A Generalised Stochastic Volatility in Mean VAR

Haroon Mumtaz

Working Paper No. 856 March 2018 ISSN 1473-0278

School of Economics and Finance

Queen Mary
University of London

A generalised stochastic volatility in mean VAR

Haroon Mumtaz*

March 6, 2018

Abstract

This paper introduces a VAR with stochastic volatility in mean where the residuals of the volatility equations and the observation equations are allowed to be correlated. This implies that exogeneity of shocks to volatility is not assumed apriori and structural shocks can be identified ex-post by applying standard SVAR techniques. The paper provides a Gibbs algorithm to approximate the posterior distribution and demonstrates the proposed methods by estimating the impact of financial uncertainty shocks on the US economy.

Key words: VAR, Stochastic volatility in mean, error covariance

JEL codes: C2,C11, E3

1 Introduction

This paper describes the estimation of a VAR with stochastic volatility in mean (VARSVOL) where the residuals of the transition equations are allowed to be correlated with those of the observation equation. This generalises existing VARSVOL models where it is typically assumed that shocks to stochastic volatility are independent of shocks to the endogenous variables.[1] From an economic point of view, such a correlation may reflect shocks that affect the level and conditional variance of a variable. For example, one might expect recessions to be periods of low growth and high output uncertainty. In econometric terms, allowing for such a correlation implies that the model has a structure akin to a reduced form VAR where the structural shocks

*Queen Mary College. Email: h.mumtaz@qmul.ac.uk

[1] Some exceptions to this existing literature are discussed below.

are identified in a second step. This allows the researcher to distinguish amongst uncertainty and level shocks by using SVAR techniques rather than imposing exogeneity of the former apriori.

While generalising the VARSVOL model in this manner makes the state-space more complex, we show that an extended version of the existing MCMC algorithms can be used to approximate the posterior distribution. The algorithm works well on simulated data.

As an application, we estimate a small VAR for the US economy that incorporates financial and macroeconomic variables. The time-varying variances in the model have shocks that are allowed to be contemporaneously correlated both mutually and with the residuals of the observation equations. In order to identify a financial uncertainty shock from these residuals, we use three identification schemes based respectively on short run, medium run and inequality restrictions. Our results indicate that financial uncertainty shocks can have a negative impact on output growth.

The paper is organised as follows: The model is described in section 2 with the estimation algorithm summarised in section 2.1. Details of this Gibbs algorithm are provided in the appendix. Finally, the empirical exercise is described in section 3.

2 Empirical model

We consider the following state-space model:

$$\tilde{h}_t = \alpha + \theta \tilde{h}_{t-1} + \sum_{j=1}^{Q} d_j Z_{t-j} + S^{1/2} \eta_t \tag{1}$$

$$Z_t = c + \sum_{j=1}^{P} \beta_j Z_{t-j} + \sum_{k=1}^{K} b_k \tilde{h}_{t-k} + H_t^{1/2} e_t \tag{2}$$

where Z_t is a matrix of N endogenous variables.

The stochastic volatilities are denoted by $\tilde{h}_t = [h_{1t}, h_{2t}, ..h_{N,t}]$ and $H_t = diag\left(\exp\left(\tilde{h}_t\right)\right)$. The shocks to the transition equation 1 have a variance $S = diag(\tilde{s})$ with $\tilde{s} = [s_1, s_2, .., s_N]$. Note that θ can be a full matrix with the elements of $\tilde{h}_t$ allowed to have a dynamic relationship amongst themselves.

The observation equation of the system is the VAR model in equation 2. As evident, $\tilde{h}_t$ is allowed to have a lagged impact on the endogenous variables.

The disturbances $\varepsilon_t = \begin{pmatrix} \eta_t \\ e_t \end{pmatrix}$ are distributed normally $N(0, \Sigma)$ where the diagonal elements of Σ are restricted to equal 1:

$$\Sigma = \begin{pmatrix} \Sigma_\eta & \Sigma'_{\eta_t e_t} \\ \Sigma_{\eta_t e_t} & \Sigma_{e_t} \end{pmatrix}$$

In other words, the time-varying covariance matrix of the reduced form residuals of the system in equations 1 and 2 can be written as $\Omega_t = \begin{pmatrix} S^{1/2} & 0 \\ 0 & H_t^{1/2} \end{pmatrix} \begin{pmatrix} \Sigma_\eta & \Sigma'_{\eta_t e_t} \\ \Sigma_{\eta_t e_t} & \Sigma_{e_t} \end{pmatrix} \begin{pmatrix} S^{1/2} & 0 \\ 0 & H_t^{1/2} \end{pmatrix}'$. Thus the model allows for correlation between the shocks to the level of the endogenous variables and volatilities.

There are two main differences between the model proposed here and VARSVOL models used in recent papers such as Mumtaz and Surico (n.d.), Mumtaz and Theodoridis (2015b), JO (2014) and Mumtaz and Zanetti (2013). First, the model allows for lagged feedback effects from the endogenous variables to the stochastic volatilties (see also Mumtaz and Theodoridis (2015a)). Second, the covariance beween level shocks and those to second moments is allowed to be non-zero. This implies that in order to identify structural shocks u_t from the $M = 2N$ reduced form disturbances in the system additional assumptions are required. In particular, the structural shocks can be estimated as $u_t = A_{0,t}^{-1}\varepsilon_t$ where $A_{0,t}A'_{0,t} = \Omega_t$. The contemporaneous impact matrix $A_{0,t}$ could be obtained using one of the techniques developed in the large literature on structural VARs. For example, timing restrictions could be incorporated by calculating $A_{0,t}$ as the Cholesky decomposition of Ω_t.

In a recent related contribution, Carriero *et al.* (2016) estimate a VARSVOL model that incorporates both unobserved idiosyncratic volatility and a measure of uncertainty that is common across VAR equations and possibly observed. They show that the time-varying volatility allows them to identify the contemporanoeus impact of level shocks on uncertainty. The current paper differs from Carriero *et al.* (2016), in that it does not attempt to directly estimate the contemporaneous impulse response. Instead, our focus is on the reduced form of the model with identification confined to a second step. The model proposed above is a multivariate extension of stochastic volatility models with leverage considered in Omori *et al.* (2007) and Pitt *et al.* (2014).

2.1 Gibbs sampling algorithm

We approximate the marginal posterior distribution of the parameters and states $B, S, \Sigma, \tilde{h}_t$ using a Gibbs sampling algorithm. While we provide the details of prior and conditional posterior distributions in the appendix, a sketch of the algorithm is provided here. The algorithm samples from the following conditional posterior distributions:

1. $G\left(B|S, \Sigma, \tilde{h}_t\right)$. The conditional posterior distribution of the coefficients $B = vec\left([\alpha, \theta, d_1, .., d_Q, c, \beta_1, .., \beta_P, b_1, .., b_K]\right)$ can be obtained by writing equations 1 and 2 as SUR system with conditionally normal disturbances with covariance matrix Σ. With a normal prior, the conditional posterior of B is also normal. The Kalman filter can be used find the mean and the variance of the conditional posterior taking into account the time-variation in H_t.

