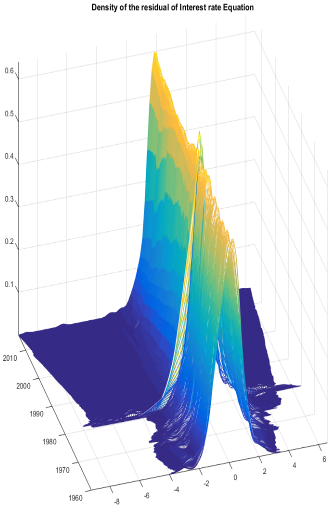
VAR models with non-Gaussian shocks

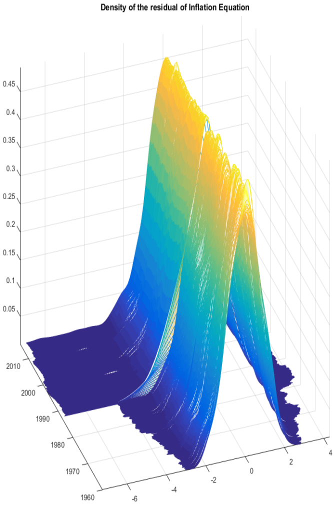
Ching-Wai (Jeremy) Chiu   Haroon Mumtaz   Gabor Pinter

September 27, 2016

Motivation and aims

- Density of the residuals from a recursive VAR(13) (1960m1-2015m6)





Motivation and aims

- Describe a BVAR model that features non-normal disturbances
- The non-normality is introduced in the model through a finite mixture of normals
- Consider if this specification can improve point and density forecasting performance
- Application to yield curve forecasting.

Related papers

- Closely related to Chiu *et al.* (2014).
- Kalliovirta *et al.* (2016) explore a similar model in a frequentist setting. Related to multivariate STAR model of Dueker *et al.* (2011).
- Non-normality explored in Curdia *et al.* (2013) for DSGE models.
- Villani *et al.* (2009) describe regression models with Gaussian mixtures.
- Econometric background in Koop (2003), Geweke (2005).
- Canova (1993), Sims (1992) explore the causes/implications of non-normality.

BVAR model with Non-normal disturbances

The BVAR model is defined as follows

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

The covariance matrix of the residuals is defined as

$$cov(u_t) = \Sigma = A^{-1} H A^{-1'} \tag{2}$$

The orthogonalised residuals are given by

$$e_t = A u_t$$

BVAR model with Non-normal disturbances

The shock to the *ith* equation is assumed to follow:

$$e_{it} = \alpha_{i,S_{it}} + \sigma_{i,S_{it}}\varepsilon_{it}, \varepsilon_{it}\tilde{} N(0,1) \tag{3}$$

where $S_{it} = 1, 2, ...M$ denotes the unobserved components or regimes.

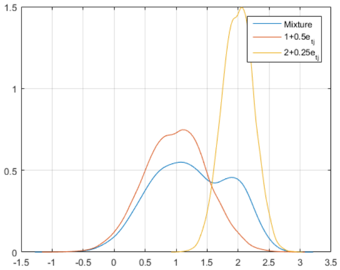
- The formulation in equation 3, describes a mixture of M distributions where each component is $N\left(\alpha_i, \sigma_i^2\right)$. The state variable S determines the component that is active at a particular point in time.
- The law of motion for S_{it} is chosen to be first order Markov process with transition probabilities

$$P_i\left(S_{i,t} = J | S_{i,t-1} = I\right) = p_{i,IJ} \tag{4}$$

- $p_{i,IJ}$ can depend on a vector of unobserved variables $P_i\left(S_{i,t} = J | S_{i,t-1} = I\right) = p_{i,IJ}\left(z_t\right)$

BVAR model with Non-normal disturbances

- The orthogonal shocks are described by a finite mixture of normals.
- If $\alpha_{i,S_{it}}$ are the same across components, the distribution is symmmetric but can have fat tails
- If $\alpha_{i,S_{it}}$ vary, the distribution can be skewed and have a kurtosis less than 3.



BVAR model with Non-normal disturbances

- The VAR model proposed above can also be interpreted as a Markov Switching VAR model (see Hamilton (1994)).

