVP-VARSVOL model F.1 – F.6 is estimated with Bayesian methods as described in Blake and Mumtaz (2012) amongst others.