2. $G\left(S|B, \Sigma, \tilde{h}_t\right)$. The correlation amongst the disturbances of the transition equation $\tilde{\eta}_t = S^{1/2}\eta_t$ implies that the conditional posterior for the elements of S is non-standard. Thus we use a Metropolis step to sample from this distribution. A candidate density that displays satisfactory performance in simulations is the inverse Gamma (IG) distribution centered at the posterior moments calculated under the assumption that $\tilde{\eta}_t$ are uncorrelated, i.e. $IG\left(v_1, T_1\right)$ where the parameter v_1 is set to $\tilde{\eta}_{it}'\tilde{\eta}_{it} + v_0$ and $T_1 = T_0 + T$ where v_0, T_0 denote prior moments and T is the sample size. In practice, this can also be combined with a IG distribution centered on the previous draw to obtain a mixture proposal density $\varkappa IG\left(v_1, T_1\right) + (1 - \varkappa) IG\left(v\left(S_{j-1}\right), T\left(\bar{V}\right)\right)$ where $v\left(S_{j-1}\right), T\left(\bar{V}\right)$ denotes the parameters of the IG consistent with a mean of S_{j-1} and standard deviation of $\bar{V}$. The latter proposal may be useful if $\tilde{\eta}$ are highly correlated. Note that given $B, \Sigma, \tilde{h}_t$ and a draw of S from the candidate density, the likelihood can be easily calculated with the process described in the appendix.

3. $G\left(\Sigma|B, \tilde{h}_t, S\right)$. Given B and the variances $S, \tilde{h}_t$, the residuals ε_t. The draw of the restricted covariance matrix is obtained via the independence Metropolis algorithm described in Chan and Jeliazkov (2009).

4. $G\left(\tilde{h}_t|\Sigma, B, S\right)$. The observation equation of the state-space system can be written as:

$$\begin{aligned} Z_t - H_t^{1/2}\mu_{e_t|\eta_t} &= c + \sum_{j=1}^{P}\beta_j Z_{t-j} + \sum_{k=1}^{K} b_k\tilde{h}_{t-k} + \tilde{e}_t \\ var\left(\tilde{e}_t\right) &= \Omega_t = H_t^{1/2}\Sigma_{e_t|\eta_t}H_t^{1/2\prime} \end{aligned}$$

where $\mu_{e_t|\eta_t}$ denotes the conditional mean of e_t and $\Sigma_{e_t|\eta_t}$ is the conditional variance:

$$\begin{aligned} \mu_{e_t|\eta_t} &= \eta_t\Sigma_{\eta_t}^{-1}\Sigma'_{\eta_t e_t} \\ \Sigma_{e_t|\eta_t} &= \Sigma_{e_t} - \Sigma_{\eta_t e_t}\Sigma_{\eta_t}^{-1}\Sigma'_{\eta_t e_t} \end{aligned}$$

We treat η_t as a state variable in this step and write the transition equation as

$$F_t = C + \Psi F_{t-1} + N_t$$

where $F_t = \begin{pmatrix} \eta_{t+1} \\ \eta_t \\ \tilde{h}_t \\ . \\ \tilde{h}_{t-k} \end{pmatrix}$. Note that the residual of the transformed observation equation $\tilde{e}_t$ is uncorrelated with N_t. As described in the appendix, we employ a particle Gibbs step (see Andrieu *et al.* (2010) and Lindsten *et al.* (2014)) to sample F_t from its conditional posterior distribution.

We conduct a small Monte-Carlo experiment to evaluate the performance of the algorithm. We generate data from the following DGP

$$\begin{pmatrix} \ln h_{1t} \\ \ln h_{2t} \end{pmatrix} = \begin{pmatrix} 0.85 & -0.1 \\ 0.1 & 0.85 \end{pmatrix}\begin{pmatrix} \ln h_{1t-1} \\ \ln h_{2t-1} \end{pmatrix} + \begin{pmatrix} -0.05 & 0.01 \\ -0.05 & 0.01 \end{pmatrix}\begin{pmatrix} \ln h_{1t-1} \\ \ln h_{2t-1} \end{pmatrix} + \begin{matrix} s_{11}^{1/2}e_{1t} \\ s_{22}^{1/2}e_{2t} \end{matrix}$$

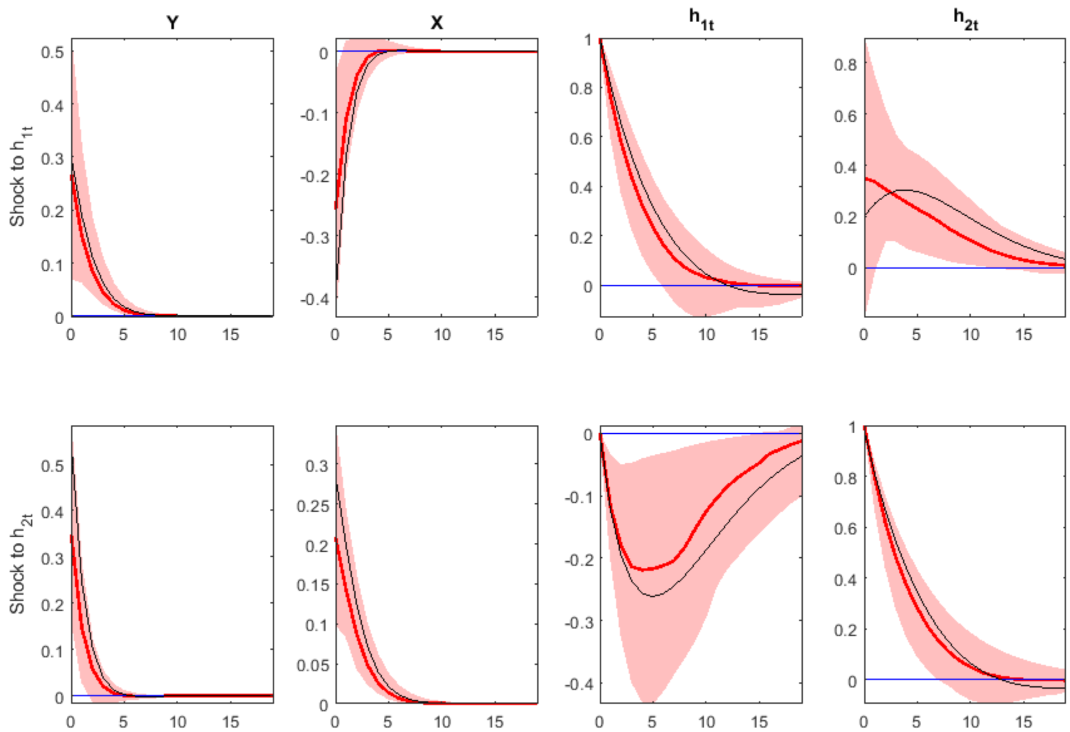

Figure 1: Impulse response to a 1 unit uncertainty shock. Black line represents true responses. The red line and shaded area represent the median estimate anf the 95% error band across the Monte-Carlo replications