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (5)$$

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ A_{21} & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (6)$$

where $e_{1t} = \alpha_{1,S_{1t}} + \sigma_{1,S_{1t}} \varepsilon_{1t}$ and $e_{2t} = \alpha_{2,S_{2t}} + \sigma_{2,S_{2t}} \varepsilon_{2t}$

BVAR model with Non-normal disturbances

Combining these equations we get:

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} =$$

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ A_{21} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \alpha_{1,S_{1t}} \\ \alpha_{2,S_{2t}} \end{pmatrix} +$$

$$\begin{pmatrix} 1 & 0 \\ A_{21} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{1,S_{1t}} & 0 \\ 0 & \sigma_{2,S_{2t}} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Intercepts: $\begin{pmatrix} \alpha_{1,S_{1t}} \\ \alpha_{2,S_{2t}} - A_{21}\alpha_{1,S_{1t}} \end{pmatrix}$

Covariance: $\begin{pmatrix} \sigma^2_{1,S_{1t}} & -A_{21}\sigma^2_{1,S_{1t}} \\ -A_{21}\sigma^2_{1,S_{1t}} & \sigma^2_{1,S_t}A^2_{21} + \sigma^2_{2,S_{2t}} \end{pmatrix}$

- n different Markov chains govern the regime switches of the reduced form

Estimation

We adopt a Bayesian approach to estimation. This has two advantages

- The marginal likelihood can be used to select the number of components
- A small extension of the MCMC algorithm provides the predictive density.

Estimation

Given B,A the model can be written as

$$e_{it} = \alpha_{i,S_{it}} + \sigma_{i,S_{it}} \varepsilon_{it}$$

- Given S_{it}, this is sequence of linear regressions and the draws from the conditional posterior of α and σ^2 are standard. Impose $\alpha_{i,S_{it}=1} < \alpha_{i,S_{it}=2} < ... < \alpha_{i,S_{it}=M}$
- Given α and σ^2, S_{it} can be drawn using the Hamilton filter and the backward recursion described in Kim and Nelson (1999) and transition probabilities from the Dirichlet distribution.
- Given α and σ, the remaining parameters involve VARs, regressions with heteroscedasticity and standard methods apply after a GLS transformation

Model Selection

- We carry out model selection by comparing the marginal likelihood across models with a different number of components. The marginal likelihood is defined as:

$$f(\tilde{y}) = \int f(\tilde{y}|\Xi)\, p(\Xi)\, d\Xi \tag{7}$$

  where $\tilde{y} = [y_1, y_2, .., y_T]$, Ξ denotes the unknown parameters of the model, $f(\tilde{y}|\Xi)$ is the likelihood and $p(\Xi)$ is the proper prior distribution.
- As is well known, the integration problem in equation 7 is non-trivial and several numerical methods have been proposed for this
- We use two methods Gelfand and Dey (1994) estimator and the bridge estimator Meng and Wong (1996)
- We also consider AIC, BIC and DIC.

Estimation on Artificial data

- The paper presents a Monte-Carlo experiment which shows some evidence that:
  1. The MCMC algorithm displays a reasonable performance
  2. ML estimators select the correct model almost 100% of the time
  3. AIC and SIC perform well, but DIC tends to select the more complex model.

Empirical Application: Modelling and Forecasting the yield curve

- The Yield curve and the economy are closely related and thus it is interesting to forecast yields (Diebold *et al.* (2006))
- Note, however, that some recent papers have pointed out that the dynamics of the yield curve are subject to structural shifts. (For example, Mumtaz and Surico (2009) and Bianchi *et al.* (2009))
- In addition, it is well known that yields at longer maturities have been trending downwards in recent years (i.e. 'Greenspan's conundrum').

Empirical Application: Modelling and Forecasting the yield curve

- Following Diebold and Li (2006) we use the Nelson and Siegel (1987) specification.
- Letting $y_t(\tau)$ denote zero coupon government bond yields at maturity τ, the Nelson and Siegel (1987) model is defined as:

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau}\right) + \beta_{3t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau}\right) \tag{8}$$

  where λ_t controls the exponential decay rate.
- The level, slope and the curvature of the yield curve are captured by β_{1t}, β_{2t} and β_{3t}, respectively which can be estimated via OLS.

Empirical Application: Modelling and Forecasting the yield curve

- The dynamics of the factors can be modelled as a VAR process which is used in Diebold and Li (2006) to produce out of sample forecasts. In this section, we compare this VAR specification with the following extended model:

$$Z_t = \sum_{l=1}^{L} B_l Z_{t-l} + u_t$$

where $Z_t = \{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ and $e_t = Au_t$. The orthogonal shock to the *ith* equation of the VAR is a Markov mixture of normals:

$$e_{it} = \alpha_{i,S_{it}} + \sigma_{i,S_{it}} \varepsilon_{it}, \varepsilon_{it} \sim N(0,1) \tag{9}$$

