UNIVERSITY OF LEEDS

This is a repository copy of *Discrete-response state space models with conditional heteroscedasticity: An application to forecasting the federal funds rate target*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/141639/

Version: Accepted Version

Article:

Dimitrakopoulos, S and Dey, DK (2017) Discrete-response state space models with conditional heteroscedasticity: An application to forecasting the federal funds rate target. Economics Letters, 154. pp. 20-23. ISSN 0165-1765

https://doi.org/10.1016/j.econlet.2017.02.012

© 2017 Elsevier B.V. All rights reserved. Licensed under the Creative Commons Attribution-Non Commercial No Derivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Discrete-response state space models with conditional heteroscedasticity: An application to forecasting the federal funds rate target

Stefanos Dimitrakopoulos*[1] and Dipak K. Dey[2]

[1]Department of Accounting, Finance and Economics, Oxford Brookes University, UK

[2]Department of Statistics, University of Connecticut, USA

Abstact

We propose a state space mixed model with stochastic volatility for ordinal-response time series data. For parameter estimation, we design an efficient Markov chain Monte Carlo algorithm. We illustrate our method with an empirical study on the federal funds rate target. The proposed model provides better forecasts than alternative specifications.

Keywords: conditional heteroscedasticity, Markov chain Monte Carlo, discrete responses, state-space model

JEL CODE: C11, C15, C22

*Correspondence to: Stefanos Dimitrakopoulos, Oxford Brookes University, Department of Accounting, Finance and Economics, Oxford, OX33 1HX, UK, Tel: +44(0) 1865 485478, E-mail: sdimitrakopoulos@brookes.ac.uk.

1 Introduction

Generalized linear state space (GLSS) models for discrete-response time series observations have been well studied in Bayesian literature (West et al., 1985; Fahrmeir, 1992; Song, 2000; Czado and Song, 2008; Stefanescu et al., 2009; Abanto-Valle and Dey, 2014). This class of models consists of two processes. In the first process, an observation or measurement equation defines the conditional mean of a time series of discrete observations as a nonlinear function (known as the inverse link function) of a sequence of latent state variables. In the second process, a transition or state equation describes the (stationary or non-stationary) dynamic process of the randomly time-varying state variables.

GLSS models can capture, through a time-varying parameter specification, the structural instability which may be present in time series of macro(financial) variables. A second well-known characteristic of (macro)financial time series is conditional heteroscedasticity. Researchers have highlighted the importance of allowing for time-varying conditional variances when analyzing discrete-response time series data (Hausman et al., 1992; Bollerslev et al., 1992; Dueker, 1999). However, the Bayesian literature on GLSS models has assumed homoscedastic errors so far.

In this paper, we extend the Bayesian literature on GLSS models by introducing a new class of models, the generalized nonlinear state space (GNLSS) models. The term "nonlinear" is justified by the presence of conditional heteroscedasticity. In the context of our empirical application we show that by accounting for conditional heteroscedasticity we achieve an increase in the forecast performance of GLSS models.

In particular, we develop methods of Bayesian inference in a state space mixed model with stochastic volatility (SV) (Taylor, 1986) for ordinal-valued time series. The stochastic volatility component accounts for some stylized facts of (macro)financial time series such as volatility clustering, heavy tails and high-peakedness. For the proposed ordinal-response model, the inverse link function is assumed to be a normal cumulative distribution function (c.d.f). The term "mixed" refers to the inclusion of both constant and time-varying coefficients in the model. The parameter transitions are captured by a random walk process.

The proposed model contributes also to the literature on discrete-response time series models with conditional heteroscedasticity (Müller and Czado (2009), Hsieh and Yang (2009),Yang and Parwada (2012), Ahmed (2015)). In the context of our empirical application, we show that by not accounting for time-varying parameters, the forecasting ability of discrete-choice models with conditional heteroscedasticity deteriorates.

Our model entails estimation challenges due to its latent nature, the presence of stochastic volatilities as well as the presence of the latent time-varying parameters. Therefore, we resort to Markov chain Monte Carlo methods and devise an efficient algorithm in order to estimate all parameters of interest.

In terms of our empirical application, our point of departure is the famous model of Hamilton and Jorda (2002) who examined the direction and magnitude of change of the Federal funds rate target in the context of an ordered probit specification. We built upon this model to account for time-varying parameters as well as conditional heteroscedasticity and conduct a forecasting exercise. Forecast evaluation is conducted, using point and density forecasts.

The resulting empirical model is inspired by the paper of Dueker (1999) who highlighted the importance of accounting for conditional heteroscedasticity in modelling discrete changes in the bank prime lending rate and the paper of Huang and Lin (2006) who examined the same issue, using an ordered probit model with time-varying parameters.

2 Econometric set up

Consider the following latent time-varying parameter regression model with stochastic volatility

$$y_t^* = \mathbf{x}_t' \boldsymbol{\beta} + \mathbf{z}_t' \boldsymbol{\alpha}_t + \varepsilon_t,\ \varepsilon_t \sim N(0, \exp(h_t)),\ t = 1, ..., T, \tag{1}$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \mathbf{u}_t,\ \mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}),\ t = 0, 1, ..., T-1, \tag{2}$$

$$h_t = \mu_h + \phi(h_{t-1} - \mu_h) + \eta_t,\ |\phi| < 1,\ \eta_t \sim N(0, \sigma_\eta^2). \tag{3}$$

Equation (1) contains the constant coefficient vector, $\boldsymbol{\beta}$, of dimension $k \times 1$ and time-varying coefficients, $\boldsymbol{\alpha}_t$, of dimension $p \times 1$. The design matrix $\mathbf{x}_t$ includes an intercept. The parameter-driven dynamics follow a random walk process which is given in equation

(2). This process is initialized with $\boldsymbol{\alpha}_0 = \mathbf{0}$ and $\mathbf{u}_0 \sim N(\mathbf{0}, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\Sigma}_0$ is a known initial state error variance.

In expression (3) time-varying volatility is captured by a stochastic volatility model, where h_t is the log-volatility at time t. The dynamics of h_t is governed by a stationary ($|\phi| < 1$) first-order autoregressive stochastic process with unconditional mean μ_h and unconditional variance $\sigma_\eta^2/(1-\phi^2)$; the parameter ϕ measures the persistence in log-volatilities and σ_η^2 is the variance of shock to the log-volatility. We also assume that both the error terms ε_t and η_t are independent for all t.

The variable y_t^* is latent. Instead, we observe the ordinal response variable y_t, where each y_t takes on any one of the J ordered values in the range $1, ..., J$. The unobserved variable y_t^* and the observed variable y_t are connected by

$$y_t = j \Leftrightarrow \quad \zeta_{j-1} < y_t^* \le \zeta_j, \;\; 1 \le \; j \le J. \tag{4}$$

To ensure a properly defined cumulative distribution function for y_t we assume $\zeta_j > \zeta_{j-1}$, $\forall j$, with $\zeta_0 = -\infty$ and $\zeta_J = +\infty$.

The model, given by the expressions (1)-(4) is the ordinal-response state space mixed model with stochastic volatility (OSSMM-SV model).

