CReMFi
Centre for Research in Macroeconomics and Finance

ISSN: 2058-5160

Discussion Paper 2

Forecasting with VAR Models: Fat Tails and Stochastic Volatility

Ching-Wai (Jeremy) Chiu (Bank of England)

Haroon Mumtaz (Queen Mary, University of London)

Gabor Pinter (Bank of England)

JEL classification: C11, C32, C52

Keywords: Bayesian VAR, Fat-tails, Stochastic volatility, Great Recession.

Forecasting with VAR Models: Fat Tails and Stochastic Volatility

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

February 19, 2015

Abstract

In this paper, we provide evidence that fat tails and stochastic volatility can be important in improving in-sample fit and out-of-sample forecasting performance. Specifically, we construct a VAR model where the orthogonalised shocks feature Student's t distribution and time-varying variance. We estimate this model using US data on output growth, inflation, interest rates and stock returns. In terms of in-sample fit, the VAR model that features both stochastic volatility and Student's t-distributed disturbances outperforms restricted alternatives that feature either attributes. The VAR model with Student's t disturbances results in density forecasts for industrial production and stock returns that are superior to alternatives that assume Gaussianity, and this difference appears to be 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)

*Chiu: Bank of England, e-mail: jeremy.chiu@bankofengland.co.uk; Mumtaz (Corresponding author): Queen Mary University of London; Pinter: Bank of England, e-mail: gabor.pinter@bankofengland.co.uk.
The authors would like to thank, but do not implicate, Andy Blake, Fabio Canova, Christopher Sims, Pawel Zabczyk, an anonymous referee and seminar participants at the Bank of England and Birkbeck College for useful comments. The views expressed in this paper are those of the authors and should not be held to represent those of the Bank of England.

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 output 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 zero 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)

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

showing 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 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 Student's 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.[2] 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 is defined as

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

[2] In an important paper, Jacquier, Polson, and Rossi (2004) provides a detailed analysis of this issue in a univariate framework.

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

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

for $k = 1, 2, ..n$. As shown by Geweke (1993), assuming a Gamma prior for $\lambda_{k,t}$ of the form $p(\lambda_{k,t}) = \prod_{t=1}^{T} \tilde{\Gamma}(1, v_{\lambda,k})$ leads to a scale mixture of normals for the orthogonal residuals $\tilde{\varepsilon}_t = Au_t$ where $\tilde{\varepsilon}_t = \{\tilde{\varepsilon}_{1,t}, \tilde{\varepsilon}_{2,t}, ..\tilde{\varepsilon}_{n,t}\}$. Note that $\tilde{\Gamma}(a, b)$ denotes a Gamma density with mean a and degrees of freedom b. 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, 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.[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.

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; Sims and Zha, 1998). We assume Normal priors, $p(B) \sim N(B_0, S_0)$, where $B = vec([c; B_j])$, where $j = 1, ..., p$, $B_0 = (x_d' x_d)^{-1}(x_d' y_d)$ 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:

$$y_D = \begin{bmatrix} \frac{diag(\gamma_1 s_1 ... \gamma_n s_n)}{\tau} \\ 0_{n \times (p-1) \times n} \\ ..... \\ diag(s_1 \dots s_n) \\ ..... \\ 0_{1 \times n} \end{bmatrix}, \qquad x_D = \begin{bmatrix} \frac{J_P \otimes diag(s_1 ... s_n)}{\tau} & 0_{np \times 1} \\ 0_{n \times np} & 0_{n \times 1} \\ ..... & \\ 0_{1 \times np} & c \end{bmatrix}, \tag{2.4}$$

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

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:

$$y_S = \frac{diag\left(\gamma_1\mu_1 \dots \gamma_n\mu_n\right)}{\lambda}, \qquad x_S = \left[\frac{\left(1_{1\times p}\right) \otimes diag\left(\gamma_1\mu_1 \dots \gamma_n\mu_n\right)}{\lambda} \quad 0_{n\times 1}\right], \tag{2.5}$$

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) in setting a hierarchical prior on the parameter controlling the degree of freedom of the Student's t distributions $v_{\lambda,n}$ and the weighting vector $\lambda_{k,t}$,

$$p\left(v_{\lambda,n}\right) \sim \tilde{\Gamma}\left(v_0, 2\right) \tag{2.6}$$

$$p\left(\lambda_{k,t}\right) \sim \tilde{\Gamma}\left(1, v_{\lambda,n}\right), \tag{2.7}$$

where $\tilde{\Gamma}$ is a Gamma density with the corresponding shape and scale parameters. 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 2.3, and propose an inverse-gamma distribution, $p\left(g_k\right) \sim IG(D_0, T_0)$, 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 $P(A) \sim N\left(0, 1000\right)$.

2.1.2 The Gibbs Sampling Algorithm

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

1. $G(\lambda_{k,t} \backslash \Psi)$ where Ψ denotes the remaining parameters of the model.
2. $G\left(v_{\lambda,k} \backslash \Psi\right)$
3. $G\left(g_{k,t} \backslash \Psi\right)$
4. $G\left(\sigma^2_{k,t} \backslash \Psi\right)$
5. $G\left(A \backslash \Psi\right)$
6. $G\left(B \backslash \Psi\right)$

The details of each conditional posterior density is provided below:

2.1.2.1 Drawing $G(\lambda_{k,t}\backslash\Psi)$ 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 $\tilde{\varepsilon}_t$.