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} 0.3 \\ -0.3 \end{pmatrix} + \begin{pmatrix} 0.5 & -0.1 \\ 0.1 & 0.5 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} -0.1 & 0.1 \\ -0.1 & 0.1 \end{pmatrix} \begin{matrix} \ln h_{1t-1} \\ \ln h_{2t-1} \end{matrix} + \begin{pmatrix} h_{1t}^{1/2} e_{3t} \\ h_{2t}^{1/2} e_{4t} \end{pmatrix}$$

$$\begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \\ e_{4t} \end{pmatrix} \tilde{} N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 & 0.3 & -0.4 \\ 0.2 & 1 & 0.6 & 0.2 \\ 0.3 & 0.6 & 1 & -0.2 \\ -0.4 & 0.2 & -0.2 & 1 \end{pmatrix} \right]$$

where $e_{it}, v_{it} \tilde{} N(0,1)$ for $i = 1, 2, .., 4$. We set $s_{11} = 1$ and $s_{22} = 1$. We generate 500 observations and discard the first 100 to remove the effect of initial conditions. The experiment is repeated 50 times. At each iteration we estimate the model using an MCMC run of 10000 iterations, with a burn-in of 5000 iterations. The particle Gibbs step employs 20 particles. For the retained draws, we calculate the response to unit shocks to the equations for h_{1t} and h_{2t} with the residuals orthogonalised using a Cholesky decomposition.

Figure 1 shows a comparison between the true impulse responses and the estimates based on the Monte-Carlo. The true responses are fairly close to the median estimates and lie within the 95% error bands. This provides some evidence that that the algorithm displays a reasonable performance.

3 Empirical results

We use the proposed model to estimate the impact of financial uncertainty shocks on the US economy. The model is estimated using 3 variables: (1) the spread between BAA corporate bond yield and the ten year treasury bill rate (S), (2) quarterly real GNP growth (Y) and (3) quarterly GNP deflator inflation (P). We employ a long sample of data running from 1920Q1 to 2015Q4. This allows us to exploit a larger number of events of high financial volatility aiding in shock identification. Prior to 1983, the data on real GNP and GNP deflator is obtained from from Gordon (1986) with data on subsequent years obtained from the Federal Reserve Bank of St Louis (FRED) database (codes GNPC96 and GNPDEF). The source for the BAA yield is also FRED (code BAA), while the 10 year rate is obtained from Global Financial database.

In terms of model specification, we set the lag lengths P, K, Q to $4, 2, 2$ respectively. The prior distributions are fairly standard and described in detail in the appendix. We employ 150,000 MCMC iterations, discarding the first 25,000 as burn-in with inference based on every 25th remaining draw. The inefficiency factors (available on request) are fairly low providing some evidence in favour of convergence.

Before discussing the identification of the financial uncertainty shock, we consider the reduced form estimates. The posterior median of Σ is displayed in the form of a heat map in Figure 2. It is interesting to note that the contemporaneous relationship between the shocks to h_{1t}, the variance of the shock to the spread, and h_{2t}, the variance of the shock to output growth is positive and relatively large with the 68% highest posterior density interval (HPDI) given by $[0.61, 0.87]$. In contrast, the covariance between shocks to h_{1t},h_{2t} and the residual of the equation for h_{3t} (the variance of the shock to inflation) is imprecisely estimated with the HPDI including a value of 0. The contemporaneous relationship between shocks to h_{1t} and h_{2t} and the residual of the equation for S is estimated to be positive with the HPDI given by $[0.33, 0.59]$ and $[0.27, 0.60]$, respectively. In contrast, the null hypothesis of zero covariance between the volatility residuals and those of the output and inflation equation cannot be rejected. These results are consistent with the view that

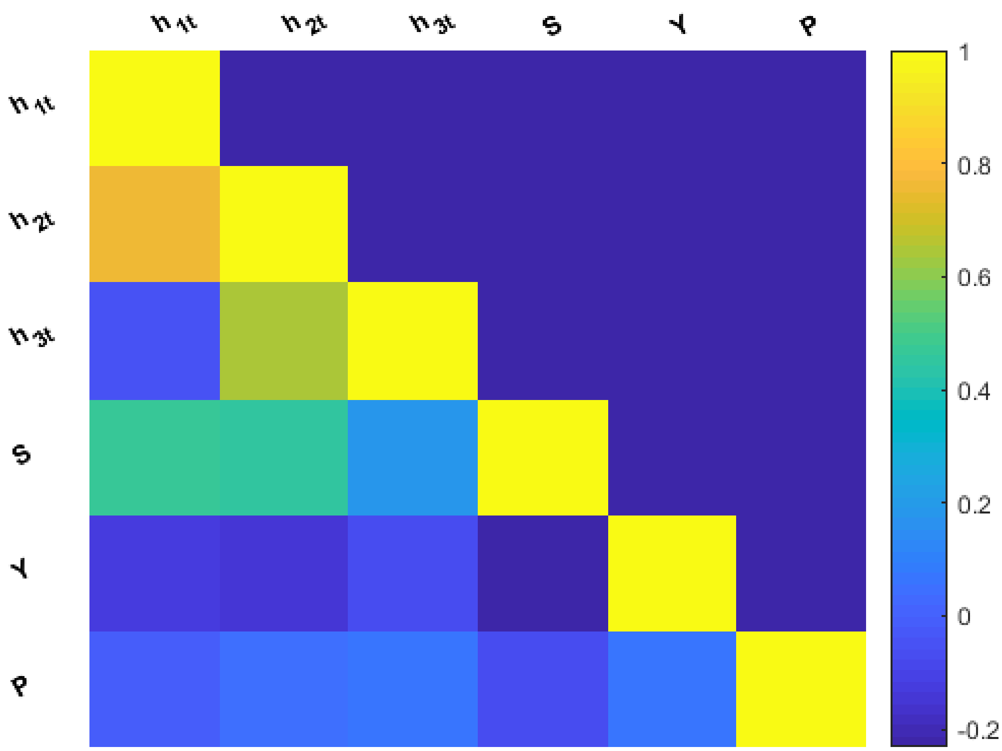

Figure 2: Elements on and below the diagonal of the posterior mean of Σ

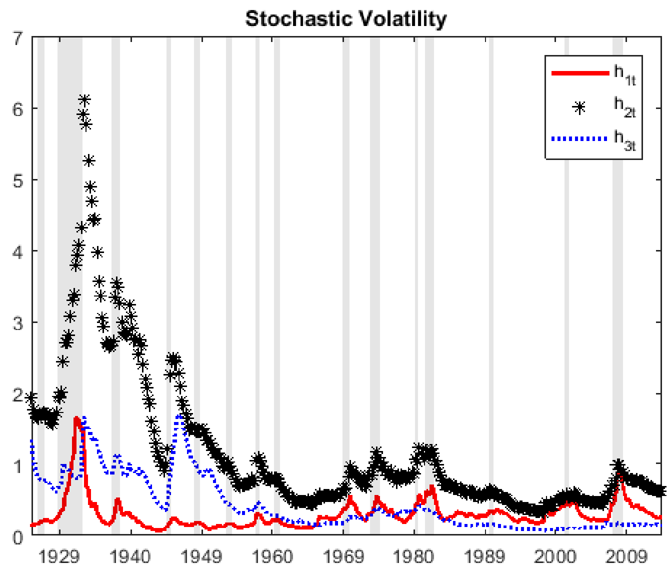

Figure 3: Posterior median estimates of the square root of stochastic volatilities. Shaded areas represent NBER recessions.

uncertainty shocks and financial shocks are closely related.