For identification reasons, some restrictions need to be imposed on the model. As a location normalization, we set $\zeta_1 = 0$. As a scale normalization we fix an additional cutpoint, setting $\zeta_{J-1} = 1$ (Chen and Dey, 2000)[1]. We also transform the cutpoints as follows

$$\zeta_j^* = \log\left(\tfrac{\zeta_j - \zeta_{j-1}}{1-\zeta_j}\right), \; j = 2, .., J-2, \tag{5}$$

with $\boldsymbol{\zeta}^*_{(2,J-2)} = (\zeta_2^*, ..., \zeta_{J-2}^*)'$. This reparameterization, due to Chen and Dey (2000) allows for an efficient way of simulating the ζ_j's.

We assume the following independent priors over the set of parameters ($\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\zeta}^*_{(2,J-2)}$, $\sigma_\eta^2, \mu_h, \phi$),

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \mathbf{B}), \boldsymbol{\Sigma} \sim IW(\delta, \Delta^{-1}), \; \boldsymbol{\zeta}^*_{(2,J-2)} \sim N(\boldsymbol{\mu}_{\zeta^*}, \Sigma_{\zeta^*}),$$

[1]For various identification schemes of ordinal-response models see Chen and Khan (2003), Hasegawa (2009) and Muller and Czado (2009).

$$\sigma_\eta^2 \sim \mathcal{IG}(v_a/2, v_\beta/2), \mu_h \sim N(\bar{\mu_h}, \bar{\sigma_h}^2), (\phi+1)/2 \sim Beta(\phi_a, \phi_\beta),$$

where IW and $\mathcal{IG}$ denote the Inverse-Wishart distribution and the inverse gamma distribution, respectively. The prior on $(\phi+1)/2$ ensures that the prior on ϕ has support on $(-1, 1)$. Furthermore, the reparametization in (5) allows us to place unrestricted priors over $\boldsymbol{\zeta}^*_{(2,J-2)}$. Therefore, for the transformed cutpoints $\boldsymbol{\zeta}^*_{(2,J-2)}$ we assume a multivariate normal prior.

3 Posterior analysis

3.1 MCMC algorithm

Define

$$\boldsymbol{y} = (y_1, ..., y_T),\ \boldsymbol{y}^* = (y_1^*, ..., y_T^*),\ \boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_T),\ \mathbf{h} = (h_1, ..., h_T).$$

The likelihood function of the proposed model is given by

$$L = p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\zeta}_{(2,J-2)}, \mathbf{h}) = \prod_{t=1}^{T}\prod_{j=1}^{J} P(y_t = j|\boldsymbol{\beta}, \boldsymbol{\alpha}_t, \zeta_{j-1}, \zeta_j, h_t)^{1(y_t=j)},$$

where

$$P(y_t = j|\boldsymbol{\beta}, \boldsymbol{\alpha}_t, \zeta_{j-1}, \zeta_j, h_t) = \Phi(\tfrac{\zeta_j - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t}{\exp(h_t/2)}) - \Phi(\tfrac{\zeta_{j-1} - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t}{\exp(h_t/2)}),$$

with $1(y_t = j)$ being an indicator function that equals one if $y_t = j$ and zero otherwise. Φ is the standard Gaussian c.d.f and $\boldsymbol{\zeta}_{(2,J-2)} = (\zeta_2, ..., \zeta_{J-2})'$.

The MCMC scheme for the OSSMM-SV model consists of updating the parameters ($\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\alpha}$, σ_η^2, μ_h, ϕ, $\boldsymbol{\zeta}^*_{(2,J-2)}$, $\boldsymbol{y}^*$, $\mathbf{h}$). We sample the state vector $\boldsymbol{\alpha}$, using the precision sampler of Chan and Jeliazkov (2009). To update the volatility vector $\mathbf{h}$ we apply the approach of Chan (2015). We update $\boldsymbol{\zeta}^*_{(2,J-2)}$ and $\boldsymbol{y}^*$ in one block, within an independence Metropolis-Hastings step in order to improve efficiency.

Details of the MCMC algorithm, along with a simulation study, are provided in the Online Appendix.

3.2 Forecast evaluation

To evaluate the performance of the proposed model we conduct a recursive out-of-sample forecasting exercise, using predictive likelihoods. Let $\boldsymbol{\Theta} = (\boldsymbol{y}^{*}, \boldsymbol{\alpha}, \mathbf{h}, \sigma_{\eta}^{2}, \mu_{h}, \phi, \boldsymbol{\zeta}_{(2,J-2)}^{*})$ denote the vector of all parameters in the model and $\boldsymbol{\Theta}^{(m)}$ be an MCMC sample of $\boldsymbol{\Theta}$ at iteration $m = 1...M$, after the burn-in period. The conditional predictive density for the (one-step ahead) y_{t+1} given $\boldsymbol{\Theta}^{(m)}$ and the data $\boldsymbol{\Omega}_{t} = (\boldsymbol{y}_{t}, \mathbf{X}_{t}, \mathbf{Z}_{t})$, where $\mathbf{X}_{t} = (\mathbf{x}_{1}, ..., \mathbf{x}_{t})$ and $\mathbf{Z}_{t} = (\mathbf{z}_{1}, ..., \mathbf{z}_{t})$ is given by

$$p(y_{t+1}|\boldsymbol{\Omega}_{t}, \boldsymbol{\Theta}^{(m)}) = \Phi(\tfrac{\zeta_{j}^{(m)} - \mathbf{x}'_{t+1}\boldsymbol{\beta}^{(m)} - \mathbf{z}'_{t+1}\boldsymbol{\alpha}_{t+1}^{(m)}}{\exp(h_{t+1}^{(m)}/2)}) - \Phi(\tfrac{\zeta_{j-1}^{(m)} - \mathbf{x}'_{t+1}\boldsymbol{\beta}^{(m)} - \mathbf{z}'_{t+1}\boldsymbol{\alpha}_{t+1}^{(m)}}{\exp(h_{t+1}^{(m)}/2)}).$$

By taking the average over the MCMC samples we can integrate out the model parameters to obtain the predictive density defined as

$$p(y_{t+1}|\boldsymbol{\Omega}_{t}) = \tfrac{1}{M}\textstyle\sum_{m=1}^{M} p(y_{t+1}|\boldsymbol{\Omega}_{t}, \boldsymbol{\Theta}^{(m)}).$$

Replacing y_{t+1} by the observed value y_{t+1}^{o}, we obtain the value $p(y_{t+1} = y_{t+1}^{o}|\boldsymbol{\Omega}_{t})$ which is called the predictive likelihood of y_{t+1}. Next we move one period ahead and repeat the same forecasting exercise with $\boldsymbol{\Omega}_{t+1}$ data. The log predictive score of the model for the evaluation period $t = t_{0} + 1, ..., T$ is the sum of the log predictive likelihoods $\sum_{t=t_{0}}^{T-1} \log p(y_{t+1} = y_{t+1}^{o}|\boldsymbol{\Omega}_{t})$. Higher values indicate better (out-of-sample) forecasting ability of the model.