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

$$G\left(v_{\lambda,k}\backslash\Psi\right) \propto \left(\frac{v_{\lambda,k}}{2}\right)^{\frac{Tv_{\lambda,n}}{2}} \Gamma\left(\frac{v_{\lambda,n}}{2}\right)^{-N} \exp\left(-\left(\frac{1}{v_0} + 0.5\sum_{t=1}^{T}\left[\ln\left(\lambda_{t,n}^{-1}\right) + \lambda_{t,n}\right]\right)v_{\lambda,n}\right). \quad (2.8)$$

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 $\frac{G\left(v_{\lambda,n}^{new}\backslash\lambda_n\right)}{G\left(v_{\lambda,n}^{old}\backslash\lambda_n\right)}$ with c chosen to keep the acceptance rate around 40%.

2.1.2.3 Drawing $G\left(g_k\backslash\Psi\right)$ The conditional posterior of $G\left(g_k\backslash\Psi\right)$ is inverse Gamma as 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 $T + T_0$.

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

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

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

2.1.2.5 Drawing $G\left(A\backslash\Psi\right)$ The conditional posterior $G\left(A\backslash\Psi\right)$ 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:

$$\begin{bmatrix} v_t \\ v_{2t} + v_{1t}\alpha_{21,t} \\ v_{3t} + v_{2t}\alpha_{32,t} + v_{1t}\alpha_{31,t} \\ v_{4t} + v_{3t}\alpha_{43,t} + v_{2t}\alpha_{42,t} + v_{1t}\alpha_{41,t} \end{bmatrix} = \begin{bmatrix} \left(\sigma_{1,t}\frac{1}{\lambda_{1t}}\right)^{1/2}\varepsilon_{1t} \\ \left(\sigma_{2,t}\frac{1}{\lambda_{2t}}\right)^{1/2}\varepsilon_{2t} \\ \left(\sigma_{3,t}\frac{1}{\lambda_{3t}}\right)^{1/2}\varepsilon_{3t} \\ \left(\sigma_{4,t}\frac{1}{\lambda_{4t}}\right)^{1/2}\varepsilon_{4t} \end{bmatrix}. \tag{2.9}$$

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

$$\begin{aligned} v_{2t} &= -v_t\alpha_{21,t} + \left(\sigma_{2,t}\varpi_{2,t}\right)^{1/2} e_{2t} \\ v_{3t} &= -v_{2t}\alpha_{32,t} - v_{1t}\alpha_{31,t} + \left(\sigma_{3,t}\varpi_{3,t}\right)^{1/2} e_{3t} \\ v_{4t} &= -v_{3t}\alpha_{43,t} - v_{2t}\alpha_{42,t} - v_{1t}\alpha_{41,t} + \left(\sigma_{4,t}\varpi_{4,t}\right)^{1/2} e_{4t}, \end{aligned} \tag{2.10}$$

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.

2.1.2.6 Drawing $G(B\backslash\Psi)$ Finally, the posterior distribution of the VAR coefficients is linear and Gaussian, $G(B\backslash\Psi) \sim N\left(B_{T\backslash T}, P_{T\backslash T}\right)$. 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:

$$Y_t = c + \sum_{j=1}^{P} b_j Y_{t-j} + \Sigma_t^{1/2} e_t, \tag{2.11}$$

$$\Sigma_t = A^{-1} H_t A^{-1'} \tag{2.12}$$

$$H_t = diag\left(\sigma_{1t}^2 \left(1/\lambda_1\right) ...\right). \tag{2.13}$$

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

$$\ln G(Y_t) = \ln F\left(Y_t\backslash\hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) + \ln P\left(\hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) - \ln H\left(\hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right), \tag{2.14}$$

where the subscript $\hat{}$ denotes the posterior mean, $F(.)$ denotes the likelihood function, $P(.)$ is the joint prior density, $H(.)$ is the posterior distribution and Ξ denotes the state variables in the model. Equation 2.14 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 straightforward to evaluate. The likelihood and posterior are more involved and described in Section A of the Appendix.

2.1.4 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. 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 orthogonolised shocks are Gaussian and consider a VAR model with stochastic volatility only. This model (VARSVOL) is defined as

$$Y_t = c + B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t, \tag{2.15}$$

where

$$\begin{aligned} \Sigma_t &= A^{-1} H_t A^{-1'} && (2.16) \\ H_t &= diag\left(\sigma_{1t}^2, \sigma_{2t}^2, ..\sigma_{nt}^2\right), && (2.17) \end{aligned}$$

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 orthogonolised residuals follow an independent t distribution (TVAR). This model, therefore, is defined as:

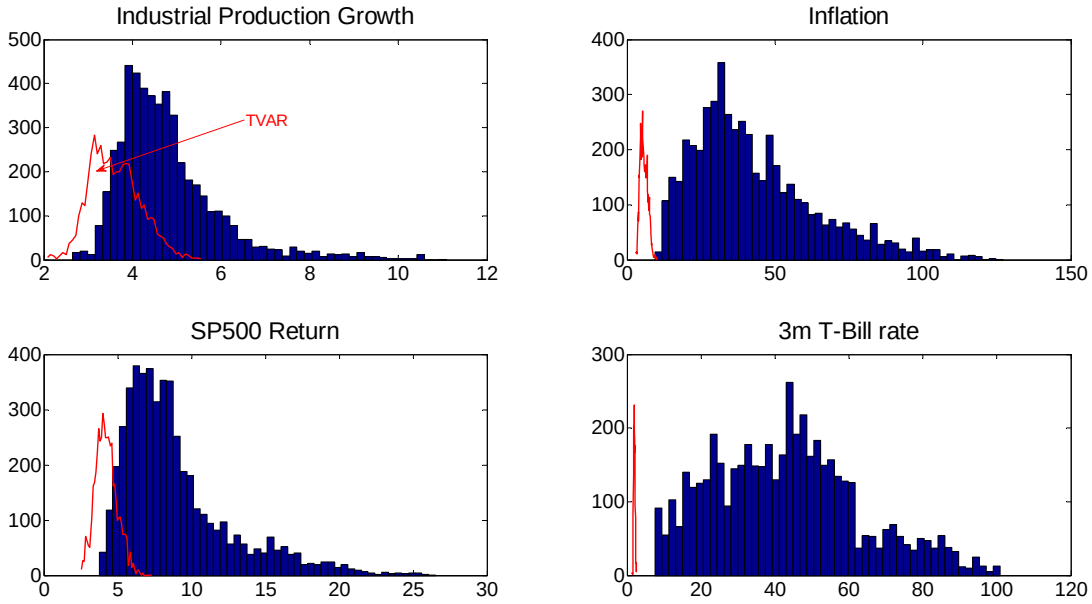
$$\begin{aligned} Y_t &= c + B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t && (2.18) \\ \Sigma_t &= A^{-1} H_t A^{-1'} && (2.19) \\ H_t &= diag\left(\sigma_1^2 \frac{1}{\lambda_{1t}}, \sigma_2^2 \frac{1}{\lambda_{2t}}, ..\sigma_n^2 \frac{1}{\lambda_{nt}}\right). && (2.20) \end{aligned}$$