Figure 3 shows that early 1930s saw the largest peaks in the volatility of shocks to output and the spread, with h_{1t} substantially larger than the remaining variances. In the post-1970 period, episodes of recession appear to be associated with a positive co-movement between h_{1t} and h_{2t} while the volatility of the shock to inflation is estimated to be smoother.

In order to identify financial uncertainty shocks, we adopt three approaches:

1. Cholesky: As a simple benchmark, the contemporaneous impact matrix is calculated using the Cholesky decomposition assuming the ordering $h_{1t}, h_{2t}, h_{3t}, S, Y, P$ with the first shock labelled as financial uncertainty.

2. 'Max FEV': Following Uhlig (2004), we place restrictions on the contribution of the shocks to the forecast error variance (FEV). We start from the observation that, the conditional variance of the spread is driven by a number of structural shocks including a the financial uncertainty shock ε_t^*:

$$h_{1t} = f\left(\varepsilon_t^*, \tilde{\varepsilon}_t\right) \tag{3}$$

Here, $\tilde{\varepsilon}_t$ denotes other structural disturbances (e.g. demand, supply, financial and policy shocks) that can also potentially lead to a change in volatility h_{1t} and are proxied by the innovations in the VAR model. In order, to seperate ε_t^* from $\tilde{\varepsilon}_t$ we assume that the former makes the largest contribution to the FEV of h_{1t} at short and medium horizons (assumed to be up to 20 quarters after the shock). In other words, shocks that explain the bulk of short and medium term movements in the conditional variance of the spread are labelled as financial uncertainty shocks.[2].

3. Sign restrictions: In the spirit of Ludvigson *et al.* (2015), we impose sign restrictions on the shocks ε_t. In particular we assume that the contemporaneous impact matrix implies shocks to h_{1t}, h_{2t} and Y_t that satisfy the following conditions:

    (a) Shocks to h_{1t} (financial uncertainty shocks) and h_{2t} (Macro uncertainty shocks) have a negative correlation with the residual obtained via an AR(1) regression using stock returns. Denoting these correlation coefficients as ρ_1 and ρ_2 and letting$\rho_{1,2} = \sqrt{\rho_1^2 + \rho_2^2}$, this conditions further requires $\rho_1 < -0.05$, $\rho_2 < -0.05$, $|\rho_1| - 2|\rho_2| > 0$ and $\rho_{1,2} > 0.2$. These conditions ensure that uncertainty shocks have a negative correlation with the reduced form shock to returns, with shocks labelled as financial uncertainty displaying a correlation that is larger in magnitude. As discussed in Ludvigson *et al.* (2015), these restrictions are consistent with the view that uncertainty shocks affect risk premia and should, therefore, be correlated with returns.

    (b) Financial uncertainty shocks are restricted to be at least 2 standard deviations larger than their mean at least once during the great depression period (1929-1933) and the recent financial crisis (2007Q4-2009Q2). This restriction is also imposed on the stock market crash in 2007Q4. We assume that shocks to Y during 2007Q4-2009Q2 must be less than two standard deviations in order to rule out draws that imply implausibly large positive output shocks over that period.

Figure 4 displays the posterior median estimates of the financial uncertainty shock under the three identification schemes. The estimates using the Cholesky and max FEV schemes are fairly similar with major peaks during the 1930s, 1950s, 1980s and the recent financial crisis. The estimate based on narrative

[2] To calculate the implied structural impact matrix we use the linear approximation to the impulse responses. As shown in Uhlig (2004), the maximisation can be written as an eigenvector eigenvalue problem and an analytical solution is available.

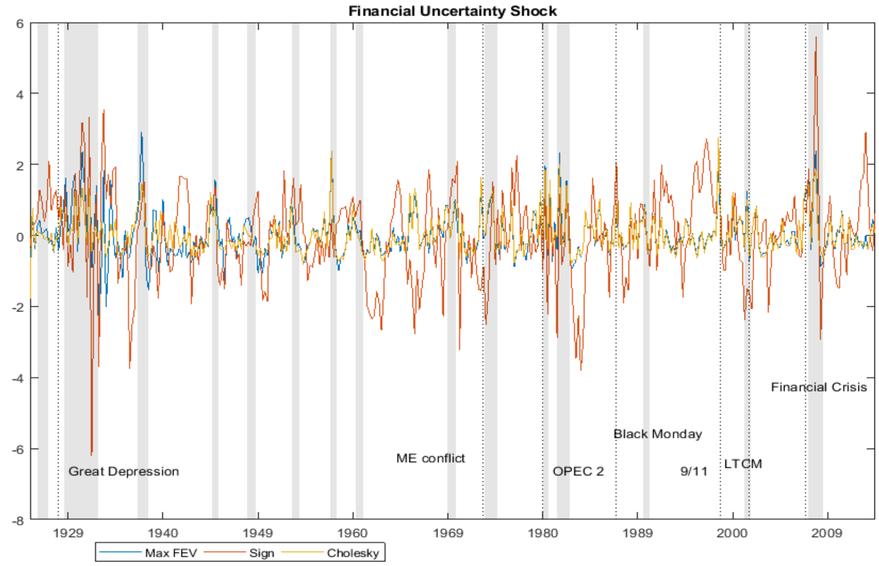

Figure 4: Estimated financial uncertainty shocks (posterior median). The shaded area are recessions while vertical lines indicate key historical events. 'ME' conflict refers the Arab Israeli war in 1973.

sign restrictions is more volatile and reaches its largest values during the recent recession.

The estimated impulse response to the financial uncertainty shock based on the three identification schemes are shown in Figure 5. The shock leads to a rise in conditional output volatility but appears to have little impact on inflation variance. The spread rises in response with the largest increase estimated when sign restrictions are used. The median response of output growth is negative under all identification schemes. However, the error bands for this response are large when sign restrictions are used for identification.

Figure 6 shows that the contribution of the shock identified via sign restrictions to the FEV is small in all cases except the spread. This might suggest that the narrative restrictions are insufficient to fully seperate uncertainty from financial shocks. In the case of max FEV identification, the contribution to volatility and output growth is estimated to be larger.

In summary, the impulse responses and FEV decomposition indicates that, unsurprisingly, identification assumptions play a big role in determining the results. However, on average across identification schemes, there is moderately strong evidence that financial uncertainty shocks have negative real and financial consequences.

