Bayesian Analysis (2014) **9**, Number 4, pp. 793–804

Comment on Article by Windle and Carvalho

Roberto Casarin *

Abstract. This article discusses Windle and Carvalho's (2014) state-space model for observations and latent variables in the space of positive symmetric matrices. The present discussion focuses on the model specification and on the contribution to the positive-value time series literature. I apply the proposed model to financial data with a view to shedding light on some modeling issues.

Keywords: Exponential Smoothing, Positive-Valued Processes,State-Space Models, Stochastic Volatility.

1 Introduction

The authors are to be congratulated on their excellent intuition, which has culminated in the development of a tractable solution of the filtering and estimation problems for a state-space model on the manifold of symmetric positive-definite matrices. Their Bayesian approach to state-space modeling is in the spirit of the seminal papers of Quintana (1985), Quintana and West (1987) and Quintana et al. (1995) (see also West and Harrison. (1997) ch. 14-16, and Prado and West (2010)). The proposed model provides a pragmatic alternative to some multivariate stochastic volatility models recently proposed in the literature, and the ease and speed of implementation of the inference procedure are appealing in many applied statistics contexts, especially in economics and finance.

Serving as a discussant for the Windle and Carvalho's article gives me the opportunity to welcome the authors' contribution. Such enthusiasm is also motivated by the difficulties that many of us - including myself first - have often been experiencing in Bayesian analysis of multivariate time series. Thus, some parts of this note emphasize the advantages of the proposed model and discuss its possible applications to econometrics. Suggestions will also be given on further references to include and points will be made in relation to unsatisfactory model properties for financial data and to missing investigation of the properties of the multi-step-ahead forecast distribution.

2 An exponential family state-space model

Tractability is the strength of the proposed state-space model, and the authors provide, in the introduction, and in the background section, a discussion on the inferential difficulties that one might encounter in multivariate stochastic volatility modeling. In order to avoid these inferential difficulties, the authors consider a new stochastic volatility model, which is the extended Uhlig's type model in equations (UE). This model is

*Professor of Econometrics, University Ca' Foscari of Venice, Italy r.casarin@unive.it

 DOI:10.1214/14-BA918

simpler in terms of dynamics, and more parsimonious, in terms of number of parameters, than many existing multivariate stochastic volatility models. The proposed UE model allows the authors to exploit the conjugacy property of the Wishart distribution family to find the optimal filter and design a backward sampling procedure for the hidden states (see proof of Propositions 1-3). Section 3 presents the UE model and the main contributions. The results in Propositions 1 and 3 represent a new extension to the matrix-variate context, of the exact filter and prediction formulas for univariate state-space models in the exponential family given in West and Harrison (1997), ch. 14. Previous articles on tractable, matrix-variate exponential family state-space models focused mainly on the Gaussian family of distributions, with the exceptions of Soyer and Tanyeri (2006), Prado and West (2010), Triantafyllopoulos (2012), and Fox and West (2011), whereas the UE model presented by the author is in the Wishart family. The authors focused on the relationship between the UE and the Prado and West (2010) model. I would suggest to complete the literature review by discussing the relationship of the UE model with the Wishart autoregressive models of Triantafyllopoulos (2012) and Fox and West (2011).

2.1 A model for positive-valued observations

The univariate UE model itself represents a contribution to the literature, since it specializes the West and Harrison (1997) results to an exponential family model for positive-valued observations. Previous attempts to provide filtering and prediction rules for exponential family state-space models with gamma distribution for the observations have been carried out by Smith and Miller (1986),Grunwald et al. (1993), Grunwald et al. (1997), and Hyndman et al. (2008), ch. 15. Let y_t be the observable variable, then the univariate UE model (i.e., $m = 1$) can be written as

$$y_t|x_t \sim \mathcal{G}a(k/2, kx_t/2) \quad (1)$$
$$x_t|x_{t-1} \sim x_{t-1}\psi_t/\lambda, \quad \psi_t \sim \mathcal{B}e(n/2, k/2) \quad (2)$$