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

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



Notes: The red lines (blue bars) represent the frequency distribution of the DOF parameters from the TVAR (TVAR-SVOL) 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.

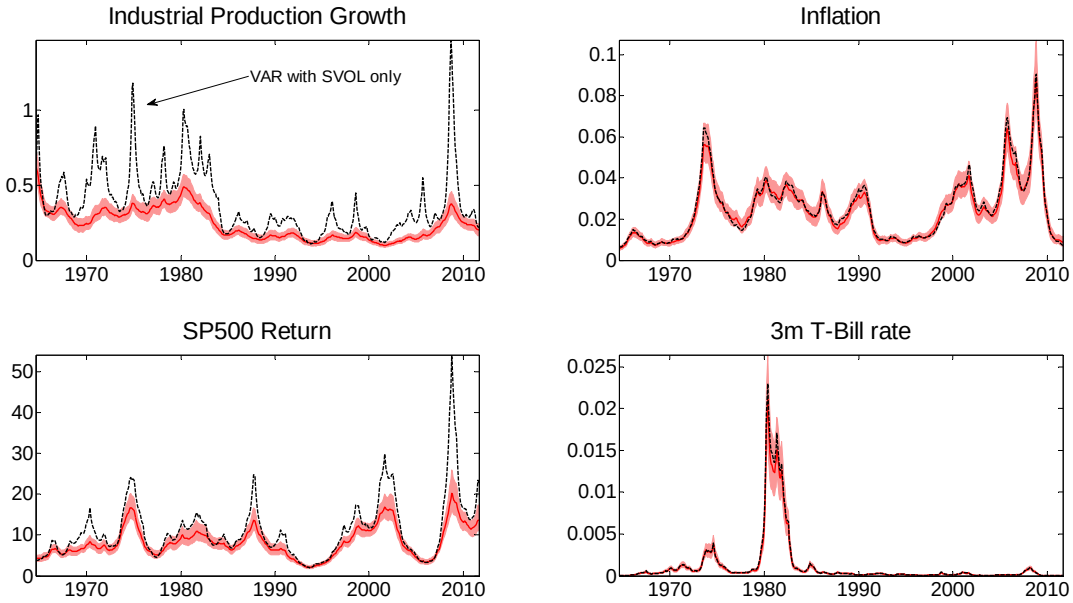
3.1 A Summary of the Posterior

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 orthogonolised 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 orthogonolised 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 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,

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



Notes: The estimated median time-series of volatility in the model with Student's t-distributed shocks and stochastic volatility (in red with 32nd and 68th percentiles of the distribution) and in the model with stochastic volatility only (in black).

thereby overestimating the fatness of tails in the shock distributions.

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.

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.

The implication of these results is that the unit root assumption for stochastic volatility models, widely adopted since Primiceri (2005), while ignoring high-frequency movements of volatility may be a less favourable choice. Section B of the Appendix provides a Monte Carlo analysis 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. The simulation results confirm that (i) the VARSVOL model overestimates the volatility whereas the TVAR-SVOL model can come close to recovering the true DGP, and (ii) this problem of misspecificiation becomes more severe as the data feature more fatness (captured by the DOF parameter) in the shock distribution.

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 lists the estimated log marginal likelihood for each model using the full sample.

Table 1: Marginal Likelihood

Model	Log Marginal Likelihood
TVAR-SVOL	-1725.3
VAR-SVOL	-1757.9
TVAR	-2444.6
BVAR	-2852.2

Notes: Each model includes 13 lags and is estimated using 50.000 iterations (40.000 burns). The computation of the marginal likelihood uses 20.000 iterations (10.000 burns).

The marginal likelihood is estimated via the Chib (1995) method as described in Section 2.1.3 earlier. We use 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-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:

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

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. The point forecast is obtained as the mean of the the forecast density. The recursive estimation delivers 490 forecast densities.

Table 2: RMSE relative to the BVAR model

	1M	3M	6M	12M
IP Growth				
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 Return				
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

Notes: The numbers are computed from the mean values obtained from the 490 rolling estimations for each of the four models (1960 estimated models in total). Each rolling estimation uses 13 lags and 11.000 iterations.

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 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 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.

3.2.2.1 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 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.

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

	1M	3M	6M	12M
IP Growth				
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 Return				
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

Notes: The numbers are computed from the mean values obtained from the 490 rolling estimations for each of the four models (1960 estimated models in total). Each rolling estimation uses 13 lags and 11.000 iterations.