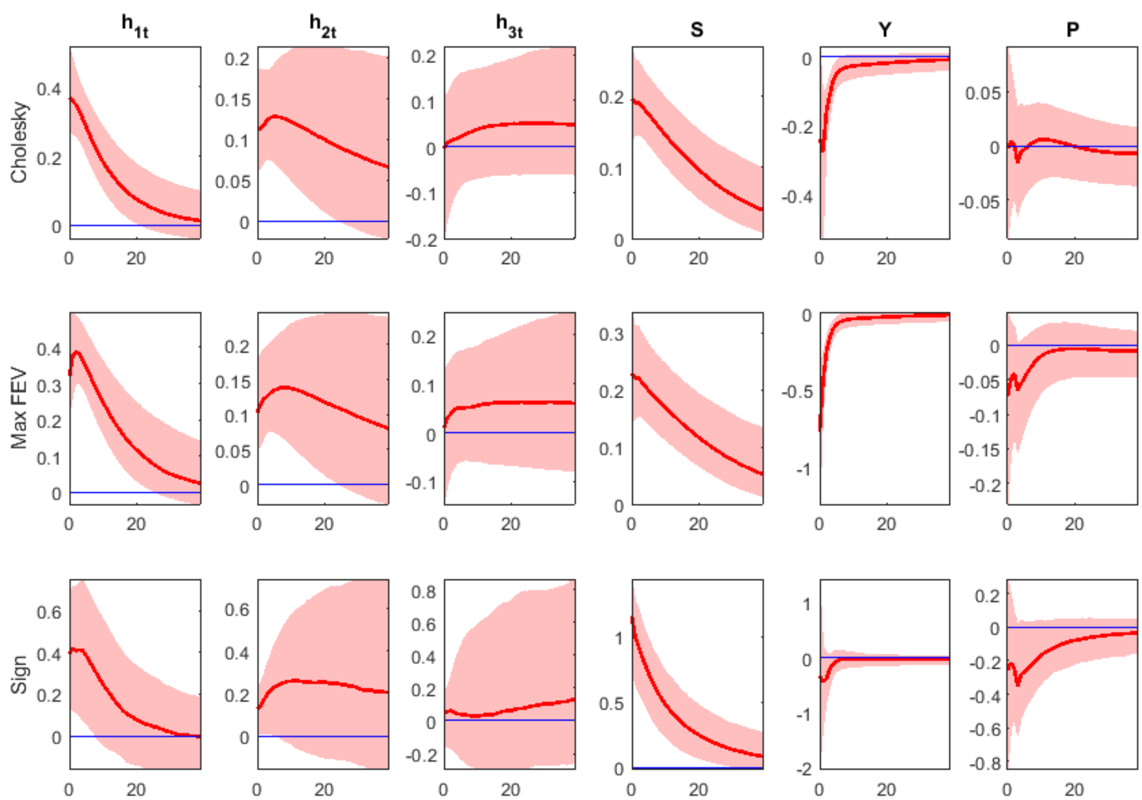

Figure 5: Impulse response to a financial uncertainty shock

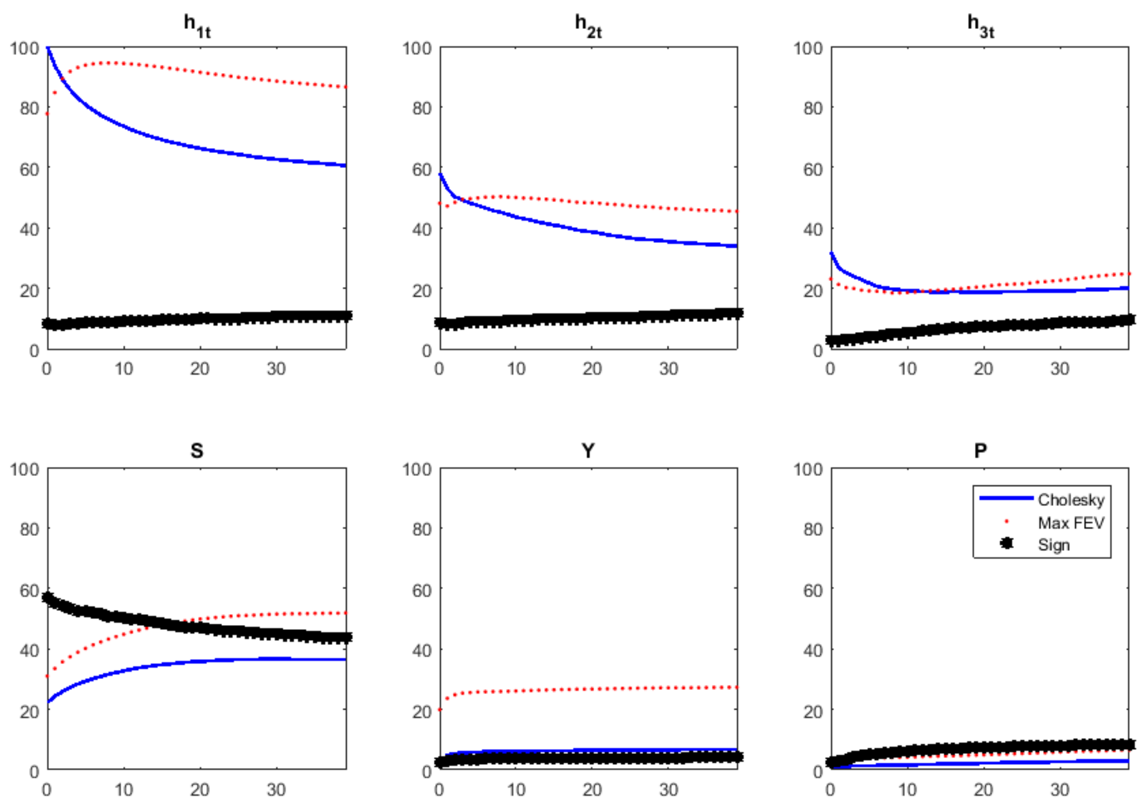

Figure 6: Contribution of the financial uncertainty shock to the FEV

4 Conclusions

This paper introduces a VAR with stochastic volatility in mean where the disturbances of the observation and transition equation are allowed to be correlated. This removes the need to assume exogeneity of volatility shocks apriori or to estimate the model in structural form. While the literature has shown that both of these approaches are feasible and useful in certain cases, the model developed in this paper provides an alternative that is closer to standard VAR models. In future work it may be interesting to extend the proposed model by incorporating time-varying parameters as it is likely that the reduced correlation amongst volatility and level shocks and their structural impact has changed over time.

References

Andrieu, Christophe, Arnaud Doucet and Roman Holenstein, 2010, Particle Markov chain Monte Carlo methods, *Journal of the Royal Statistical Society Series B* **72**(3), 269–342.

Banbura, Marta, Domenico Giannone and Lucrezia Reichlin, 2007, Bayesian VARs with Large Panels, *CEPR Discussion Papers 6326*, C.E.P.R. Discussion Papers.

Carriero, Andrea, Todd Clarke and Massimiliano Marcellino, 2016, Endogenous Uncertainty?, *Technical report*.

Carter, C and P Kohn, 2004, On Gibbs sampling for state space models, *Biometrika* **81**, 541–53.

Chan, Joshua Chi-Chun and Ivan Jeliazkov, 2009, MCMC Estimation of Restricted Covariance Matrices, *Journal of Computational and Graphical Statistics* **18**(2), 457–480.

Cogley, T. and T. J. Sargent, 2005, Drifts and Volatilities: monetary policies and outcomes in the Post WWII U.S., *Review of Economic Dynamics* **8**, 262–302.

Gordon, Robert J., 1986, *The American Business Cycle: Continuity and Change*, number gord86-1 in *NBER Books*, National Bureau of Economic Research, Inc.

JO, SOOJIN, 2014, The Effects of Oil Price Uncertainty on Global Real Economic Activity, *Journal of Money, Credit and Banking* **46**(6), 1113–1135.