for $t = 1, \ldots, T$, with $x_0 = x$, where $\mathcal{G}a(a, b)$ denotes the gamma distribution with probability density function

$$f(x) = \Gamma(a)^{-1} b^a \exp\{-bx\} x^{a-1}$$

and $\mathcal{B}e(a, b)$ the beta distribution. By applying propositions 1 and 2, one gets the following filtering and prediction distributions:

$$x_t|\mathcal{D}_t \sim \mathcal{G}a((n + k)/2, k\sigma_t^2/2) \quad (3)$$
$$x_{t+1}|\mathcal{D}_t \sim \mathcal{G}a(n/2, \lambda k\sigma_t^2/2) \quad (4)$$

where $\sigma_t^2 = y_t + \lambda\sigma_{t-1}^2$, and the backward sampling distribution

$$x_t|x_{t+1}, \mathcal{D}_t \sim \lambda x_{t+1} + z_{t+1}, \quad z_{t+1} \sim \mathcal{G}a(k/2, k\sigma_t^2/2). \quad (5)$$

Note that the univariate UE model, defined in (1)-(2), is a special case of the exponential family state-space model in equations (14.1) and (14.5)-(14.7) of West and Harrison.

(1997), ch. 14; that is

$$p(y_t|\eta_t, v_t) = \exp\{v_t^{-1}(y(y_t)\eta_t - a(\eta_t))\}b(y_t, v_t)$$

for the observational distribution, where $\eta_t = 1/x_t$, $a(\eta_t) = k^2 \log(\eta_t)$, $y(y_t) = y_t$, $v_t = -2k$, and $b(y_t, v_t) = y_t^{-v_t/4-1/2}/\Gamma(-v_t/4)(-v_t)^{v_t/4}$, and

$$p(\eta_t|\mathcal{D}_{t-1}) = \exp\{(r_t\eta_t - s_t a(\eta_t))\}c(r_t, s_t)$$

for the state posterior predictive distribution, where $r_t = \lambda k^2 \sigma_{t-1}^2/4$, $s_t = n/2$ and $c(r_t, s_t) = r_t^{s_t+1}/\Gamma(s_t + 1)$.

On one hand, this representation shows that the proposed model is in the spirit of West and Harrison's (1997). On the other hand, it tells us that the authors did not fully exploit the generality of West and Harrison's (1997) approach. More specifically, possible extensions of the UE model include not only the time-varying scale parameter, mentioned by the authors in Section 6, but also the use of exogenous or predetermined variables, time-varying λ, n, and k. Later on in this note, I will discuss the interest in considering time-variations in all these parameters.

2.2 Potential applications

In Sections 5 and 6, the authors mainly focus on the one-step-ahead forecasting of the realized volatility series. The deficient rank case ($k = 1$) is motivated primarily by the application to the squared normal returns and the full rank case by the application to the realized volatility, which is obtained from the aggregation of the intra-day squared returns. Nevertheless, the normality assumption is usually violated in many financial series. Moreover, many financial quantities, such as financial indexes, are weighted sums of series with different features. From my empirical experience (see later on in this note) there is strong evidence of $k > 1$ or $k < 1$ in many squared-return series. Thus, in my opinion, the UE model with unknown k can be used as a model for the squared returns, or for the latent volatility process of the returns.