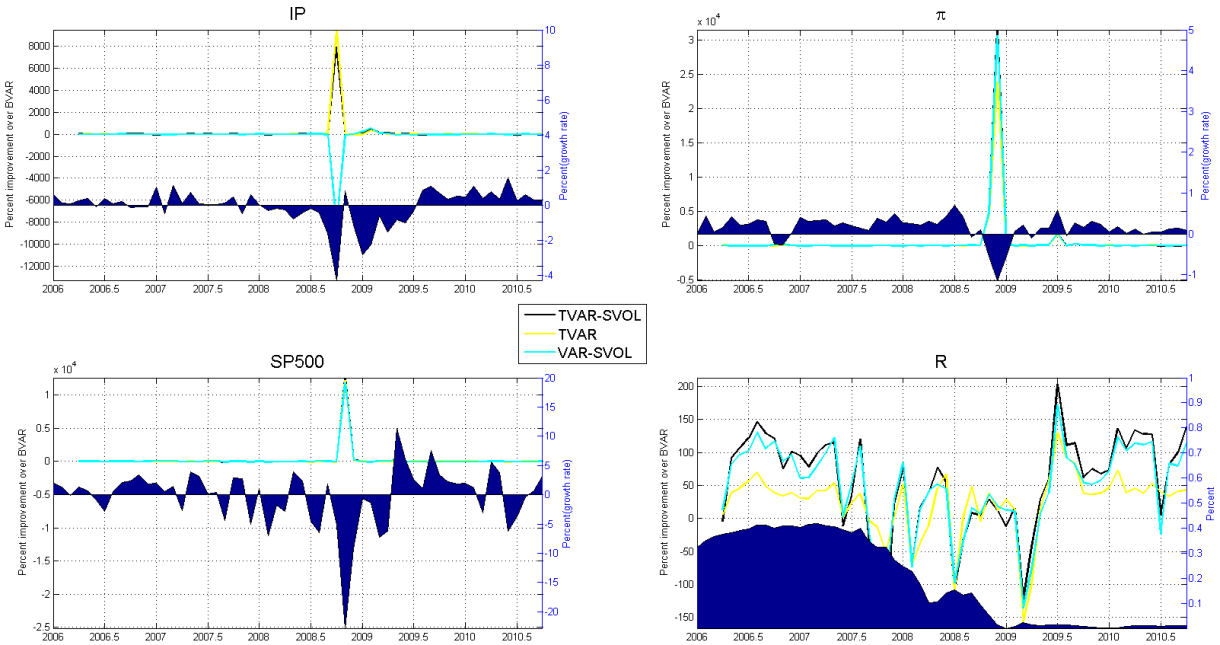
For inflation and interest rates, both stochastic volatility and fat tails appear to be important. The TVAR-SVOL model produces the largest improvement over the BVAR for inflation at the 6 and the 12 month horizon. At the 1 and the 3 month horizon, the benchmark model produces the best performance, with the VARSVOL model delivering the largest improvement over the BVAR at longer horizons.

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 4 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

Table 4: Percentage Improvement in Bivariate Log Scores over the BVAR Model

	1M	3M	6M	12M
IP-π pair				
TVAR-SVOL	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				
TVAR-SVOL	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

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



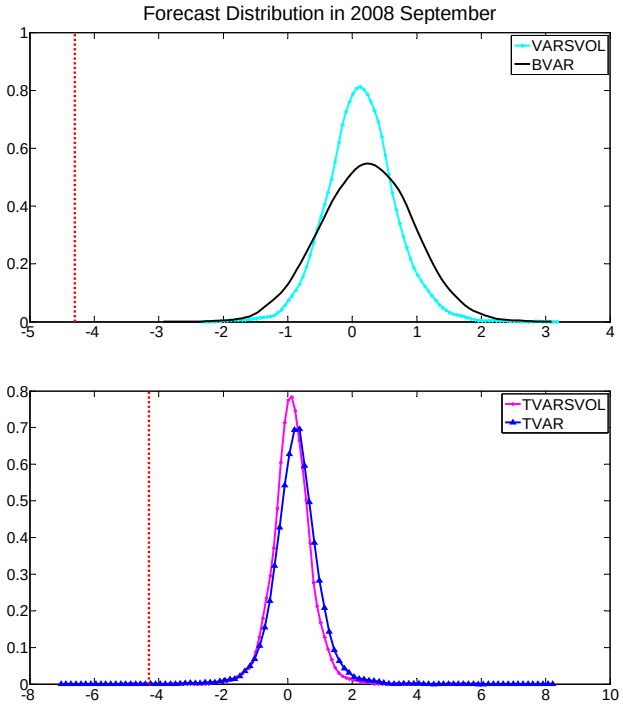
increases forecast accuracy.

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 led 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 3.

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.

Figure 4: 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.

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.

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 4 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 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 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.

3.2.4 Sensitivity Analysis

In this subsection, we ask whether our results are robust to choosing alternative priors for the degree of fatness of shock distribution and to choosing different orderings of the variables. In the benchmark model, we used the prior $v_0 = 20$ for $p(v_{\lambda,n}) \sim \Gamma(v_0, 2)$ 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 presents the estimated log scores (relative to the BVAR) from a version of the benchmark model that uses an alternative prior. The results indicate that for industrial production and stock market returns (the variables for which the orthogonolised 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.