Lindsten, Fredrik, Michael I. Jordan and Thomas B. Schön, 2014, Particle Gibbs with Ancestor Sampling, *Journal of Machine Learning Research* **15**, 2145–2184.

Ludvigson, Sydney C., Sai Ma and Serena Ng, 2015, Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?, *NBER Working Papers 21803*, National Bureau of Economic Research, Inc.

Mumtaz, Haroon and Francesco Zanetti, 2013, The Impact of the Volatility of Monetary Policy Shocks, *Journal of Money, Credit and Banking* **45**(4), 535–558.

Mumtaz, Haroon and Konstantinos Theodoridis, 2015a, Dynamic Effects of Monetary Policy Shocks on Macroeconomic Volatility, *Working Papers 760*, Queen Mary University of London, School of Economics and Finance.

Mumtaz, Haroon and Konstantinos Theodoridis, 2015b, The International Transmission Of Volatility Shocks: An Empirical Analysis, *Journal of the European Economic Association* **13**(3), 512–533.

Mumtaz, Haroon and Paolo Surico, n.d., Policy uncertainty and aggregate fluctuations, *Journal of Applied Econometrics* pp. n/a–n/a. jae.2613.

Nonejad, Nima, 2015, REPLICATING THE RESULTS IN ŚA NEW MODEL OF TREND INFLATIONŠ USING PARTICLE MARKOV CHAIN MONTE CARLO, *Journal of Applied Econometrics* pp. n/a–n/a.

Omori, Yasuhiro, Siddhartha Chib, Neil Shephard and Jouchi Nakajima, 2007, Stochastic volatility with leverage: Fast and efficient likelihood inference, *Journal of Econometrics* **140**(2), 425–449.

Pitt, Michael, Sheheryar Malik and Arnaud Doucet, 2014, Simulated likelihood inference for stochastic volatility models using continuous particle filtering, *Annals of the Institute of Statistical Mathematics* **66**(3), 527–552.

Uhlig, Harald, 2004, What moves GNP?, *Econometric Society 2004 North American Winter Meetings 636*, Econometric Society.

A Appendix: Model Estimation

Consider the VAR model

$$\tilde{h}_t = \alpha + \theta \tilde{h}_{t-1} + \sum_{j=1}^{Q} d_j Z_{t-j} + S^{1/2} \eta_t \tag{4}$$

$$Z_t = c + \sum_{j=1}^{P} \beta_j Z_{t-j} + \sum_{k=1}^{K} b_k \tilde{h}_{t-k} + H_t^{1/2} e_t \tag{5}$$

where Z_t is a matrix of endogenous variables, $\tilde{h}_t = [h_{1t}, h_{2t}, ..h_{N,t}]$, $H_t = diag\left(\exp\left(\tilde{h}_t\right)\right)$ and $\tilde{s} = [s_1, s_2, .., s_N]$ $S = diag(\tilde{s})$. The disturbances $\varepsilon_t = \begin{pmatrix} \eta_t \\ e_t \end{pmatrix}$ are distributed normally $N(0, \Sigma), \Sigma = \begin{pmatrix} \Sigma_\eta & \Sigma'_{\eta_e} \\ \Sigma_{\eta_e} & \Sigma_e \end{pmatrix}$ where the diagonal elements of Σ are restricted to equal 1. For example for $M = 4, N = 2 : \Sigma = \begin{pmatrix} 1 & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ \Sigma_{21} & 1 & \Sigma_{23} & \Sigma_{24} \\ \Sigma_{31} & \Sigma_{32} & 1 & \Sigma_{34} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & 1 \end{pmatrix}$

A.1 Prior distributions and starting values

A.1.1 VAR coefficients

Let $\Gamma = vec\left([c; \beta_j; b_k]\right)$. Following Banbura *et al.* (2007), we employ a Normal prior. The priors are implemented by the dummy observations y_D and x_D that are defined as:

$$y_D = \begin{bmatrix} \frac{diag(\gamma_1 s_1 ... \gamma_n s_n)}{\tau} \\ 0_{N \times (P-1) \times N} \\ .... \\ 0_{EX \times N} \end{bmatrix}, \quad x_D = \begin{bmatrix} \frac{J_P \otimes diag(s_1 ... s_n)}{\tau} & 0_{NP \times EX} \\ 0_{N \times (NP) + EX} & \\ .... & \\ 0_{EX \times NP} & I_{EX} \times 1/c \end{bmatrix} \tag{6}$$

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 exogenous and pre-determined regressors. EX denotes the number of exogenous and pre-determined regressors in each equation. N denotes the total number of endogenous variables and P is the lag length. We set $\tau = 0.1$. We use a different value of c for the coefficients on the lagged volatilities and for the coefficients on the lagged volatilities c is set equal to 0.1. A flat prior is used for the intercept terms and the corresponding tightness is set equal to $c = 1000$. Note that these dummies do not directly implement a prior belief on the VAR error covariance matrix which is time-varying in our setting.

The priors for the coefficients are thus: $N(\Gamma_0, P_0)$ where $\Gamma_0 = (x'_D x_D)^{-1} (x'_D y_D)$ and $P_0 = S \otimes (x'_D x_D)^{-1}$ where S is a diagonal matrix with an estimate of the variance of Z_t (obtained using the training sample described below) on the main diagonal.

A.1.2 Elements of H_t

Following Cogley and Sargent (2005) we use a training sample (of 20 pre-sample observations) to set the prior for the elements of the transition equation of the model. Let $\hat{v}^{ols}$ denote the OLS estimate of the VAR covariance matrix estimated on the pre-sample data. The prior for $\tilde{h}_t$ at $t = 0$ is defined as $\ln h_0 \sim N(\ln \mu_0, I_4)$ where μ_0 are the diagonal elements of the Cholesky decomposition of $\hat{v}^{ols}$.

A.1.3 Elements of Σ

Chan and Jeliazkov (2009) decompose Σ as $\Sigma = L^{-1} D L^{-1\prime}$ with the diagonal elements of D denoted by λ_k and a_{kj} denoting the lower triangular elements of L^{-1}. The prior for a_{kj} is assumed to be $N(0, 1)$ while the prior on D is implicit via the restriction that Σ have diagonal elements that equal 1.

A.1.4 Parameters of the transition equation

The prior for VAR coefficients $\tilde{\Gamma} = vec([\alpha; \theta; d_j])$ is set as above for Γ. We assume an inverse Gamma prior for $\tilde{s} : IG(v_0, T_0), v_0 = 0.001, T_0 = 1$.