Moreover, I would not limit the potential use of the UE model to realized volatility only. As said earlier in this note, the UE model provides a tractable state-space model for positive-valued data; therefore, it is well suited for statistical applications to business, industry, and economics, where positive time series (e.g., sale volumes, demand level, insurance claims, weather derivatives, etc.) are very common (see Hyndman et al. (2008), ch. 15). For illustrative purposes, I consider here a database on the historical liquidity conditions in the euro area (i.e. the Eurosystem's supply of and the credit institution's demand for liquidity, publicly available from the ECB (website: `https://www.ecb.europa.eu/mopo/liq/html/index.en.html`). I focus on the daily volumes (in millions of euros) of recourse to the marginal lending facility and on the current accounts, from January 1999 to July 21, 2014. I assume a uniform prior on $[0, 1]$ for λ and a flat prior on $\mathbb{R}^+$ for k and n. The results are given in Figure 1. The exponential smoothing is quite good at describing the dynamics of the two variables.

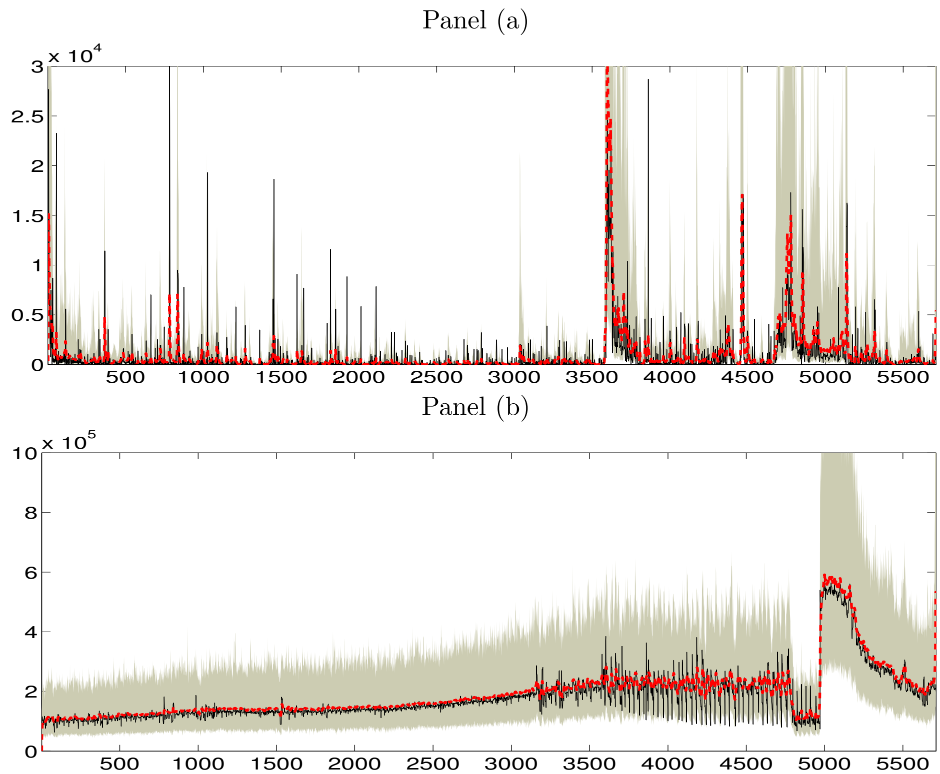

Figure 1: Observable variable y_t, which are daily volumes, in millions of euros, of recourse to the marginal lending facility (black line, panel (a)) and current accounts (black line, panel (b)), from January 1999 to July 21, 2014; y_t's marginal posterior predictive mean (red line in each panel); and the 95% highest posterior density interval (gray area in each panel), given $\mathcal{D}_T = \{y_1, \ldots, y_T\}$.

Shifts and jumps in the level belong to the 95% highest posterior density region of the model.

I find rather surprising that the authors do not mention Bayesian vector autoregressive (BVAR) models with stochastic volatility as a possible application of their UE model. These kinds of applications are in the spirit of Uhlig (1997) and are very useful in economics. The advantages in forecasting with stochastic volatility BVAR models have been confirmed empirically by many parties (e.g., see Clark (2011)). The application to panel BVAR models represents another challenging issue that calls for the use of hierarchical structures, blocking, and sparsity constraints within a Wishart framework.