The predictive likelihood $p(y_{t+1} = y_{t+1}^{o}|\boldsymbol{\Omega}_{t})$ is a natural measure to evaluate the density forecast $p(y_{t+1}|\boldsymbol{\Omega}_{t})$. We can also obtain the point forecast for y_{t+1} by producing an estimate for the predictive mean $E(y_{t+1}|\boldsymbol{\Omega}_{t})$. A usual metric for the evaluation of point forecasts is the root mean squared forecast error (RMSFE) defined as

$$RMSFE = \sqrt{\frac{\sum_{t=t_{0}}^{T-1}(y_{t+1}^{o} - E(y_{t+1}|\boldsymbol{\Omega}_{t}))^{2}}{T - t_{o}}}.$$

4 Empirical application

4.1 Data

To illustrate the proposed methodology we focus on the Federal funds rate target changes. In particular, we exploit the data set of the seminal paper of Hamilton and Jorda (2002). Using the Federal Open Market Committee (FOMC) meeting days, Hamilton and Jorda (2002) estimated an ordered probit model of monetary policy with five ordinal responses in order to capture the magnitude and direction of the target changes when they occurred. They used weekly data covering the period from the 1st week of February 1984 to the last week of April 2001. The explanatory variables used in their analysis included the magnitude of the last target change as of the previous week (y_{t-1}) and the spread between the 6-month Treasury bill rate and the Federal funds rate (SP_{t-1}).

We use the same explanatory variables but allow their coefficients to be time-varying, that is, $\mathbf{z}_{t-1} = (y_{t-1}, SP_{t-1})'$. Following also Hamilton and Jorda (2002) we characterize the monetary policy in terms of five regimes over the period 1984-2001, ranging from -0.50% (extreme easying) to +0.50% (extreme tightening), in steps of 0.25%; see Table 1 which displays the frequency of each monetary regime in our data set.

The OSSMM-SV model is compared against the same model but without the SV component (OSSMM model) and an ordinal-response with SV model that assumes time-constant coefficients (OR-SV model). The last 6 observations were used to calculate the log predictive scores (LPS) and the RMSFE.

In Table 2 we present our estimation results along with the Geweke (1992)'s Convergence Diagnostics (CD) and the Inefficiency Factor (IF). We run the sampler 150000 times after throwing away the first 50000 iterations. We use the same hyperparameters for the priors of the OSSMM-SV model as those used in the simulation study (Online Appendix).

4.2 Results

Based on the log predictive scores, reported in Table 2, the OSSMM-SV model, which accounts for conditional heteroscedasticity and time-varying coefficients provides better density forecasts than the rest of the models. By assuming time-constant conditional variance in the OSSMM-SV model, the forecast performance of the resulting model, the

OSSMM model, deteriorates. Similarly, by assuming time-constant coefficients in the OSSMM-SV model, the resulting model, the OR-SV model performs quite badly, failing to produce good density forecasts. The produced values of the RMSFE verify the above findings.

All the parameters across all models of Table 2 are statistically significant. Figures 1 displays the path of the estimated posterior means of the time-varying parameters along with their two standard deviation bands, obtained from the OSSMM-SV model. As can be seen from Figure 1 the effect of the previous target change (y_{t-1}) on the current's week target change is positive throughout the time period in question. So, it is more possible to have an increase of the target in this week than a decrease, if there was a target increase previously. Furthermore, the effect of the spread between the 6-month Treasury bill rate and the Federal funds rate (SP_{t-1}) is positive most of the time while it can be larger than the effect of y_{t-1}; see Figure 1.

Similar results were obtained from the OSSMM model (Figure 2).

5 Conclusions

We set up and estimated a discrete-response state space model with stochastic volatility. Bayesian methods were used to estimate the model parameters. We found that this model had better forecast performance than alternative specifications.

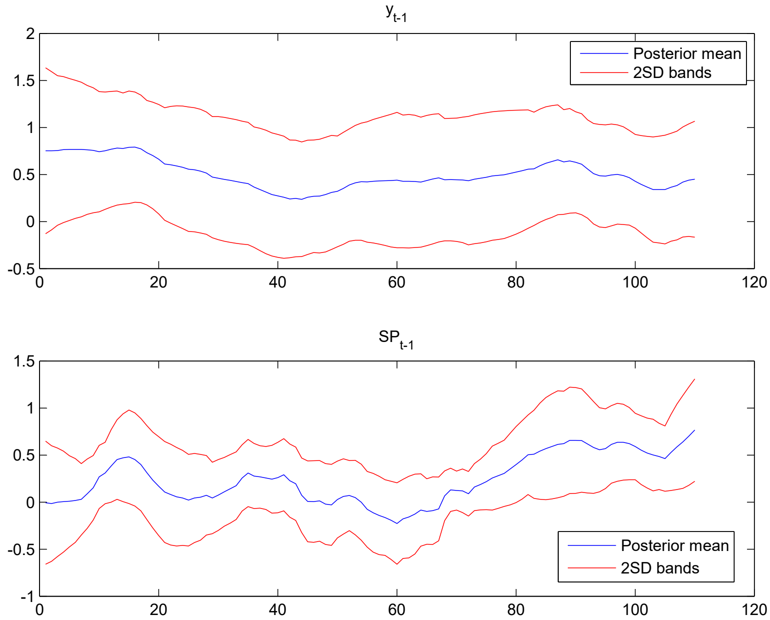

Figure 1: Empirical analysis: Path of the posterior means of the time-varying parameters obtained from the OSSMM-SV model

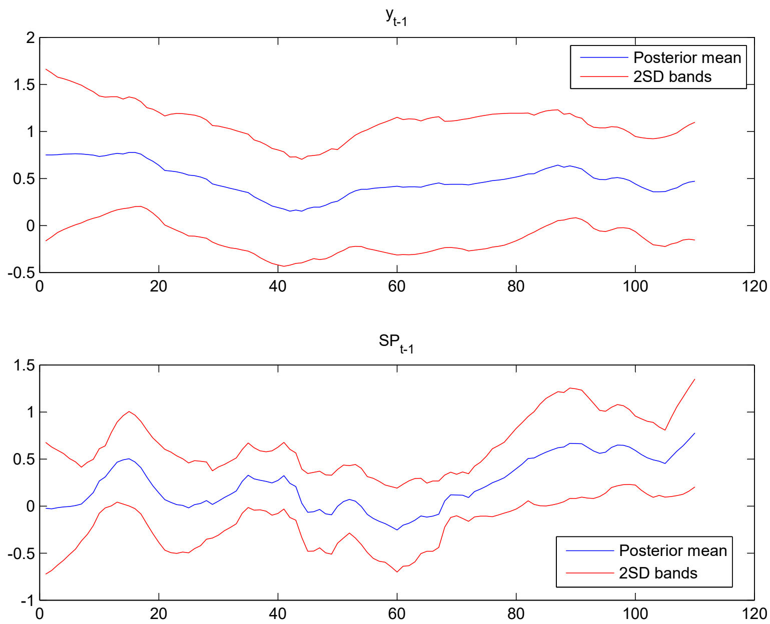

Figure 2: Empirical analysis: Path of the posterior means of the time-varying parameters obtained from the OSSMM model.