A.2 Simulating the posterior distributions

A.2.1 Coefficients

Conditional on S, H_t and Σ, the model can be written as a SUR system with heteroscedasticity

$$\begin{aligned} Y_t &= X_t \Pi_t + E_t \\ var\left(E_t\right) &= G_t \Sigma G_t' \end{aligned} \tag{7}$$

where $G_t = diag\left([\tilde{s}^{1/2}, \exp\left(\tilde{h}_t\right)^{1/2}]\right)$, $Y_t = \begin{pmatrix} h_{1t} \\ h_{2t} \\ . \\ Z_t \end{pmatrix}$, $X_t = \begin{pmatrix} x_{1t} & 0 & 0 & 0 \\ 0 & x_{2t} & 0 & 0 \\ 0 & 0 & . & 0 \\ 0 & 0 & 0 & . \end{pmatrix}$ where x_{it} denotes the coefficients in the ith equation of the system. Given a draw for the time-varying error covariance matrix, the coefficients have a conditional posterior that is normal: $N\left(\Pi_{T\backslash T}, P_{T\backslash T}\right)$. Following Carter and Kohn (2004) we use the Kalman filter to estimate the mean and variance of the conditional posterior where we account for the fact that the covariance matrix of the VAR residuals changes through time. To use the Kalman filter we define the transition equation as

$$\Pi_t = \Pi_{t-1}$$

The Kalman filter is initialised at Π_0 and $P_{0|0}$ which are based on the priors for the coefficients introduced above and the recursions are given by the following equations for $t = 1, 2..T$

$$
\begin{aligned}
\Pi_{t\backslash t-1} &= \Pi_{t-1\backslash t-1} \\
P_{t\backslash t-1} &= P_{t-1\backslash t-1} \\
\eta_{t\backslash t-1} &= Y_t - X_t \Pi_{t\backslash t-1} \\
f_{t\backslash t-1} &= X_t P_{t\backslash t-1} X_t' + (G_t \Sigma G_t') \\
K_t &= P_{t\backslash t-1} X_t' f_{t\backslash t-1}^{-1} \\
\Pi_{t\backslash t} &= \Pi_{t\backslash t-1} + K_t \eta_{t\backslash t-1} \\
P_{t\backslash t} &= P_{t\backslash t-1} - K_t X_t P_{t\backslash t-1}
\end{aligned}
$$

The final iteration of the Kalman filter at time T delivers $\Pi_{T\backslash T}$ and $P_{T\backslash T}$.

A.3 Element of S

Given the residuals e_t, and Σ the transition equations can be written as VAR:

$$
\begin{aligned}
\tilde{h}_t - S^{1/2}\mu_{\eta_t|e_t} &= \alpha + \theta \tilde{h}_{t-1} + \sum_{j=1}^{Q} d_j Z_{t-j} + \eta_t^* \\
var\left(\eta_t^*\right) &= S^{1/2} \Sigma_{\eta_t|e_t} S^{1/2\prime}
\end{aligned} \tag{8}
$$

$$
\begin{aligned}
\mu_{\eta_t|e_t} &= e_t \Sigma_e^{-1} \Sigma_{\eta_e} \\
\Sigma_{\eta_t|e_t} &= \Sigma_\eta - \Sigma_{\eta_e}' \Sigma_e^{-1} \Sigma_{\eta_e}'
\end{aligned}
$$

Note that η_t^* is uncorrelated with the residuals of the observation equation. The proposal density $q\left(.\right)$ is defined as:

$$
S_j = \varkappa IG\left(v_1, T_1\right) + (1-\varkappa) IG\left(v\left(S_{j-1}\right), T\left(\bar{V}\right)\right)
$$

where the parameter v_1 is set to $\tilde{\eta}_{it}'\tilde{\eta}_{it} + v_0$ and $T_1 = T_0 + T$ where T is the sample size and $\tilde{\eta}_{it}$ denotes the residuals of the ith transition equation. Letting the mean of the IG distribution equal S_{j-1} and standard deviation $\bar{V}$, the implied parameters of the second component of $q(.)$ are defined as:

$$
\begin{aligned}
v(S_{j-1}) &= 2S_{j-1}\left(1+\frac{S_{j-1}^2}{\bar{V}^2}\right) \\
T(\bar{V}) &= 2\left(2+\frac{S_{j-1}^2}{\bar{V}^2}\right)
\end{aligned}
$$

In the benchmark model, we set $\varkappa = 0.5$ and $\bar{V}$ is chosen to obtain a satisfactory acceptance rate. The draws accepted with probability:

$$
\alpha = \frac{g(E_t|S_j)\, q(S_{j-1})}{g(E_t|S_{j-1})\, q(S_j)}
$$

where $g(E_t|S_j)$ denotes the posterior distribution evaluated at the jth draw of S given all other parameters at their values drawn in previous steps. With the model in the form of a VAR (equation 8), the likelihood can be evaluated easily.

A.3.1 Elements of Σ

Chan and Jeliazkov (2009) describe how to sample covariance matrices with restrictions on some of the elements and we follow their method in implementing the draw from this conditional posterior. Chan and Jeliazkov (2009) decompose Σ as $\Sigma = L^{-1}DL^{-1\prime}$. They show that when the diagonal elements of Σ are restricted to equal 1, then the diagonal elements of D (denoted by λ_k) satisfy

$$
\begin{aligned}
\lambda_1 &= 1 \\
\lambda_k &= 1-\sum_{j=1}^{k-1}\left(a^{kj}\right)^2\lambda_j, k=2,3,..N
\end{aligned}
\tag{9}
$$

where a^{kj} are lower diagonal elements of L^{-1}. They propose an independence Metropolis step to sample a^{kj} with a proposal density of the form:

$$
f\left(a^{kj}|\varepsilon_t\right) = N(\mu_k, \tau V_k)
$$

where $V = \left(A_0^{-1} + \sum_{t=1}^{T} U_t \hat{D}^{-1} U_t\right)$ and $\mu = V\left(A_0^{-1} a_0 + \sum_{t=1}^{T} U_t \hat{D}^{-1} \varepsilon_t\right)$. Here U_t is defined as the matrix:

$$U_t = -\begin{pmatrix} 0 & . & . & . & & & & & 0 \\ \varepsilon_{t,1} & 0 & . & . & . & & & & \\ 0 & \varepsilon_{t,1} & \varepsilon_{t,2} & . & . & . & & & \\ 0 & . & . & \varepsilon_{t,1} & \varepsilon_{t,2} & \varepsilon_{t,3} & 0 & & . \\ . & . & . & & & & & & . \\ . & . & . & & & & & & . \\ . & . & . & & & & & & 0 \\ 0 & . & . & . & 0 & 0 & \varepsilon_{t,1} & . & \varepsilon_{t,N} \end{pmatrix}$$

and the diagonal elements of $\hat{D}$ can be obtained by iterating between the equation for μ and equation 9. The draw is accepted with probability:

$$\alpha = \frac{g\left(\varepsilon_t|\Sigma_{new}\right) f\left(a_{old}|\varepsilon_t\right)}{g\left(\varepsilon_t|\Sigma_{old}\right) f\left(a_{new}|\varepsilon_t\right)}$$

with λ_k restricted to be greater than zero to ensure that Σ is positive definite. The expression for the likelihood function used to construct the posterior $g\left(\varepsilon_t|\Sigma\right)$ is given in equation 2.7 in Chan and Jeliazkov (2009).