2.3 Multi-step-ahead forecasting

In economics and business, researchers and practitioners are usually interested in multi-step-ahead forecasting. Thus, having some tractable formulas for forecasting both the latent and observable variables is an advantage of the forecast approach used. Unfortunately, and as stated by the authors in Section 8, the predictive distribution of x_{t+k} given $\mathcal{D}_t$ is available only one-step-ahead, i.e., $k = 1$. Nevertheless, I shall remark that, following Smith and Miller (1986), Section 2, Hyndman et al. (2008), ch. 15, one can focus on the moments of the predictive distribution.

For the univariate UE model, by applying the properties of the harmonic mean for the beta distribution and the inverted gamma distribution, I find that the r-th order moment, $m_{r,h} = \mathbb{E}(x_{t+h}^{-r}|\mathcal{D}_t)$ of the h-step-ahead forecast distribution, $r = 1, 2, 3, \ldots$, $h = 1, 2, 3, \ldots, H$, satisfies the following difference equation

$$m_{r,h} = m_{r,h-1} \frac{\Gamma(n/2 - r)}{\Gamma(n/2)} \frac{\Gamma((n+k)/2 - r)}{\Gamma((n+k)/2)} \lambda^r \tag{6}$$

with the initial condition

$$m_{r,h} = \left(\frac{k\sigma_t^2}{2}\right)^r \frac{\Gamma((n+k)/2 - r)}{\Gamma((n+k)/2)}. \tag{7}$$

Using this set of recursions, one can find the predictive mean and variance

$$\begin{aligned} \mathbb{E}(x_{t+h}^{-1}|\mathcal{D}_t) &= \left(\frac{n+k-2}{n-2}\lambda\right)^h \frac{k\sigma_t^2}{n+k-2} && (8) \\ &= (1-\lambda)\sigma_t^2 \\ \mathbb{V}(x_{t+h}^{-1}|\mathcal{D}_t) &= \sigma_t^4 \lambda^{2h} \left(\left(\frac{(n+k-4)(n+k-2)}{(n-4)(n-2)}\right)^h \frac{k^2}{(n+k-2)(n+k-4)} \right. && (9) \\ &\quad \left. - \left(\frac{n+k-2}{(n-2)}\right)^{2h} \left(\frac{k}{n+k-2}\right)^2 \right) \\ &= \sigma_t^4 \left(\left(\lambda \frac{n+k-4}{n-4}\right)^h \frac{k^2}{(n+k)(n+k-4)} - (1-\lambda)^2 \right) \end{aligned}$$

where the last step in the two equations is obtained by assuming $\lambda = (n-2)/(n+k-2)$. One can see that the h-step-ahead forecast mean is increasing, decreasing, or constant depending on the value of λ. If one assumes the constraint in Equation 3 of the paper is satisfied, then the prediction mean is constant in h. This feature may be appealing in short-term forecasts, but may be not realistic in long-term forecasts where one would expect an increase or decrease in x_t^{-1} (or y_t). See, for example, the second application in Figure 1.

Moreover, in the parametrization chosen by the authors, the constraints $k > m - 1$ and $n > m - 1$ (for $m = 1$) discussed in Section 3 are not sufficient to have well-defined forecasting variances. For this case, one should consider $n > 4$. A discussion