Table 5: Percentage Improvement in Log Scores of the TVARSVOL over the BVAR Model

	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

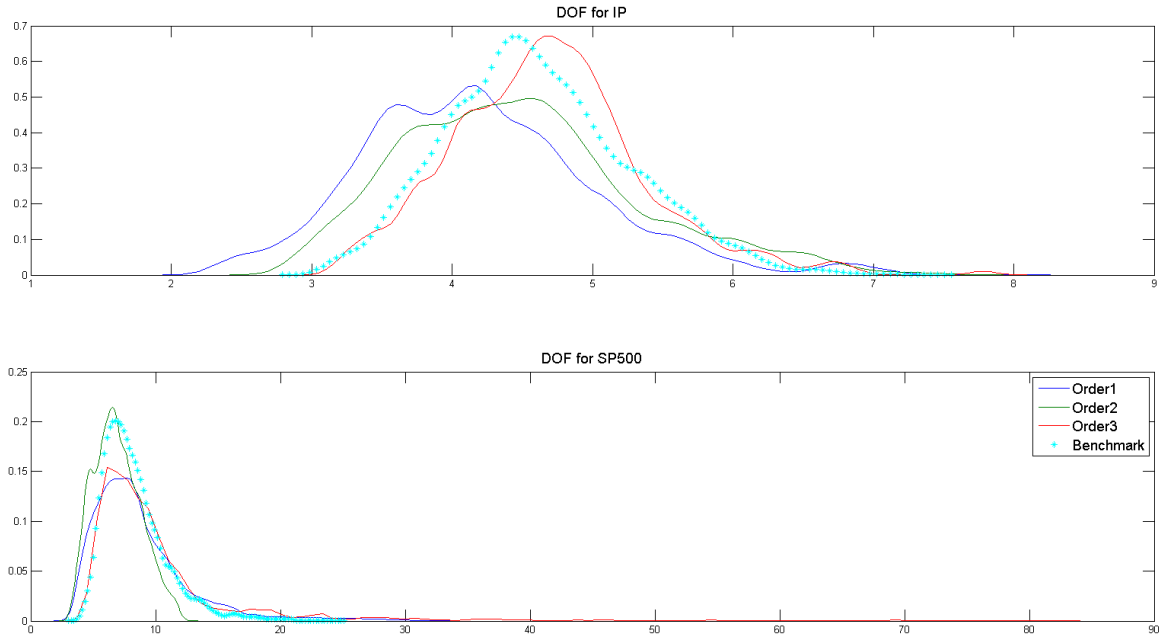
Notes: The numbers are computed from the mean values obtained from the 490 rolling estimations for each of the four models (1960 estimated models in total). Each rolling estimation uses 13 lags and 11.000 iterations.

Moreover, Figure 5 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' 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 centered around a value of DOF less than 10 in all cases.

4 Results for Canada, Germany and the UK

In this section, we carry out further robustness checks by estimating the four model specifications for other countries including the UK, Germany and Canada. Monthly data for the UK

Figure 5: Sensitivity of the DOF Posterior to Alternative Orderings



are taken from the Global Financial Database is the source for the following data: (i) monthly data for the UK covering the period 1959M1-2011M9, including the FTSE all-share Index, CPI, industrial production, the 3-month Treasury Bill yield and the industrial production Index; (ii) monthly data for Germany covering the period 1961M1-2011M9, including the CDAX composite index, CPI and industrial production. The OECD Economic Outlook is the data source for monthly data for Canada covering the period 1961M1-2011M9 and for the short-term German interest rate. 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.[4]

Table 6 presents the results for Canada that are similar to those found for the US. With the exception of inflation, 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 SVOL model, the TVAR tends to outperform in terms of forecasting inflation and stock returns and tends to under-perform in terms of forecasting output growth and interest rates. Over the 1 month horizon, the TVAR-SVOL model comes first in predicting stock returns and the interest rate.

Results for the UK are presented in table 7 providing some interest results. First, the calculated log score values for the interest rates are infinity. The reason for this is as follows: 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

[4]The recursive algorithm starts with the estimation of each of the models up to 1970 January in the case of the UK and up to 1972 January in the case of Canada and Germany. This delivers 490 forecast densities for the UK and 464 forecast densities for Germany and Canada.

Table 6: Forecast Comparison Results for Canada

(a) Relative RMSE values

	1M	3M	6M	12M
IP Growth				
TVAR-SVOL	0.956	0.959	0.963	0.971
TVAR	0.959	0.963	0.968	0.977
VARSVOL	0.949	0.956	0.961	0.969
π				
TVAR-SVOL	1.006	1.013	1.011	1.003
TVAR	1.003	1.011	1.010	1.004
VARSVOL	0.998	1.007	1.004	0.996
Stock Return				
TVAR-SVOL	0.973	0.967	0.972	0.982
TVAR	0.971	0.964	0.969	0.979
VARSVOL	0.978	0.966	0.970	0.980
R				
TVAR-SVOL	0.969	0.942	0.932	0.912
TVAR	0.956	0.924	0.911	0.900
VARSVOL	0.967	0.939	0.926	0.901

(b) Relative Log-scores values

	1M	3M	6M	12M
IP Growth				
TVAR-SVOL	3.239	3.631	2.037	0.627
TVAR	4.025	3.135	1.555	1.081
VARSVOL	5.087	3.261	1.156	1.453
π				
TVAR-SVOL	-14.672	-3.064	-35.606	-31.201
TVAR	12.735	12.585	9.244	1.793
VARSVOL	-23.246	-48.188	-35.477	-19.110
Stock Return				
TVAR-SVOL	41.383	15.236	20.681	3.683
TVAR	24.024	15.749	38.415	18.718
VARSVOL	23.419	6.610	1.814	-29.917
R				
TVAR-SVOL	203.198	66.803	28.984	24.943
TVAR	165.721	49.443	17.478	20.992
VARSVOL	202.175	68.106	32.803	28.563

Note: The mean RMSE and log-scores values obtained from the 466 rolling estimations for each of the four models (1864 estimated models in total). Each rolling estimation uses 13 lags and 11.000 iterations.

produced by the BVAR simply has no probability mass to cover such a jump, leading to a –Inf in log score. Hence, 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. Furthermore, the results highlight the dangers of ignoring the slow-moving component of volatility when using fat-tailed VAR models for forecasting purposes. Outliers such as the aforementioned interest rate jump lead 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), 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 TVAR-SVOL model outperforms the other competing volatility models, with the exception of the 1 month ahead forecast of stock returns and output 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 TVAR-SVOL model tends to dominate other volatility models, at least in the short-run.