A.3.2 Elements of H_t

Conditional on the VAR coefficients and the parameters of the transition equation, the model has a multivariate non-linear state-space representation. It is convenient to express the state-space as:

$$\begin{aligned} F_t &= C + \Psi F_{t-1} + N_t \\ Z_t - H_t^{1/2} \mu_{e_t|\eta_t} &= c + \sum_{j=1}^{P} \beta_j Z_{t-j} + \sum_{k=1}^{K} b_k \tilde{h}_{t-k} + \tilde{e}_t \\ var\left(\tilde{e}_t\right) &= \Omega_t = H_t^{1/2} \Sigma_{e_t|\eta_t} H_t^{1/2\prime} \end{aligned}$$

where

$$
\begin{aligned}
F_t &= \begin{pmatrix} \eta_{t+1} \\ \eta_t \\ \tilde{h}_t \\ . \\ \tilde{h}_{t-k} \end{pmatrix} \\
C &= \begin{pmatrix} 0 \\ 0 \\ \alpha + \sum_{j=1}^{Q} d_j Z_{t-j} \\ . \\ 0 \end{pmatrix} \\
\Psi &= \begin{pmatrix} 0 & 0 & . & . & 0 \\ 1 & 0 & . & . & . \\ S^{1/2} & 0 & \theta & . & 0 \\ 0 & 1 & . & . & 0 \\ . & . & 1 & . & 0 \end{pmatrix} \\
N_t &= \begin{pmatrix} \eta_{t+1} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}
\end{aligned}
$$

$\mu_{e_t|\eta_t}$ denotes the conditional mean of e_t while $\Sigma_{e_t|\eta_t}$ is the conditional variance. These can be easily calculated using results for multi-variate normal distributions. Partitioning Σ as:

$$
\Sigma = \begin{pmatrix} \Sigma_\eta & \Sigma'_{\eta_e} \\ \Sigma_{\eta_e} & \Sigma_e \end{pmatrix}
$$

the conditional mean and variance are given by:

$$\begin{aligned}
\mu_{e_t|\eta_t} &= \eta_t \Sigma_\eta^{-1} \Sigma'_{\eta_e} \\
\Sigma_{e_t|\eta_t} &= \Sigma_e - \Sigma_{\eta_e} \Sigma_\eta^{-1} \Sigma'_{\eta_e}
\end{aligned}$$

Moreover:

$$var\left(N_t\right) = \tilde{Q} = \begin{pmatrix} \Sigma_\eta & . & . & . & 0 \\ 0 & . & . & & . \\ 0 & . & . & & . \\ 0 & . & & & . \\ 0 & . & 0 & & 0 \end{pmatrix}$$

Following recent developments in the seminal paper by Andrieu *et al.* (2010), we employ a particle Gibbs step to sample from the conditional posterior of F_t. Andrieu *et al.* (2010) show how a version of the particle filter, conditioned on a fixed trajectory for one of the particles can be used to produce draws that result in a Markov Kernel with a target distribution that is invariant. However, the usual problem of path degeneracy in the particle filter can result in poor mixing in the original version of particle Gibbs. Recent development, however, suggest that small modifications of this algorithm can largely alleviate this problem. In particular, Lindsten *et al.* (2014) propose the addition of a step that involves sampling the 'ancestors' or indices associated with the particle that is being conditioned on. They show that this results in a substantial improvement in the mixing of the algorithm even with a few particles.[3]As explained in Lindsten *et al.* (2014), ancestor sampling breaks the reference path into pieces and this causes the particle system to collapse towards something different than the reference path. In the absence of this step, the particle system tends to collapse to the conditioning path. We employ particle Gibbs with ancestor sampling in this step.

Let $F_t^{(i-1)}$ denote the fixed the fixed trajectory, for $t = 1, 2, ..T$ obtained in the previous draw of the Gibbs algorithm. We denote all the parameters of the model by Ξ, and $j = 1, 2, ..M$ indexes the particles. The conditional particle filter with ancestor sampling proceeds in the following steps:

[3]See Nonejad (2015) for a recent application of this algorithm.

1. For $t = 1$

   (a) Draw $F_1^{(j)} \backslash F_0^{(j)}$,$\Xi$ for $j = 1, 2, ..M - 1$. Fix $F_1^{(M)} = F_1^{(i-1)}$

   (b) Compute the normalised weights $p_1^{(j)} = \frac{w_1^{(j)}}{\sum_{j=1}^{M} w_1^{(j)}}$ where $w_1^{(j)}$ denotes the conditional likelihood: $\left|\Omega_1^{(j)}\right|^{-0.5} - 0.5 \exp\left(\tilde{e}_1 \left(\Omega_1^{(j)}\right)^{-1} \tilde{e}_1'\right)$ where $\Omega_1^{(j)} = H_1^{(j)} \Sigma_{e_t|\eta_t} H_1^{(j)\prime}$ with $H_1^{(j)} = diag\left(\exp\left(\tilde{h}_{1,[0]}^{(j)}\right)\right)$ and $\tilde{e}_1 = Z_1 - (H_1^{(j)1/2} \mu_{e_t|\eta_t} + c + \sum_{j=1}^{P} \beta_j Z_{t-j} + \sum_{k=1}^{K} b_k \tilde{h}_{1,[-k]}^{(j)})$ The subscript [0] denotes the contemporaneous value in the state vector while $[-k]$ denote the k lagged states.

2. For $t = 2$ to T

   (a) Resample $F_{t-1}^{(j)}$ for $j = 1, 2, ..M - 1$ using indices $a_t^{(j)}$ with $\Pr\left(a_t^{(j)} = j\right) \propto p_{t-1}^{(j)}$

   (b) Draw $F_t^{(j)} \backslash F_{t-1}^{(a_t^{(j)})}$,$\Xi$ for $j = 1, 2, ..M - 1$ using the transition equation of the model. Note that $F_{t-1}^{(a_t^{(j)})}$ denotes the resampled particles in step (a) above.

   (c) Fix $F_t^{(M)} = F_t^{(i-1)}$

   (d) Sample $a_t^{(M)}$ with $\Pr\left(a_t^{(M)} = j\right) \propto p_{t-1}^{(j)} \Pr\left(F_t^{(i-1)} \backslash F_{t-1}^{(j)}, C, \Psi, \tilde{Q}\right)$ where the density $\Pr\left(F_t^{(i-1)} \backslash F_{t-1}^{(j)}, C, \Psi, \tilde{Q}\right)$ is computed as $\left|\tilde{Q}\right|^{-0.5} - 0.5 \exp\left(N_t^{(j)} \left(\tilde{Q}\right)^{-1} N_t^{(j)}\right)$. This constitutes the ancestor sampling step. If $a_t^{(M)} = M$ then the algorithm collapses to the simple particle Gibbs.

   (e) Update the weights $p_t^{(j)} = \frac{w_t^{(j)}}{\sum_{j=1}^{M} w_t^{(j)}}$ where $w_1^{(j)}$ denotes the conditional likelihood: $\left|\Omega_t^{(j)}\right|^{-0.5} - 0.5 \exp\left(\tilde{e}_t \left(\Omega_t^{(j)}\right)^{-1} \tilde{e}_t'\right)$

3. End

4. Sample $F_t^{(i)}$ with $\Pr\left(F_t^{(i)} = F_t^{(j)}\right) \propto p_T^{(j)}$ to obtain a draw from the conditional posterior distribution

We use $M = 20$ particles in our application. The initial values μ_0 defined above are used to initialise step 1 of the filter.

School of Economics and Finance

Queen Mary
University of London

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

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