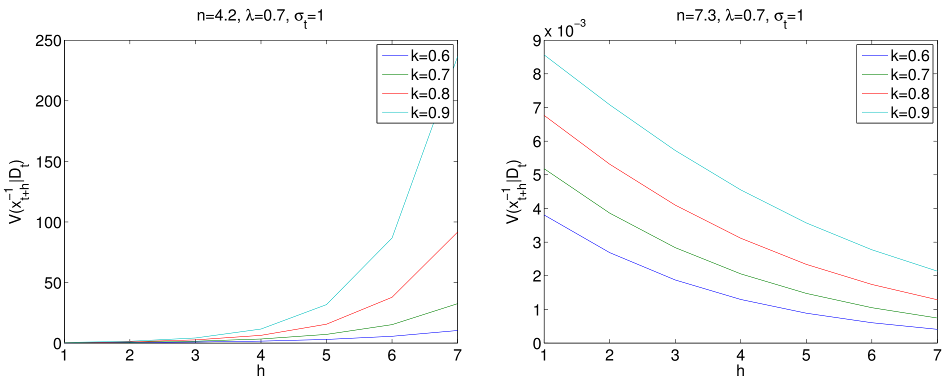

Figure 2: Forecasting variance $\mathbb{V}(x_{t+h}^{-1}|\mathcal{D}_t)$ (vertical axis) at different forecast horizons $h = 1, \ldots, 7$ (horizontal axis) for $\lambda = 0.7$, and for different values of k (different lines) and n (different plots).

on the behavior of the variance as a function of the parameters λ, n and k, would be appreciated. Figure 2 shows the behaviour of the forecast variance for different parameter settings and forecast horizons. I assumed $\lambda = 0.7$, which does not satisfy the constraint in Equation (3) of the paper. The left chart ($n = 4.2$) shows an increasing profile, whereas the right chart ($n = 0.7$) exhibits a decreasing behaviour of the variance, which may be less appealing for applications to economics.

In conclusion, the properties of the predictive distribution generated by the UE model are not explored by the authors. Moreover, following the recent literature by Patton and Timmermann (2011) and in order to have a fair comparison between the UE model and alternative models, such as the stochastic volatility factor models, different forecast horizons, and not only a one-period horizon should be considered.

3 Financial modeling issues

3.1 Heterogeneity in large panel

I will consider the daily log-returns of the financial sector constituents of the Eurostoxx600 index sampled from May 27, 2013 to May 12, 2014 (about one year). This is a panel dataset of 199 series and 250 daily log-returns. The analysis shows evidence of substantial differences in the λ, k, and n across assets (left plot in Figure 3), confirming a strong cross-sectional heterogeneity. I would expect to find heterogeneity also across sectors and countries. Taking into account this feature seems necessary particularly in strategic and global asset allocation contexts. In Section 5 of their paper, the authors conclude that the model in Section 4 is equivalent to the restricted UE model with $R = \sqrt{\lambda} I_m$. I suspect that the validity of the result is limited to the empirical applica-

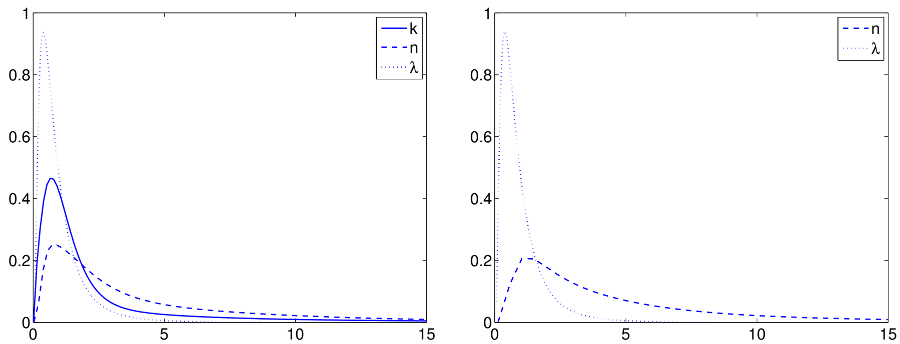

Figure 3: Cross-section estimates (smoothed histograms) for the unconstrained rank ($k > 0$, left) and the deficient rank ($k = 1$, right) UE model.

tion considered in their paper. The performance of the UE model on large dimensions (hundreds or thousands of assets) has not yet been explored, and more realistic modeling strategies, such as blocking or hierarchical modeling, are needed in order to have interesting models for large portfolios.