As in the case of the UK, the presence of the inf log-score values is due to extreme values. Specifically, the infinite relative improvements of the volatility models in predicting 3 month ahead inflation are related to the extremely large negative observation of January 1991

Table 7: Forecast Comparison Results for the UK

(a) Relative RMSE values

	1M	3M	6M	12M
IP Growth				
TVAR-SVOL	0.955	0.973	0.986	1.00
TVAR	0.960	1.394	1.302	1.234
VARSVOL	0.970	0.980	0.988	1.131
π				
TVAR-SVOL	0.969	0.952	0.944	0.922
TVAR	0.951	1.916	2.010	1.862
VARSVOL	0.930	0.923	0.926	1.083
Stock Return				
TVAR-SVOL	0.985	0.989	0.995	1.001
TVAR	3.060	2.149	1.849	4.309
VARSVOL	0.990	0.991	0.999	2.108
R				
TVAR-SVOL	0.986	0.965	0.968	0.969
TVAR	12.817	7.976	5.795	8.591
VARSVOL	0.965	0.943	0.964	3.029

(b) Relative Log-scores values

	1M	3M	6M	12M
IP Growth				
TVAR-SVOL	28.354	32.823	35.220	36.373
TVAR	26.456	29.045	27.649	26.557
VARSVOL	31.983	30.677	30.949	28.351
π				
TVAR-SVOL	63.984	78.263	68.334	79.877
TVAR	59.271	67.140	51.538	61.716
VARSVOL	63.364	80.726	65.190	72.388
Stock Return				
TVAR-SVOL	109.816	109.552	99.838	105.163
TVAR	101.013	95.805	99.265	106.761
VARSVOL	112.098	108.108	96.489	98.820
R				
TVAR-SVOL	Inf	87.846	64.693	39.441
TVAR	Inf	67.655	45.342	23.607
VARSVOL	Inf	83.237	57.088	25.683

Note: The mean RMSE and log-scores values obtained from the 490 rolling estimations for each of the four models (1960 estimated models in total). Each rolling estimation uses 13 lags and 11.000 iterations.

Table 8: Forecast Comparison Results for Germany

(a) Relative RMSE values

	1M	3M	6M	12M
IP Growth				
TVAR-SVOL	0.946	0.955	0.971	0.983
TVAR	0.948	0.957	0.973	0.984
VARSVOL	0.952	0.956	0.970	0.982
π				
TVAR-SVOL	0.989	0.993	0.997	1.001
TVAR	0.982	0.985	0.990	0.994
VARSVOL	0.998	0.996	1.000	1.003
Stock Return				
TVAR-SVOL	0.973	0.978	0.983	0.986
TVAR	0.978	0.982	0.986	0.989
VARSVOL	0.971	0.980	0.983	0.987
R				
TVAR-SVOL	0.925	0.847	0.848	0.855
TVAR	0.924	0.864	0.868	0.878
VARSVOL	0.900	0.836	0.840	0.853

(b) Relative Log-scores values

	1M	3M	6M	12M
IP Growth				
TVAR-SVOL	23.165	16.330	13.709	20.284
TVAR	21.639	14.465	11.694	18.903
VARSVOL	14.637	12.250	10.145	21.281
π				
TVAR-SVOL	76.055	Inf	79.753	5.820
TVAR	111.258	Inf	121.558	104.502
VARSVOL	128.675	Inf	113.001	117.387
Stock Return				
TVAR-SVOL	53.356	35.559	37.985	41.023
TVAR	40.707	33.865	38.310	37.958
VARSVOL	51.325	34.981	37.943	40.467
R				
TVAR-SVOL	50.090	31.095	31.269	44.232
TVAR	28.100	19.721	17.103	57.359
VARSVOL	38.159	30.262	31.774	52.103

Note: The mean RMSE and log-scores values obtained from the 466 rolling estimations for each of the four models (1864 estimated models in total). Each rolling estimation uses 13 lags and 11.000 iterations.

(−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.

Moreover, table 9 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 TVAR-SVOL and VAR-SVOL models. The TVAR-SVOL has the best in-sample fit in the case of Canada, whereas the VAR-SVOL model comes first in the case of Germany and the UK.

Table 9: Marginal Likelihood

Model	Canada	Germany	UK
TVAR-SVOL	-1857.2	-2428.7	-3201.6
VAR-SVOL	-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 (40.000 burns). The computation of the marginal likelihood uses 20.000 iterations (10.000 burns).

To sum up, our findings suggest that models which account for heteroscadisticity in the error structure can exhibit considerably improved forecast accuracy relative to the baseline BVAR model. These results are consistent with those recently presented by Clark and Ravazzolo (2015). An additional result is that explicitely 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 BVAR 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 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, BVAR models with Gaussian disturbances fail to attach any probability to low values of industrial production 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.

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:

$$X_t = H\Gamma_t \tag{A.1}$$
$$\Gamma_t = \mu + F\Gamma_{t-1} + Q_t^{1/2}\varepsilon_t \tag{A.2}$$
$$\ln q_{Kt} = \ln q_{Kt-1} + v_t, \tag{A.3}$$

where $\varepsilon_t = \{\varepsilon_{1t}, ..\varepsilon_{Nt}\}$ with ε_{Kt} follows a Student's 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):

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

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

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

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

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

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

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

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

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

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

where $f(X_1 \backslash z_0) = \int f(X_1 \backslash \Phi_1, z_0) \times f(\Phi_1 \backslash z_0)\, d\Phi_1$ 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 $\ln lik = \ln f(X_1 \backslash z_0) + \ln f(X_2 \backslash z_1) + ... \ln f(X_T \backslash z_{t-1})$

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 $f\left(\Phi_1 \backslash z_0\right)$ 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 $W^m = f\left(X_1 \backslash z_0\right)$ 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 $f\left(\Phi_1 \backslash z_1\right)$. 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 $\ln lik_t = \ln \frac{\sum_{m=1}^M W^m}{M}$.