Table 1: Regimes of monetary policy

Dependent variable y_t	target change	frequency
1	-0.50 (extreme easing)	16
2	-0.20 (easing)	43
3	0 (no change)	14
4	0.20 (tightening)	34
5	0.50 (extreme tightening)	9

Table 2: Empirical results

Model	OSSMM-SV			OSSMM			OR-SV		
	Mean	CD	IF	Mean	CD	IF	Mean	CD	IF
$Const$	0.5550*	0.385	13.43	0.5439*	0.900	8.67	0.5622*	0.247	3.76
	(0.0587)			(0.0544)			(0.0698)		
y_{t-1}							0.7586*	0.398	2.30
							(0.1153)		
SP_{t-1}							0.2017*	0.979	2.22
							(0.0502)		
$\mathbf{\Sigma}_{y_{t-1}}$	0.0200*	0.685	35.93	0.0206*	0.657	34.03			
	(0.0136)			(0.0139)					
$\mathbf{\Sigma}_{SP_{t-1}}$	0.0199*	0.477	36.15	0.0227*	0.233	39.43			
	(0.0118)			(0.0134)					
ϕ	0.9108*	0.037	6.72				0.9082*	0.231	6.52
	(0.0597)						(0.0603)		
μ_h	-2.7447*	0.618	3.74				-2.3186*	0.643	2.75
	(0.5760)						(0.5477)		
σ_η	0.2007*	0.412	30.22				0.2044*	0.480	30.64
	(0.0515)						(0.0535)		
σ				0.2483*	0.250	6.80			
				(0.0266)					
ζ_2	0.4358*	0.842	5.61	0.4340*	0.704	4.38	0.4156*	0.035	4.30
	(0.0477)			(0.0457)			(0.0540)		
ζ_3	0.5749*	0.803	6.05	0.5709*	0.696	4.83	0.5516*	0.030	4.24
	(0.0492)			(0.0469)			(0.0561)		
LPS	-4.8703			-4.8893			-8.9303		
$RMSFE$	0.5596			0.5618			2.3113		

*Significant based on the 95% highest posterior density interval. Standard errors in parentheses.

References

Abanto-Valle, C.A., Dey, D.K. (2014). State space mixed models for binary responses with scale mixture of normal distributions links. *Computational Statistics & Data Analysis* **71**: 274 – 287.

Ahmed, J. (2016). A conditionally heteroskedastic binary choice model for macro-financial time series. *Journal of Statistical Computation and Simulation* **86**(10): 2007–2035.

Bollerslev, T., Chou, R.Y., Kroner, K.F. (1992). ARCH modelling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* **52**(1-2): 5–59.

Chan, J.C.C. (2015). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics, forthcoming.*

Chan, J.C.C, Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation* **1**: 101–120.

Chen, M.H., Dey, D.K. (2000) *. Bayesian analysis for correlated ordinal data models.* In Dipak DK, Ghosh SK and Mallik BK, (Eds.), Generalized Linear Models: A Bayesian Perspective, New York: Marcel-Dekker, pp. 133-157.

Chen, S., Khan, S. (2003). Rates of convergence for estimating regression coefficients in heteroskedastic discrete response models. *Journal of Econometrics* **117**(2): 245 – 278.

Czado, C., Song, P.X.K. (2008). State space mixed models for longitudinal observations with binary and binomial responses. *Statistical Papers* **49**(2): 691714.

Dueker, M. (1999). Conditional heteroscedasticity in qualitative response models of time series: A Gibbs-sampling approach to the bank prime rate. *Journal of Business and Economic Statistics* **17**(4): 466–472.

Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *Journal of the American Statistical Association* **87**(418): 501–509.

Hamilton, J., Jorda, O. (2002). A model for the federal funds rate target. *Journal of Political Economy* **110**(5): 1135–1167.

Hasegawa, H. (2009). Bayesian dynamic panel-ordered probit model and its application to subjective well-being. *Communications in Statistics - Simulation and Computation* **38**(6): 1321–1347.

Hausman, J.A., Lo, A.W., MacKinlay, A.C. (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics* **31**(3): 319 – 379.

Hsieh, PH., Yang, JJ. (2009). A censored stochastic volatility approach to the estimation of price limit moves. *Journal of Empirical Finance* **16**(2): 337–351.

Huang, HC., Lin, SC. (2006). Time-varying discrete monetary policy reaction functions. *Applied Economics* **38**(4): 449–464.

Müller, G., Czado, C. (2009). Stochastic volatility models for ordinal-valued time series with application to finance. *Statistical Modelling* **9**(1): 69–95.

Song, P.X.K. (2000). Monte Carlo kalman filter and smoothing for multivariate discrete state space models. *Canadian Journal of Statistics* **28**(3): 641–652.

Stefanescu, C., Tunaru, R., Turnbull, S. (2009). The credit rating process and estimation of transition probabilities: A Bayesian approach. *Journal of Empirical Finance* **16**(2): 216 – 234.

Taylor, S. (1986). *Modelling Financial Time Series.* Chichester: Wiley.

West, M., Harrison, P.J., Migon, H.S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association* **80**(389): 73–83.

Yang, JW., Parwada, J. (2012). Predicting stock price movements: An ordered probit analysis on the Australian securities exchange. *Quantitative Finance* **12**(5): 791–804.

Online Appendix for: Discrete-response state space models with conditional heteroscedasticity: An application to forecasting the federal funds rate target

Stefanos Dimitrakopoulos*[1] and Dipak K. Dey[2]

[1]Department of Accounting, Finance and Economics, Oxford Brookes University, UK

[2]Department of Statistics, University of Connecticut, USA

1 A simulation experiment

In order to assess the performance of the proposed MCMC estimation procedure we generated T=1000 observations from the OSSMM-SV model. We assumed $J = 7$ ordered choices, $k = 3$ constant coefficients (including the intercept), $p = 2$ time-varying coefficients and the following set of true parameter values

$$\boldsymbol{\beta} = (0, 1, -0.8)', \boldsymbol{\Sigma} = diag(0.1, 0.03), \boldsymbol{\alpha}_1 = (2, -1)', \phi = 0.8,\ \mu_h = 0.9,$$

$$\zeta_2 = 0.2,\ \zeta_3 = 0.4,\ \zeta_4 = 0.6,\ \zeta_5 = 0.8,\ \sigma_\eta^2 = 0.01,$$

where $diag(\cdot)$ is a diagonal matrix. The elements (x_{1t}, x_{2t}) of $\mathbf{x}_t = (1, x_{1t}, x_{2t})'$ as well as $\mathbf{z}_t = (z_{1t}, z_{2t})'$ for $t = 1, ..., T$ are generated respectively as $x_{jt} \sim U(0, 1) - 0.5$ and $z_{it} \sim 2 * U(0, 1) - 0.5$, $j, i = 1, 2$, where U is the uniform distribution.

Furthermore, we assume the following proper (but sufficiently diffuse) prior distributions

*Correspondence to: Stefanos Dimitrakopoulos, Oxford Brookes University, Department of Accounting, Finance and Economics, Oxford, OX33 1HX, UK, Tel: +44(0) 1865 485478, E-mail: sdimitrakopoulos@brookes.ac.uk.

$$\boldsymbol{\beta} \sim N(0, 20 \times I_{3\times3}),\ \boldsymbol{\alpha}_1 \sim N(0, 20 \times I_{2\times2}),\ \boldsymbol{\Sigma} \sim IW(1, 20 \times I_{2\times2}),$$

$$(\phi + 1)/2 \sim Beta(80, 14),\ \sigma_\eta^2 \sim \mathcal{IG}(50/2, 0.5/2),\ \mu_h \sim N(0, 100),$$

$$\boldsymbol{\zeta}^*_{(2,5)} \sim N(0, 20 \times I_{4\times4}),$$

where $I_{i\times i}$ is a $i \times i$ identity matrix.

We run our sampler for 60000 iterations with a burn-in of 30000 iterations. Table 1 reports the estimates of the posterior means and standard deviations for all the parameters. Furthermore, we monitor convergence and the efficiency of the posterior simulators, using the CD statistics of Geweke (1992) and the inefficiency factor (IF)- see, for example, Chib (2001)-, respectively.