3.2 Structural breaks

I consider S&P500 daily percent log-returns data from January 3, 1972 to September 9, 2013, an updated version of the dataset used in Geweke and Amisano (2010, 2011) and Fawcett et al. (2014). In relation to this dataset it has been recognized that in some sub-samples heavy-tail GARCH models have a better predictive ability than Gaussian GARCH models. I focus on the period of the great financial crisis and consider data from June 12, 2008 to September 9, 2013, for a total of 1260 observations. I estimate the univariate UE using rolling samples of 60 observations (about three months) and produce Bayesian estimates for the unknown parameters k, n, and λ. The results show that many of the spikes (or jumps) in the squared return are out of the 95% highest posterior density region (gray area in Figure 4, which may suggest that for some return series with large kurtosis the model should be extended to account for heavy tails. The dotted blue line in Figure 4 shows the rolling estimates of the degrees of freedom parameter of the Student-t GARCH(1,1) model on the return series, with a window size of 90 observations. This provides evidence of heavy tails in the conditional distribution of the observations and may motivate the extension proposed by the authors in Section 6.

The results in Figure 5 support a strong evidence of breaks in the parameters, which calls for the use of time-varying parameter models. Using constant n or λ leads to over- and under-estimation of the volatility. For certain assets, the model should allow for time-varying λ, k, and n. Time variations may be due to changes in the tail behaviour of the return distribution. I shall remark that the pattern of the degrees of freedom

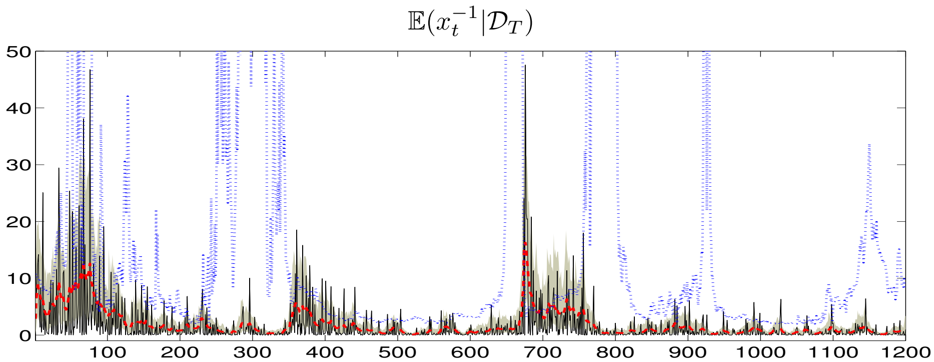

Figure 4: Daily squared returns (solid black line) on S&P500 index from January 12, 2008 to September 9, 2013, volatility posterior mean (dashed red line) and the 95% highest posterior density region (gray area), given $\mathcal{D}_T = \{y_1, \ldots, y_T\}$. Rolling estimates of the degrees of freedom parameter of the Student-t GARCH(1,1) model (dotted blue line), with a window size of 90 observations, on the original return series.

estimates in Figure 4 is similar to the one of the k and n rolling estimates in Figure 5.

From my empirical analysis, I shall conclude that the parameter dynamics exhibits persistence and is subjected to rapid changes. These facts call for the use of Markov-switching (MS) models, such as the following MS-UE model

$$Y_t|X_t, z_t \sim \mathcal{W}_m\left(k_t, (k_t X_t)^{-1}\right) \tag{10}$$

$$X_t|X_{t-1}, z_t \sim T_{t-1}^{-1\prime}\Psi_t/T_{t-1}^{-1}\lambda_t, \quad \Psi_t|z_t \sim \beta_m(n_t/2, k_t/2) \tag{11}$$

where k_t, n_t and λ_t are driven by a MS process, z_t, $t = 1, \ldots, T$, i.e. $k_t = \sum_{l=1}^{M} \bar{k}_j \mathbb{I}(z_t = j)$, $k_t = \sum_{l=1}^{M} \bar{n}_j \mathbb{I}(z_t = j)$ and $\lambda_t = \sum_{l=1}^{M} \bar{\lambda}_j \mathbb{I}(z_t = j)$, where $P(z_t = j|z_{t-1} = i) = p_{ij}$ and $\bar{k}_j$, $\bar{n}_j$, and $\bar{\lambda}_j$ are regime-specific parameters.