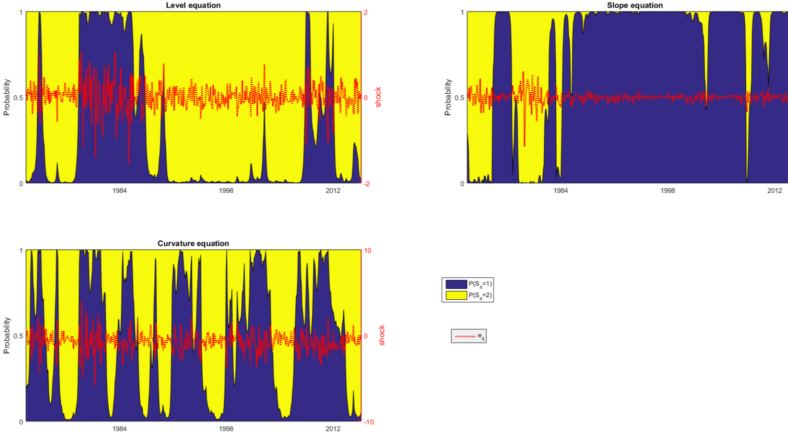
Empirical Application: Modelling and Forecasting the yield curve

- ML suggests a model with 2 components

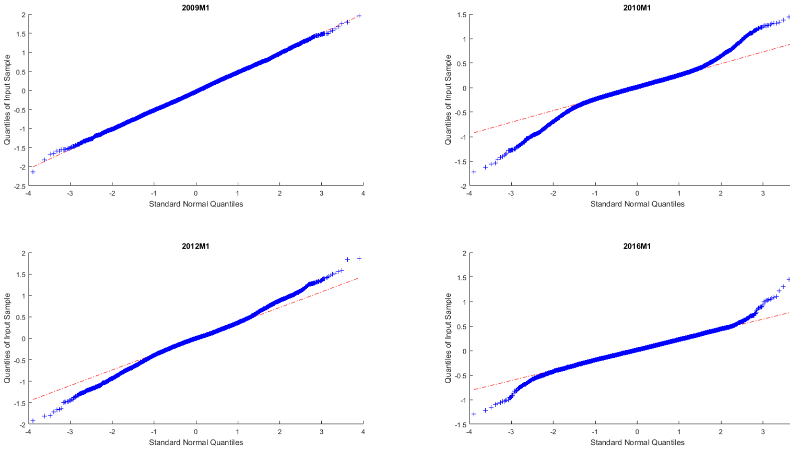
Equation	$\alpha_{i,S_{it}=1}$	$\alpha_{i,S_{it}=2}$	$\sigma^2_{i,s_{it}=1}$	$\sigma^2_{i,s_{it}=2}$	P_i
Level	-0.024	0.017	0.241	0.041	$\begin{pmatrix} 0.946 & 0.022 \\ 0.054 & 0.978 \end{pmatrix}$
	$[-0.055, -0.006]$	$[-0.008, 0.045]$	$[0.210, 0.283]$	$[0.037, 0.046]$	
Slope	-0.029	0.064	0.105	1.340	$\begin{pmatrix} 0.983 & 0.058 \\ 0.017 & 0.942 \end{pmatrix}$
	$[-0.071, -0.007]$	$[-0.006, 0.171]$	$[0.093, 0.117]$	$[1.148, 1.573]$	
Curvature	-0.935	-0.545	1.978	0.368	$\begin{pmatrix} 0.934 & 0.069 \\ 0.066 & 0.931 \end{pmatrix}$
	$[-1.235, -0.628]$	$[-0.759, -0.326]$	$[1.733, 2.340]$	$[0.306, 0.442]$	

Table 2: Estimates of regime dependent parameters. Median and 68 percent highest posterior density interval

Empirical Application: Modelling and Forecasting the yield curve



Empirical Application: Modelling and Forecasting the yield curve



Forecasting performance

- Estimate up to December 1979. Then each model is estimated recursively adding one month of data at a time until January 2015. At each recursion, we produce a 12 month density forecast for the three factors from
  1. Proposed model with 2 and 3 components (M2-VAR and M3-VAR)
  2. A BVAR with T errors
  3. A BVAR with SVOL
- We assess the point forecasts using root mean squared errors (RMSE) and the density forecasts using the continuous rank probability score (CRPS) as it is less sensitive to outlier outcomes (see Gneiting and Raftery (2007)). We consider the performance over the full sample and over the

Forecasting performance

- Over the full sample, the performance of these models is very similar and deliver a gain of about 10% over the BVAR.
  - The proposed model with two components delivers point and density forecasts for the level of the yield curve that are more accurate than those obtained from the competing models.
- Over the post-1990 forecast sample, the improvement in forecasting performance over the BVAR is substantially larger. The M2-VAR delivers RMSEs in forecasting the level factor which are more than 20% lower than those obtained from the BVAR model.
  - The performance of the SVOI-VAR is similar at short horizons. The proposed model does better at the one year horizon.