The sampler for the OSSMM-SV model provides satisfactory estimation results for all the parameters. It is worth noting that the sample autocorrelations for the cutpoints (not shown), decay very quickly within the first few iterations. This is also verified by the small values of the inefficiency factors for the cutpoints (see Table 1). Based on CD statistics, the null hypothesis that convergence to the posterior distribution has been achieved can not be rejected for the estimated parameters at the 5% significance level. The inefficiency factors are quite low for all the parameters, indicating an efficient sampling, except for the parameters Σ_{11}, Σ_{22}, μ_h and σ_η. However, due to the M=60000 iterations (after the burn-in period), we obtain sufficient uncorrelated samples for posterior inference.

The paths for the posterior means of α_{1t} and α_{2t}, obtained from the OSSMM-SV model are presented in Figure 1. As can be seen from these figure, the posterior means follow closely the true paths of α_{1t} and α_{2t} and almost all the true values are contained within the two standard deviation bands.

Table 1: Simulation results for the OSSMM-SV model

True values	Mean	CD	IF
$\beta_0 = 0(Const)$	0.0426	0.695	30.98
	(0.016)		
$\beta_1 = 1$	1.2135	0.241	81.72
	(0.2125)		
$\beta_2 = -0.8$	-0.8133	0.333	71.12
	(0.1592)		
$\mathbf{\Sigma}_{11} = 0.1$	0.1004	0.171	172.68
	(0.0483)		
$\mathbf{\Sigma}_{22} = 0.03$	0.0477	0.723	155.99
	(0.0365)		
$\phi = 0.8$	0.8573	0.499	37.76
	(0.0610)		
$\mu_h = 0.9$	0.8626	0.277	164.65
	(0.2204)		
$\sigma_\eta = 0.1$	0.1732	0.034	145.85
	(0.0355)		
$\zeta_2 = 0.2$	0.1343	0.966	2.05
	(0.0357)		
$\zeta_3 = 0.4$	0.3719	0.148	2.14
	(0.0510)		
$\zeta_4 = 0.6$	0.6214	0.080	2.22
	(0.0532)		
$\zeta_5 = 0.8$	0.7956	0.017	2.26
	(0.0486)		

Standard errors in parentheses.

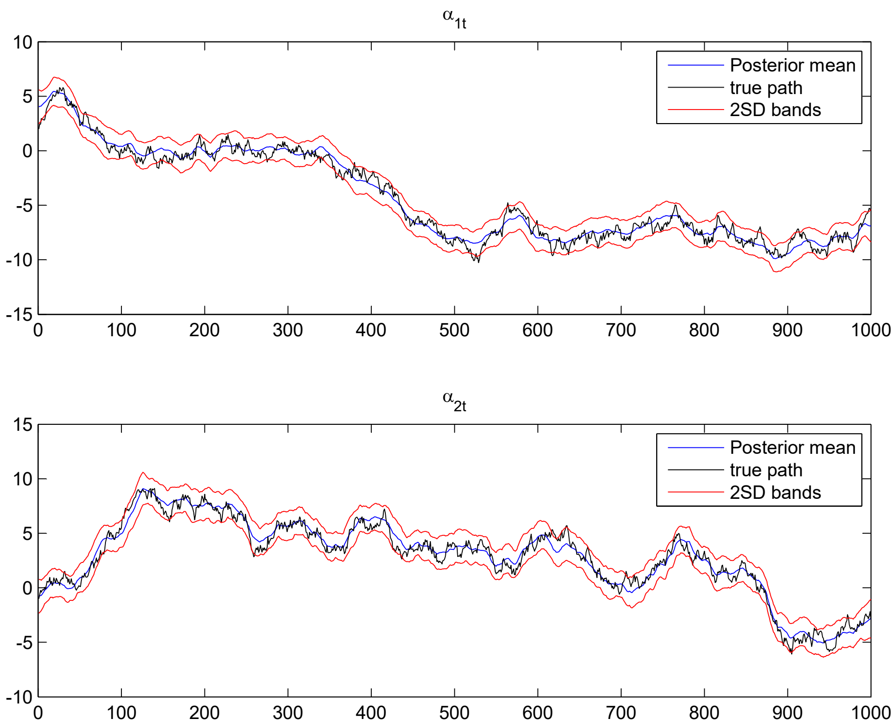

Figure 1: Simulated data: Path of the estimated α_{1t} and α_{2t} for the OSSMM-SV model; T=1000. True path (black), posterior mean (blue), two standard deviation bands (red).

2 MCMC algorithm for the OSSMM-SV model

Posterior sampling of $\boldsymbol{\beta}$

Update $\boldsymbol{\beta}$ by sampling from

$$\boldsymbol{\beta}|\mathbf{B}, \boldsymbol{\beta}_0, \boldsymbol{\alpha}, \boldsymbol{y}^{*}, \mathbf{h} \sim N(D_0 d_0, D_0),$$

where

$$D_0 = \left(\mathbf{B}^{-1} + \sum_{t=1}^{T} \frac{\mathbf{x}_t \mathbf{x}_t'}{\exp(h_t)}\right)^{-1}, \; d_0 = \mathbf{B}^{-1}\boldsymbol{\beta}_0 + \sum_{t=1}^{T} \frac{\mathbf{x}_t (y_t^{*} - \mathbf{z}_t'\boldsymbol{\alpha}_t)}{\exp(h_t)}.$$

Posterior sampling of $\boldsymbol{\Sigma}$

Update $\boldsymbol{\Sigma}$ by sampling from

$$\boldsymbol{\Sigma}|\delta, \Delta, \boldsymbol{\alpha} \sim IW\left(\delta + T - 1, \Delta^{-1} + \sum_{t=1}^{T-1} (\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t)(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t)'\right).$$

Posterior sampling of $\boldsymbol{\alpha}$

Stacking $y_t^{*} = \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{z}_t'\boldsymbol{\alpha}_t + \varepsilon_t$ over $t = 1, ..., T$, we have

$$\boldsymbol{y}^{*} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{S}_{y^{*}}) \Leftrightarrow$$

$$\boldsymbol{y}^{*} - \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \Leftrightarrow$$

$$\tilde{\boldsymbol{y}}^{*} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

where $\tilde{\boldsymbol{y}}^{*} = \boldsymbol{y}^{*} - \mathbf{X}\boldsymbol{\beta}$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_T)'$, $\mathbf{X} = (\mathbf{x}_1', ..., \mathbf{x}_T')'$,

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1' & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_2' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{z}_3' & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{z}_T' \end{pmatrix}, \boldsymbol{S}_{y^{*}} = \begin{pmatrix} \exp(h_1) & 0 & 0 & \cdots & 0 \\ 0 & \exp(h_2) & 0 & \cdots & 0 \\ 0 & 0 & \exp(h_3) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \exp(h_T) \end{pmatrix}.$$

The state equation $\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \mathbf{u}_t$, can be rewritten in a matrix notation as follows,

$$\boldsymbol{H}\boldsymbol{\alpha} = \tilde{\boldsymbol{\delta}}_{\boldsymbol{\alpha}} + \mathbf{u}, \mathbf{u} \sim N(\mathbf{0}, \boldsymbol{S}_u),$$

where $\tilde{\boldsymbol{\delta}}_{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}_0, \mathbf{0}, ..., \mathbf{0})'$, $\mathbf{u} = (\mathbf{u}_1, ..., \mathbf{u}_T)'$, $\boldsymbol{S}_u = diag(\boldsymbol{\Sigma}_0, ..., \boldsymbol{\Sigma})$ and $\boldsymbol{H}$ is the first difference matrix