In conclusion, for many financial series, there is a strong empirical evidence of time-variations and cross-section heterogeneity in the UE model parameters, which can undermine the goodness of fit and the forecasting abilities of the proposed model. Tractable extensions of the model, particularly in large dimension, would be very useful.

4 Conclusion

In their paper, the authors sketch a number of possible extensions. I would suggest as further research lines the inclusion of graphical structures in the UE model, and/or sparsity constraints, following Carvalho and West (2007) and Wang and West (2009). It is clear that this is an exciting and stimulating work. I am therefore very pleased to be able to propose the vote of thanks to the authors for their work.

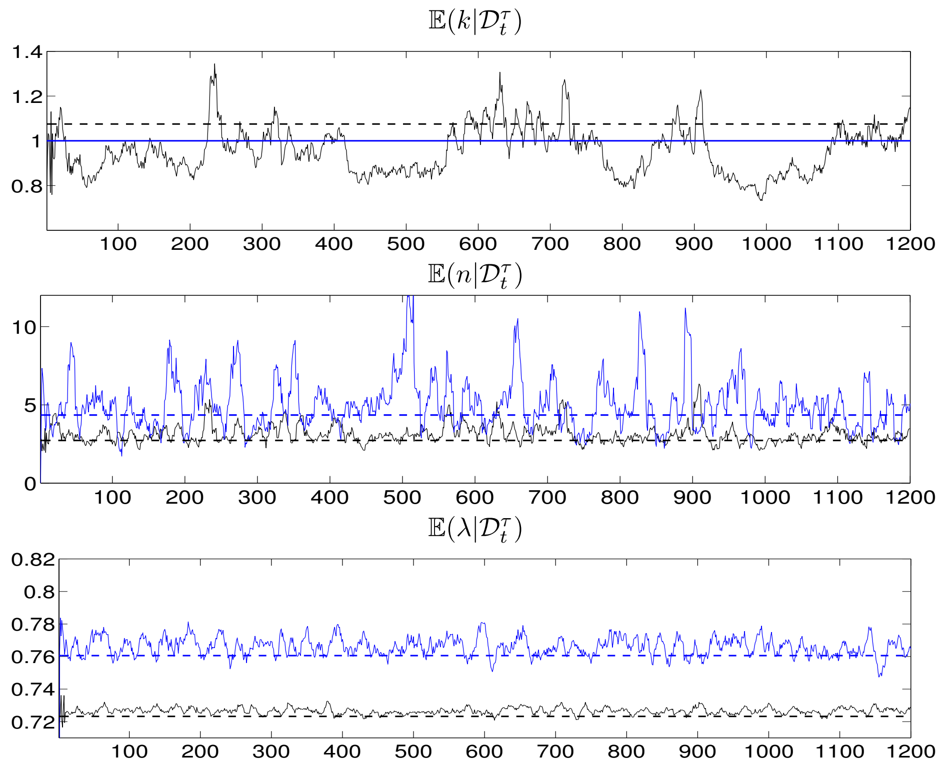

Figure 5: k, n and λ rolling posterior mean (solid lines) given the observation set $\mathcal{D}_t^\tau = \{y_{t-\tau+1}, \ldots, y_t\}$ for the unconstrained (black line) and constrained (blue line) UE model. Horizontal dashed lines represent the parameter posterior means given $\mathcal{D}_T$.