Summary

- We introduce Non-Gaussian disturbances in a BVAR model and provide the Gibbs algorithm.
- Provide evidence of non-normality in shocks using a US BVAR model.
- The proposed model may be useful in forecasting variables over more tranquil periods
  - Forecasting Macroeconomic variables
  - Looking at a wider range of countries
  - Asymmetric number of components

Appendix

	RMSE				CRPS			
	1M	3M	6M	12M	1M	3M	6M	12M
Level Factor								
VAR-T	0.787	0.774	0.780	0.762	0.775	0.763	0.788	0.769
M2-VAR	0.779	0.772	0.773	0.730	0.765	0.748	0.759	0.682
M3-VAR	0.792	0.797	0.807	0.761	0.780	0.781	0.801	0.708
SVOL-VAR	0.786	0.776	0.783	0.758	0.767	0.751	0.774	0.737
Slope Factor								
VAR-T	0.778	0.765	0.807	0.900	0.750	0.766	0.868	0.999
M2-VAR	0.777	0.761	0.790	0.858	0.729	0.730	0.816	0.918
M3-VAR	0.784	0.762	0.789	0.866	0.730	0.729	0.835	0.982
SVOL-VAR	0.760	0.735	0.765	0.856	0.717	0.705	0.805	0.967
Curvature Factor								
VAR-T	0.934	0.968	0.996	1.036	0.928	0.973	1.016	1.075
M2-VAR	0.944	0.981	1.016	1.049	0.937	1.003	1.070	1.137
M3-VAR	0.943	0.978	1.016	1.053	0.941	1.006	1.089	1.153
SVOL-VAR	0.954	1.005	1.053	1.112	0.942	1.022	1.100	1.178

Bianchi, Francesco, Haroon Mumtaz and Paolo Surico, 2009, The great moderation of the term structure of UK interest rates, *Journal of Monetary Economics* **56**(6), 856 – 871.

Canova, Fabio, 1993, Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model, *Journal of Economic Dynamics and Control* **17**(1-2), 233–261.

Chiu, Ching-Wai (Jeremy), Haroon Mumtaz and Gabor Pinter, 2014, Fat-tails in VAR Models, *Working Papers 714*, Queen Mary University of London, School of Economics and Finance.

Curdia, Vasco, Marco Del Negro and Daniel L. Greenwald, 2013, Rare shocks, Great Recessions, *Technical report*.

Diebold, Francis X. and Canlin Li, 2006, Forecasting the term structure of government bond yields, *Journal of Econometrics* **130**(2), 337–364.

Diebold, Francis X., Glenn D. Rudebusch and S. BoragËĞan Aruoba, 2006, The macroeconomy and the yield curve: a dynamic latent factor approach, *Journal of Econometrics* **131**(1âĂŞ2), 309 – 338.

Dueker, Michael, Zacharias Psaradakis, Martin Sola and Fabio Spagnolo, 2011, Multivariate contemporaneous-threshold autoregressive models, *Journal of Econometrics* **160**(2), 311–325.

Gelfand, A. E. and D. K. Dey, 1994, Bayesian Model Choice: Asymptotics and Exact Calculations, *Journal of the Royal Statistical Society. Series B (Methodological)* **56**(3), pp. 501–514.

Geweke, John, 2005, *Contemporary Bayesian Econometrics and Statistics*, Wiley.

Gneiting, Tilmann and Adrian E. Raftery, 2007, Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association* **102**, 359–378.

Hamilton, James D., 1994, *Time Series Analysis*, 1 edition, Princeton University Press.

Kalliovirta, Leena, Mika Meitz and Pentti Saikkonen, 2016, Gaussian mixture vector autoregression, *Journal of Econometrics* **192**(2), 485 – 498. Innovations in Multiple Time Series Analysis.

Kim, Chang Jin and Charles R Nelson, 1999, *State-Space Models with Regime Switching*, MIT Press.

Koop, Gary, 2003, *Bayesian Econometrics*, Wiley.

Meng, Xiao Li and Wing Hung Wong, 1996, Simulating ratios of normalizing constants via a simple identity: A theoretical exploration, *Statistica Sinica* pp. 831–860.

Mumtaz, Haroon and Paolo Surico, 2009, Time-varying yield curve dynamics and monetary policy, *Journal of Applied Econometrics* **24**(6), 895–913.

Nelson, C R. and A F Siegel, 1987, Parsimonious modeling of yield curves, *The Journal of Business* **60**(4), 473–489.

Sims, Christopher, 1992, A Nine Variable Probabilistic Macroeconomic Forecasting Model, *Cowles Foundation Discussion Papers 1034*, Cowles Foundation for Research in Economics, Yale University.

Villani, M., R. Kohn and P. Giordani, 2009, Regression density estimation using smooth adaptive Gaussian mixtures, *Journal of Econometrics* **153**(2), 155–173. cited By 29.