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{I}_p & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{I}_p & \boldsymbol{I}_p & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_p & \boldsymbol{I}_p & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{I}_p & \boldsymbol{I}_p \end{pmatrix}.$$

Hence, the prior distribution of $\boldsymbol{\alpha}$ is Gaussian, that is, $\boldsymbol{\alpha}|\boldsymbol{\Sigma} \sim N(\boldsymbol{\delta}_{\boldsymbol{\alpha}}, (\boldsymbol{H}'\boldsymbol{S}_u^{-1}\boldsymbol{H})^{-1})$, where $\boldsymbol{\delta}_{\boldsymbol{\alpha}} = \boldsymbol{H}^{-1}\tilde{\boldsymbol{\delta}}_{\boldsymbol{\alpha}}$. The conditional posterior distribution of $\boldsymbol{\alpha}$ is also Gaussian,

$$\boldsymbol{\alpha}|\boldsymbol{\Sigma}, \boldsymbol{y^*} \sim N(\hat{\boldsymbol{\alpha}}, \boldsymbol{D}_{\boldsymbol{\alpha}}^{-1}),$$

where
$$\boldsymbol{D}_{\boldsymbol{\alpha}} = \boldsymbol{H}'\boldsymbol{S}_u^{-1}\boldsymbol{H} + \mathbf{Z}'\boldsymbol{S}_{y^*}^{-1}\mathbf{Z} \tag{A.1}$$
and
$$\hat{\boldsymbol{\alpha}} = \boldsymbol{D}_{\boldsymbol{\alpha}}^{-1}(\boldsymbol{H}'\boldsymbol{S}_u^{-1}\boldsymbol{H}\boldsymbol{\delta}_{\boldsymbol{\alpha}} + \mathbf{Z}'\boldsymbol{S}_{y^*}^{-1}\tilde{\boldsymbol{y}}^*). \tag{A.2}$$

Typically, $\boldsymbol{D}_{\boldsymbol{\alpha}}$ is a high-dimensional covariance matrix and sampling from the posterior distribution of $\boldsymbol{\alpha}$ can be time-consuming. Yet, since the precision matrix $\boldsymbol{D}_{\boldsymbol{\alpha}}$ is a band matrix, we can efficiently sample from $N(\hat{\boldsymbol{\alpha}}, \boldsymbol{D}_{\boldsymbol{\alpha}}^{-1})$, using the precision sampler of Chan and Jeliazkov (2009) that exploits block-banded and sparse matrix algorithms, instead of Kalman-filter based methods

In particular, compute $\boldsymbol{D}_{\boldsymbol{\alpha}}$, using (A.1) and obtain its Cholesky factor C, such that $C'C = \boldsymbol{D}_{\boldsymbol{\alpha}}$. Then, we proceed to the smoothing step, where we solve

$$C'C\hat{\boldsymbol{\alpha}} = \boldsymbol{H}'\boldsymbol{S}_u^{-1}\boldsymbol{H}\boldsymbol{\delta}_{\boldsymbol{\alpha}} + \mathbf{Z}'\boldsymbol{S}_{y^*}^{-1}\tilde{\boldsymbol{y}}^*,$$

by forward and backward substitution to obtain $\hat{\boldsymbol{\alpha}}$. The final step is the simulation step, where we sample $\boldsymbol{z} \sim N(0, \boldsymbol{I})$, solve $\boldsymbol{C}'\boldsymbol{x} = \boldsymbol{z}$ for $\boldsymbol{x}$ by backward substitution and return

$\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}} + \boldsymbol{x}$, so that $\boldsymbol{\alpha} \sim N(\hat{\boldsymbol{\alpha}}, \boldsymbol{D}_{\boldsymbol{\alpha}}^{-1})$.

Posterior sampling of h

Apply the sampler of Chan (2015) to the following model

$$
\begin{aligned}
y_t^\star &= \exp(h_t/2)\epsilon_t,\ \epsilon_t \sim N(0,1),\ t = 1, ..., T, \\
h_t &= \mu_h + \phi(h_{t-1} - \mu_h) + \eta_t,\ |\phi| < 1,\ \eta_t \sim N(0, \sigma_\eta^2),
\end{aligned}
$$

with $cov(\epsilon_t, \eta_t) = 0$, $y_t^\star = y_t^* - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{z}_t'\boldsymbol{\alpha}_t$ and initial condition $h_1 \sim N(\mu_h, \sigma_\eta^2/(1-\phi^2))$.

To be more specific, the posterior distribution of the volatility vector $\mathbf{h}$ is given by

$$
p(\mathbf{h}|\phi, \sigma_\eta^2, \mu_h, \boldsymbol{\beta}, \boldsymbol{y}^\star, \boldsymbol{\alpha}) \propto p(\boldsymbol{y}^\star|\boldsymbol{\beta}, \mathbf{h}, \boldsymbol{\alpha}) p(\mathbf{h}|\phi, \sigma_\eta^2, \mu_h), \tag{A.3}
$$

where $\boldsymbol{y}_t^\star = (y_1^\star, ..., y_T^\star)$. In order to sample from the posterior distribution $p(\mathbf{h}|\phi, \sigma_\eta^2, \mu_h, \boldsymbol{\beta}, \boldsymbol{y}^\star, \boldsymbol{\alpha})$, we approximate it by a Gaussian distribution, which is then used as a proposal density within the Acceptance-Rejection Metropolis-Hastings (ARMH) algorithm (see, for example, Tierney (1994) and Chib and Greenberg (1995)). Candidate draws from the Gaussian approximation are generated using the precision sampler of Chan and Jeliazkov (2009), instead of Kalman-filter based methods. In particular, it can be shown that the density $p(\mathbf{h}|\phi, \sigma_\eta^2, \mu_h)$ in expression (A.3) is Gaussian; that is, $\mathbf{h}|\phi, \sigma_\eta^2, \mu_h \sim N(\boldsymbol{H}^{-1}\hat{\mathbf{h}}, (\boldsymbol{H}'\Sigma^{-1}\boldsymbol{H})^{-1})$, where $\hat{\mathbf{h}} = (\mu_h, (1-\phi)\mu_h, ..., (1-\phi)\mu_h)'$, $\Sigma = diag(\sigma_\eta^2/(1-\phi^2), \sigma_\eta^2, ..., \sigma_\eta^2)$ and $\boldsymbol{H}$ is a lower triangular sparse matrix (with determinant 1-hence, it is invertible)

$$
\boldsymbol{H} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi & 1 & 0 & \cdots & 0 \\ 0 & -\phi & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\phi & 1 \end{pmatrix}.
$$

The logarithm of $p(\mathbf{h}|\phi, \sigma_\eta^2, \mu_h)$ can be written as

$$\log p(\mathbf{h}|\phi,\sigma_\eta^2,\mu_h) = const - \tfrac{1}{2}(\mathbf{h}'\boldsymbol{H}'\Sigma^{-1}\boldsymbol{H}\mathbf{h} - 2\mathbf{h}'\boldsymbol{H}'\Sigma^{-1}\boldsymbol{H}\boldsymbol{H}^{-1}\hat{\mathbf{h}}). \tag{A.4}$$