Acknowledgement

Authors' research is supported by funding from the European Union, Seventh Framework Programme FP7/2007-2013 under grant agreement SYRTO-SSH-2012-320270, by the Institut Europlace of Finance, "Systemic Risk grant", the Global Risk Institute in Financial Services, the Louis Bachelier Institute, "Systemic Risk Research Initiative", and by the Italian Ministry of Education, University and Research (MIUR) PRIN 2010-11 grant MISURA.

References

Carvalho, C. M. and West, M. (2007). "Dynamic Matrix-Variate Graphical Models." *Bayesian Analysis*, 2: 69–98. 800

Clark, T. E. (2011). "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility." *Journal of Business & Economic Statistics*, 29(3): 327–341. 796

Fawcett, N., Kapetanios, G., Mitchell, J., and Price, S. (2014). "Generalised density forecast combinations." Technical report, Bank of England Working Paper 492, Bank of England. 799

Fox, E. B. and West, M. (2011). "Autoregressive models for variance matrices: stationary inverse Wishart processes." Technical report, ArXiv: http://arxiv.org/abs/1101.2017. 794

Geweke, J. and Amisano, G. (2010). "Comparing and evaluating Bayesian predictive distributions of asset returns." *International Journal of Forecasting*, 26: 216–230. 799

— (2011). "Optimal prediction pools." *Journal of Econometrics*, 164: 130–141. 799

Grunwald, G. K., Guttorp, P., and Raftery, A. E. (1993). "Prediction rules for exponential family state space models." *Journal of the Royal Statistical Society, Series B*, 55: 937–943. 794

Grunwald, G. K., Hamza, K., and Hyndman, R. J. (1997). "Some properties and generalizations of non-negative Bayesian time series models." *Journal of the Royal Statistical Society, Series B*, 55: 937–943. 794

Hyndman, R. J., Koehler, A. B., Ord, J., and Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer-Verlag, Berlin. 794, 795, 797

Patton, A. J. and Timmermann, A. (2011). "Forecast Rationality Tests Based on Multi-Horizon Bounds." *Journal of Business & Economic Statistics*, 30(1): 1–17. 798

Prado, R. and West, M. (2010). *Time Series: Modeling, Computation, and Inference*, chapter Multivariate DLMs and Covariance Models, 263–319. Chapman & Hall/CRC. 793

Quintana, J. M. (1985). "A dynamic linear matrix-variate regression model." Technical report, Research Report 83, Department of Statistics, University of Warwick. 793

Quintana, J. M., Chopra, V. K., and Putnam, B. H. (1995). "Global asset allocation: Stretching returns by shrinking forecasts." In *Proceedings of the ASA Section on Bayesian Statistical Science, 1995 Joint Statistical Meetings, American Statistical Association.*. 793

Quintana, J. M. and West, M. (1987). "Multivariate time series analysis: new techniques applied to international exchange rate data." *The Statistician*, 36: 275–281. 793

Smith, R. L. and Miller, J. E. (1986). "A non-Gaussian state space model and application to prediction of records." *Journal of the Royal Statistical Society, Series B*, 48(79-88). 794, 797

Soyer, R. and Tanyeri, K. (2006). "Bayesian portfolio selection with multi-variate random variance models." *European Journal of Operational Research*, 171: 977–90. 794

Triantafyllopoulos, K. (2012). "Multi-variate stochastic volatility modelling using Wishart autoregressive processes." *Journal of Time Series Analysis*, 33: 48–60. 794

Uhlig, H. (1997). "Bayesian vector autoregressions with stochastic volatility." *Econometrica*, 65(1): 59–73. 796

Wang, H. and West, M. (2009). "Bayesian analysis of matrix normal graphical models." *Biometrika*, 96(4): 821–834. 800

West, M. and Harrison., J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer Verlag. 793, 794

Windle, J. and Carvalho, C. (2014). "A tractable state-space model for symmetric positive-definite matrices." *Bayesian Analysis*, forthcoming. 793