A.2 Evaluation of the Posterior Density $H(\cdot)$

Consider the following decomposition:

$$\begin{aligned} H\left(\hat{B}, \hat{A}, \hat{g}, \hat{\lambda}, v_\lambda, \Xi\right) &= H\left(\hat{B} \backslash \hat{A}, \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) \times H\left(\hat{A} \backslash \hat{g}, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) \times H\left(\hat{g} \backslash \hat{\lambda}, \hat{v}_\lambda, \Xi\right) \quad \text{(A.7)} \\ &\quad \times H\left(\hat{\lambda} \backslash \hat{v}_\lambda, \Xi\right) \times H\left(\hat{v}_\lambda, \Xi\right). \quad \text{(A.8)} \end{aligned}$$

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

1. $H\left(\hat{B} \backslash \hat{A}, \hat{\sigma}^2, \hat{\lambda}, v_\lambda, \Xi\right)$. 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) $H\left(B_i \backslash \hat{A}, \hat{\sigma}^2, \hat{\lambda}, v_\lambda, \Xi_j\right)$

   (b) $H\left(\Xi_i \backslash \hat{A}, \hat{\sigma}^2, \hat{\lambda}, v_\lambda, B_i\right)$

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

2. $H\left(\hat{A} \backslash \hat{\sigma}^2, \hat{\lambda}, v_\lambda, \Xi\right) = \int H\left(\hat{A} \backslash \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, B, \Xi\right) \times H\left(B \backslash \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, \Xi\right) dB$. 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) $H\left(A_i \backslash \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, B_j, \Xi_j\right)$

   (b) $H\left(B_i \backslash \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, A_j, \Xi_j\right)$

   (c) $H\left(\Xi_i \backslash B_j, \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, A_j\right)$

   After a burn-in period $H\left(\hat{A} \backslash \hat{\sigma}^2, \hat{\lambda}, v_\lambda, \Xi\right) \approx \frac{1}{J} \sum_{i=1}^J H\left(\hat{A} \backslash \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, B_i, \Xi_j\right)$ where $H\left(\hat{A} \backslash \hat{\sigma}^2, \hat{\lambda}, \hat{v}_\lambda, B_i, \Xi_j\right)$ is the Normal density described above.

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

   (a) $H\left(g_i\backslash\hat{\lambda},\hat{v}_\lambda,B_j,A_j,\Xi_j\right)$
   (b) $H\left(B_i\backslash g_j,\hat{\lambda},\hat{v}_\lambda,A_j,\Xi_j\right)$
   (c) $H\left(A_i\backslash B_j,g_j,\hat{\lambda},\hat{v}_\lambda,\Xi_j\right)$
   (d) $H\left(\Xi_i\backslash A_j,B_j,g_j,\hat{\lambda},\hat{v}_\lambda\right)$

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

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

   (a) $H\left(\lambda_i\backslash g_j,\hat{v}_\lambda,B_j,A_j,\Xi_j\right)$
   (b) $H\left(g_i\backslash\lambda_j,\hat{v}_\lambda,B_j,A_j,\Xi_j\right)$
   (c) $H\left(B_i\backslash g_j,\lambda_j,\hat{v}_\lambda,A_j,\Xi_j\right)$
   (d) $H\left(A_i\backslash B_j,g_j,\lambda_j,\hat{v}_\lambda,\Xi_j\right)$
   (e) $H\left(\Xi_i\backslash A_i,B_j,g_j,\lambda_j,\hat{v}_\lambda\right)$

   After a burn-in period $H\left(\hat{\lambda}\backslash\hat{v}_\lambda\right)\approx\frac{1}{J}H\left(\hat{\lambda}\backslash\sigma_j^2,\hat{v}_\lambda,B_j,A_j,\Xi_j\right)$ which has a Gamma pdf.

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

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

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

$$\alpha\left(v_\lambda^j,\hat{v}_\lambda\backslash B,A,\sigma^2,\lambda,\Xi\right)q\left(v_\lambda^j,\hat{v}_\lambda\backslash B,A,\sigma^2,\lambda,\Xi\right),$$

   where j indexes the MCMC draws. The denominator term requires an additional Gibbs sampler as $\alpha\left(\hat{v}_\lambda,v_\lambda^j\backslash B,A,\sigma^2,\lambda,\Xi\right)$ 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 $v_\lambda^j\sim q\left(\hat{v}_\lambda,v_\lambda\backslash B,A,\sigma^2,\lambda\right)$. The average acceptance probability produces an estimate of the denominator.

B 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 TVAR-SVOL 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 and price level 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:

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

with the following variance-covariance structure:

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

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

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 $g_k = 0.001$ for both variables. The lower triangular matrix 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 VAR-SVOL and a TVAR-SVOL 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.

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 8 and 9. 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 percent points larger for variable 1, and 30 percent points larger for variable 2. Overall, the simulation results suggest that the VAR-SVOL model over-estimates the underlying true volatility, whereas TVARSVOL model provides more precise estimates of it.