The density $p(\boldsymbol{y}^\star|\boldsymbol{\beta},\mathbf{h},\boldsymbol{\alpha})$ in expression (A.3) can also be approximated by a normal density. By taking the second order Taylor expansion of the logarithm of $p(\boldsymbol{y}^\star|\boldsymbol{\beta},\mathbf{h},\boldsymbol{\alpha})$ around $\tilde{\mathbf{h}}$, which is the mode of the posterior $\log p(\mathbf{h}|\phi,\sigma_\eta^2,\mu_h,\boldsymbol{\beta},\boldsymbol{y}^\star,\boldsymbol{\alpha})$ (see below), we have,

$$\begin{aligned}\log p(\boldsymbol{y}^\star|\boldsymbol{\beta},\mathbf{h},\boldsymbol{\alpha}) &\approx \log p(\boldsymbol{y}^\star|\boldsymbol{\beta},\boldsymbol{\alpha},\tilde{\mathbf{h}}) + (\mathbf{h}-\tilde{\mathbf{h}})'\mathbf{f} - \tfrac{1}{2}(\mathbf{h}-\tilde{\mathbf{h}})'\mathbf{G}(\mathbf{h}-\tilde{\mathbf{h}}),\\ &= const - \tfrac{1}{2}(\mathbf{h}'\mathbf{G}\mathbf{h} - 2\mathbf{h}'(\mathbf{f}+\mathbf{G}\tilde{\mathbf{h}})),\end{aligned} \tag{A.5}$$

where $\mathbf{f} = (f_1,...,f_T)'$ is the gradient vector with $f_t = \frac{d\log p(y_t^\star|\boldsymbol{\beta},\boldsymbol{\alpha}_t,h_t)}{dh_t} = -\frac{1}{2}+\frac{1}{2}y_t^{2\star}\exp(-h_t)$ evaluated at $\tilde{h}_t$, $t=1,...,T$ and $\mathbf{G} = diag(G_1,...,G_T)$ is the diagonal negative Hessian of the $\log p(\boldsymbol{y}^\star|\boldsymbol{\beta},\mathbf{h},\boldsymbol{\alpha})$, with $G_t = -\frac{d^2\log p(y_t^\star|\boldsymbol{\beta},\boldsymbol{\alpha}_t,h_t)}{dh_t^2} = \frac{1}{2}y_t^{2\star}\exp(-h_t)$ evaluated at $\tilde{h}_t$, $t=1,...,T$.

From (A.4) and (A.5) the logarithm of the posterior distribution of the volatility vector becomes

$$\log p(\mathbf{h}|\phi,\sigma_\eta^2,\mu_h,\boldsymbol{\beta},\boldsymbol{y}^\star,\boldsymbol{\alpha}) \approx const - \tfrac{1}{2}(\mathbf{h}'\boldsymbol{K_h}\mathbf{h} - 2\mathbf{h}'\boldsymbol{k_h}) = \log\ g(\mathbf{h}), \tag{A.6}$$

where $\boldsymbol{K_h} = \boldsymbol{H}'\Sigma^{-1}\boldsymbol{H}+\mathbf{G}$, $\boldsymbol{k_h} = \mathbf{f}+\mathbf{G}\tilde{\mathbf{h}}+\boldsymbol{H}'\Sigma^{-1}\boldsymbol{H}\boldsymbol{H}^{-1}\hat{\mathbf{h}}$ and $g(\mathbf{h}) \propto N(\hat{\boldsymbol{m}},\boldsymbol{K_h}^{-1})$, with $\hat{\boldsymbol{m}} = \boldsymbol{K_h}^{-1}\boldsymbol{k_h}$. In other words, the posterior $p(\mathbf{h}|\phi,\sigma_\eta^2,\mu_h,\boldsymbol{\beta},\boldsymbol{y}^\star,\boldsymbol{\alpha})$ can be approximated by a normal density with mean $\hat{\boldsymbol{m}}$ and precision matrix $\boldsymbol{K_h}$. This Gaussian approximation is then used as a proposal density in the ARMH step, where candidate values are obtained using the precision sampler of Chan and Jeliazkov (2009), instead of Kalman-filter based methods.

Typically, $N(\hat{\boldsymbol{m}},\boldsymbol{K_h}^{-1})$ is a high-dimensional density and sampling from it can be time-consuming. Here, we use the precision-based sampler of Chan and Jeliazkov (2009), which exploits the fact that the precision matrix $\boldsymbol{K_h}$ is a band matrix since $\boldsymbol{H}'\Sigma^{-1}\boldsymbol{H}$ and $\mathbf{G}$ are also band matrices. In particular, $\boldsymbol{K_h}$ is a tridiagonal matrix as its non-zero elements appearing only on the main diagonal and the diagonals above and below the main one. Consequently, we can compute fast and efficiently the mean $\hat{\boldsymbol{m}}$ without calculating the inverse $\boldsymbol{K_h}^{-1}$, by solving the linear system $\boldsymbol{K_h}\hat{\boldsymbol{m}} = \boldsymbol{k_h}$. Furthermore, a draw $\tilde{\boldsymbol{m}}$ from

$N(\hat{\boldsymbol{m}}, \boldsymbol{K_h}^{-1})$ can be obtained, using the precision sampler of Chan and Jeliazkov (2009): calculate the Cholesky factor $\boldsymbol{C}$ of $\boldsymbol{K_h}$ such that $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{K_h}$, sample T independent standard normal draws, $\boldsymbol{z} \sim N(0, \boldsymbol{I})$, solve $\boldsymbol{C}'\boldsymbol{x} = \boldsymbol{z}$ for $\boldsymbol{x}$ by backward substitution and return $\tilde{\boldsymbol{m}} = \hat{\boldsymbol{m}} + \boldsymbol{x}$.

The point $\tilde{\mathbf{h}}$ around of which the second order Taylor expansion is taken in expression (A.5) is desirable to be the mode of the posterior $\log p(\mathbf{h}|\phi, \sigma_\eta^2, \mu_h, \boldsymbol{\beta}, \boldsymbol{y}^\star, \boldsymbol{\alpha})$ for an efficient sampling. The negative Hessian of this posterior distribution evaluated at $\mathbf{h} = \tilde{\mathbf{h}}$ is $\boldsymbol{K_h}$ and the gradient evaluated at $\mathbf{h} = \tilde{\mathbf{h}}$ is -$\boldsymbol{K_h}\tilde{\mathbf{h}}$ +$\boldsymbol{k_h}$. To find the mode, we apply the Newton-Raphson method as follows: 1) Initialize $\mathbf{h} = \tilde{\mathbf{h}}^{(1)}$ for some constant vector $\tilde{\mathbf{h}}^{(1)}$. 2) Set $\tilde{\mathbf{h}} = \tilde{\mathbf{h}}^{(l)}$ for $l = 1, 2, ...,$ and compute $\boldsymbol{K_h}$, $\boldsymbol{k_h}$ and $\mathbf{h}^{(l+1)} = \mathbf{h}^{(l)} + \boldsymbol{K_h}^{-1}(-\boldsymbol{K_h}\mathbf{h}^{(l)} + \boldsymbol{k_h}) = \boldsymbol{K_h}^{-1}\boldsymbol{k_h}$. This process is repeated until convergence is achieved.