Figure 6: Stochastic Volatility Estimates from the VARSVOL Model

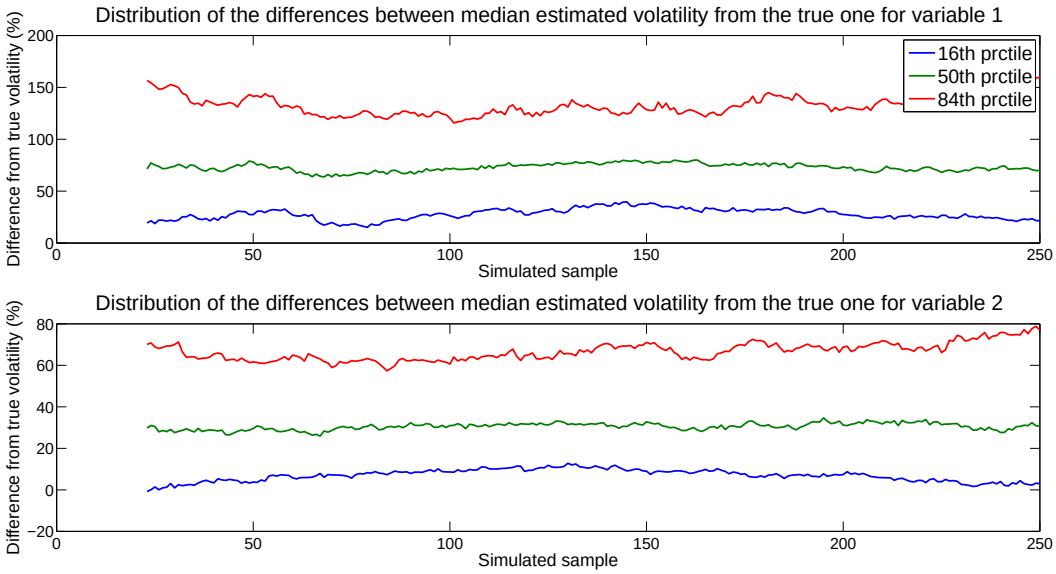


Figure 7: Stochastic Volatility Estimates from the TVARSVOL Model

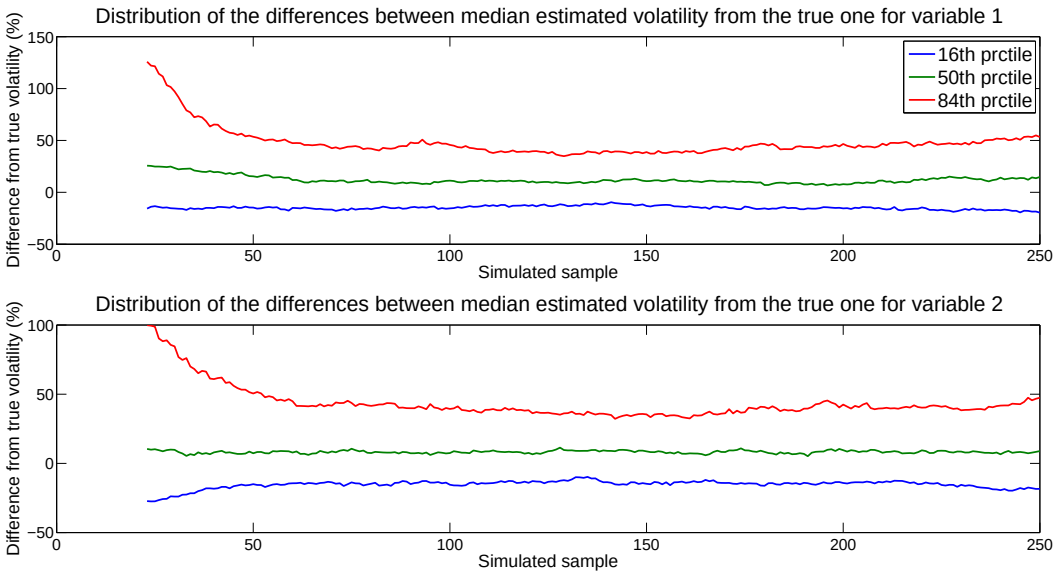


Figure 8: Stochastic Volatility Estimates : Equation 1

	TVARSVOL			VARSVOL		
selected	(% difference of the median			(% difference of the median		
data point	estimated volatility from the true volatility)			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

Figure 9: Stochastic Volatility Estimates : Equation 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

References

Amisano, G., and J. Geweke (2011): "Optimal prediction pools," *Journal of Econometrics*, 164(1), 130–141.

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

Ascari, G., G. Fagiolo, and A. Roventini (2015): "Fat-tail Distributions and Business-cycle Models," *Macroeconomic Dynamics*, Forthcoming.

Banbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25(1), 71–92.

Botev, Z. I., J. F. Grotowski, and D. P. Kroese (2010): "Kernel density estimation via diffusion," *Ann. Statist.*, 38(5), 2916–2957.

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

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

Chib, S., and I. Jeliazkov (2001): "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.

Clark, T. E., and F. Ravazzolo (2015): "Macroeconomic Forecasting Performance Under Alternative Specifications of Time-varying Volatility," *Journal of Applied Econometrics*, Forthcoming.

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.

Curdia, V., M. del Negro, and D. L. Greenwald (2014): "Rare Shocks, Great Recessions," *Journal of Applied Econometrics*, 29(7), 1031–1052.

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.

Elliott, G., and A. Timmermann (2013): "Handbook of Economic Forecasting," vol. 2, Part A. Elsevier.

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," Discussion paper.

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

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

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.

——— (2004): "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): "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98(3), 604–41.

Koop, G. (2003): *Bayesian Econometrics*. Wiley.

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.

Liu, Z., D. F. Waggoner, and T. Zha (2011): "Sources of macroeconomic fluctuations: A regime-switching DSGE approach," *Quantitative Economics*, 2(2), 251–301.

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.

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

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.

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.

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

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

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

Stock, J. H., and M. W. Watson (2012): "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.