Posterior sampling of ϕ

We sample from $p(\phi|\sigma_\eta^2, \mathbf{h}, \phi_a, \phi_\beta, \mu_h)$ using an independence Metropolis-Hastings algorithm. At the ith iteration we generate a proposed value $\phi^{*(p)}$ from the truncated normal in the interval $[-1, 1]$,

$$\phi^{*(p)}|\mathbf{h}, \sigma_\eta^2, \mu_h \sim TN_{[-1,1]}\left(\frac{\sum_{t=2}^T(h_t-\mu_h)(h_{t-1}-\mu_h)}{\sum_{t=3}^T(h_{t-1}-\mu_h)^2}, \frac{\sigma_\eta^2}{\sum_{t=3}^T(h_{t-1}-\mu_h)^2}\right).$$

Given the accepted value $\phi^{(i-1)}$ from the previous $(i-1)th$ iteration, we accept $\phi^{*(p)}$ as a valid current value ($\phi^{(i)} = \phi^{*(p)}$) with probability

$$a_p(\phi^{(i-1)}, \phi^{*(p)}) = \min\left(\frac{p(\phi^{*(p)})\sqrt{1-\phi^{2*(p)}}}{p(\phi^{(i-1)})\sqrt{1-\phi^{2(i-1)}}}, 1\right),$$

where $p(\phi) \propto (\frac{\phi+1}{2})^{\phi_a-1}(\frac{1-\phi}{2})^{\phi_\beta-1}$ is the prior of ϕ.

Posterior sampling of μ_h

We sample μ_h from

$$\mu_h|\phi, \sigma_\eta^2, \mathbf{h}, \bar{\mu_h}, \bar{\sigma_h}^2 \sim N(D_1 d_1, D_1),$$

where

$$D_1 = \left(\frac{\sigma_\eta^2+\bar{\sigma_h}^2[(T-1)(1-\phi)^2+1-\phi^2]}{\bar{\sigma_h}^2\sigma_\eta^2}\right)^{-1}, \; d_1 = \frac{\bar{\mu_h}}{\bar{\sigma_h}^2} + \frac{h_1(1-\phi^2)}{\sigma_\eta^2} + \frac{(1-\phi)\sum_{t=2}^{T}(h_t-\phi h_{t-1})}{\sigma_\eta^2}.$$

Posterior sampling of $\boldsymbol{\zeta}^*_{(2,J-2)}$ and $\boldsymbol{y}^*$ in one block

To improve the mixing of the proposed MCMC algorithm, we first sample the transformed cutpoints $\boldsymbol{\zeta}^*_{(2,J-2)}$, marginalized over the latent variables $\boldsymbol{y}^*$, using a Metropolis-Hastings algorithm. We, then, calculate the cutpoints ζ_j, from $\zeta_j = \frac{\zeta_{j-1}+\exp\zeta_j^*}{1+\exp\zeta_j^*}$, $j = 2, .., J-2$. Next, given the updated ζ_j's, we sample the latent dependent variable y_t^* , $t = 1, ..., T$ from

$$y_t^* \mid y_t = j, \boldsymbol{\beta}, \boldsymbol{\alpha}_t, \zeta_{j-1}, \zeta_j, h_t \sim TN_{(\zeta_{j-1},\zeta_j]}(\mathbf{x}_t'\boldsymbol{\beta} + \mathbf{z}_t'\boldsymbol{\alpha}_t, \exp(h_t)),$$

where TN is the truncated normal distribution with support defined by the threshold parameters ζ_{j-1} and ζ_j.

The proposed Metropolis-Hastings algorithm for sampling the $\boldsymbol{\zeta}^*_{(2,J-2)}$ works as follows. The conditional distribution of $p(\boldsymbol{\zeta}^*_{(2,J-2)}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{h})$ is defined as

$$p(\boldsymbol{\zeta}^*_{(2,J-2)}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{h}) = p(\boldsymbol{\zeta}^*_{(2,J-2)})p(\boldsymbol{\zeta}_{(2,J-2)}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{h}) \times \prod_{j=2}^{J-2} \frac{(1-\zeta_{j-1})\exp\zeta_j^*}{(1+\exp\zeta_j^*)2},$$

where the last term of the above expression is the Jacobian of the transformation from $\boldsymbol{\zeta}_{(2,J-2)}$ to $\boldsymbol{\zeta}^*_{(2,J-2)}$ and the second term is the full conditional distribution of the cutpoints $\boldsymbol{\zeta}_{(2,J-2)}$ given by

$$p(\boldsymbol{\zeta}_{(2,J-2)}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{h}) \propto \prod_{t:y_t=2} P(\zeta_1 < y_t^* \leq \zeta_2) \times ... \quad ... \times \prod_{t:y_t=J-1} P(\zeta_{J-2} < y_t^* \leq \zeta_{J-1}).$$

The multivariate Student-t distribution

$$MVt(\boldsymbol{\zeta}^*_{(2,J-2)}| \; \widehat{\boldsymbol{\zeta}^*_{(2,J-2)}}, c\widehat{\Sigma}_{\boldsymbol{\zeta}^*_{(2,J-2)}}, v),$$

is used as a proposal distribution, where $\widehat{\boldsymbol{\zeta}^*_{(2,J-2)}} = arg\ max\ p(\boldsymbol{\zeta}^*_{(2,J-2)}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{h})$ is defined to be the mode of the right hand side of $p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)$ and the term

$$\widehat{\Sigma}_{\boldsymbol{\zeta}^*_{(2,J-2)}} = \left[\left(-\frac{\vartheta^2 \log p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)}{\vartheta\boldsymbol{\zeta}^*_{(2,J-2)}\vartheta\boldsymbol{\zeta}^{*\prime}_{(2,J-2)}}\right)_{\boldsymbol{\zeta}^*_{(2,J-2)}=\widehat{\boldsymbol{\zeta}^*_{(2,J-2)}}}\right]^{-1},$$

is the inverse of the negative Hessian matrix of $\log p(\boldsymbol{\zeta}^*_{(2,J-2)}|\bullet)$, scaled by some arbitrary number $c > 0$. The term v is the degrees of freedom and is specified arbitrarily at the onset of the programming along with the scalar c and the other priors.

Let $\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}$ be the accepted value of $\boldsymbol{\zeta}^*_{(2,J-2)}$ at the previous ($l-1$-th) iteration. At the l-th iteration generate a value $\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}$ from $MVt(\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}|\bullet)$. The transition probability from $\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}$ to $\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}$ is

$$a_p(\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}, \boldsymbol{\zeta}^{*(p)}_{(2,J-2)}) = \min(\frac{p(\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{h}) \ MVt(\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}|\bullet)}{p(\boldsymbol{\zeta}^{*(l-1)}_{(2,J-2)}|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{h}) \ MVt(\boldsymbol{\zeta}^{*(p)}_{(2,J-2)}|\bullet)}, 1).$$

References

Chan, J.C.C. (2015). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics, forthcoming.*

Chan, J.C.C, Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation* **1**: 101–120.

Chib, S. (2001) . *Markov Chain Monte Carlo methods: Computation and Inference.* In Heckman JJ. and Leamer E, (Eds.), Handbook of Econometrics, Volume 5, pp. 3569-3649. Elsevier.

Chib, S., Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm. *The American Statistician* **49**(4): 327–335.

Geweke, J. (1992). *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments.* In Bernardo J, Berger J, Dawid A and Smith A, (Eds.), Bayesian Statistics 4, Oxford: Clarendon Press, pp. 641-649.

Tierney, L. (1994). Markov Chains for exploring posterior distributions. *The Annals of Statistics* **22**(4): 1701–1728.