Bayesian VEC models with Markov-switching heteroscedasticity in forecasting macroeconomic time series

Łukasz Kwiatkowski[1]

Abstract

In the paper, we examine the forecasting abilities of Bayesian vector error correction models (allowing for long-term relationships between modelled variables) featuring Markovian breaks in the conditional covariance matrix so as to capture time-varying volatility typically recognized in macroeconomic data. Such models may prove a useful tool in prediction of macroeconomic time series as a valid, empirically 'sufficient' alternative for VEC structures with more sophisticated specifications of heteroscedasticity, such as GARCH or stochastic volatility processes. The predictive performance of the models in question is evaluated here within the probabilistic paradigm of forecasting, with the accuracy of density forecasts measured by means of the log predictive score and Bayes factors, while also using Probability Integral Transform (PIT) to assess their calibration. Two empirical studies conducted in this research for the Polish and US economies, in the context of small models of monetary policy, indicate some gains in the predictive power of VEC models with Markov-switching heteroscedasticity as compared with homoscedastic VEC systems, though more sophisticated specifications (like GARCH or stochastic volatility) may still be required for further improvement.

Keywords: cointegration, probabilistic forecasting, predictive score, predictive Bayes factor

JEL Classification: C11, C32, C53

1. Introduction

For an effective forecasting of any time series by means of some statistical model it is essential for the latter to capture key characteristics of the data at hand. Macroeconomic time series, which are of this paper's main focus, typically display two features (potentially, along with some seasonal or cyclical patterns): non-stationarity (due to the presence of stochastic trends) and conditional heteroscedasticity, with the latter having already been commonly associated not only with financial and commodity markets. Dealing with non-stationary processes jointly for different variables usually requires the use of cointegration analysis, with the underlying vector autoregression (VAR) model in its vector error correction (VEC) form. Then, to account also for the other feature some time-variability needs to be introduced into the conditional covariance matrix of the observations, with typical choices including a variety of multivariate

[1] Cracow University of Economics, Rakowicka 27, 31-510 Kraków, Poland, lukasz.kwiatkowski@uek.krakow.pl.

GARCH (MGARCH) or stochastic volatility (MSV) processes, both classes enabling continuously-valued (rather than discrete) changes of conditional variances and/or correlations. Recently, Wróblewska and Pajor (2019) examined the predictive performance of VEC models equipped with hybrid structures combining Multiplicative Stochastic Factor (MSF) process (belonging to the MSV class) and Scalar BEKK (SBEKK) specification (of the MGARCH family). As evidenced in the cited work (focusing on macroeconomic data for the Polish economy), extending a 'standard', homoscedastic VEC model with the MSF or MSF-SBEKK conditional volatility structure (introduced by Osiewalski, 2009, Osiewalski and Pajor, 2009) dramatically improves the forecasting abilities of the model (as measured by the log predictive score, LPS, energy score, ES, and also mean squared forecast error, MSFE).

In this paper, we shift the attention to a qualitatively different and simpler (as compared with MGARCH, MSV or their hybrids) approach to modelling conditional heteroscedasticity in cointegrated VAR/VEC systems. We conjecture that in the case of macroeconomic (as opposed to financial) time series it may be empirically 'sufficient' (for the purpose of prediction) to enable discrete rather than continuously-valued shifts in the multivariate volatility. Following this line of reasoning, we allow the conditional covariance matrix to switch between two regimes, say, of high and low volatility, according to a homogenous and ergodic Markov chain.

Although the concept of Markov-switching (MS) time series models has been well-established and present in the literature for a long time (since a seminal paper by Hamilton 1989), papers devoted to forecast evaluation of MS-VEC models for macroeconomic data are not as much scarce as their focus limited to the point (rather than density) prediction almost exclusively (see, e.g. Clarida *et al.*, 2003, Sarno *et al.*, 2005, and Psaradakis and Spagnolo, 2005; in the latter the authors, apart from the point forecasts, also evaluate the calibration of density forecasts *via* the Probability Integral Transform, PIT). Therefore, following the recent shift of the forecasting paradigm from the point to probabilistic prediction, the aim of this study is an empirical evaluation of predictive densities performance of VEC models with Markov-switching heteroscedasticity (VEC-MSH) in comparison with homoscedastic VEC structures.

As for the statistical inference framework, we resort to the Bayesian approach to estimation and prediction, similarly as Wróblewska and Pajor (2019). Admittedly, the Bayesian setting is the most suitable while dealing with models with latent processes (like stochastic volatility or hidden Markov chains). Moreover, it handles coherently the parameter estimation uncertainty while producing predictive densities, which in the case of regime-changing models may be essential for the sake of their forecasting performance (as conjectured by Psaradakis and

Spagnolo, 2005). The latter is evaluated here by means of the log predictive score (LPS), which underlies the so-called predictive Bayes factor. We also examine PIT histograms to assess forecasts' densities calibration; see, e.g., Geweke and Amisano (2010), Gneiting and Raftery (2007), Gneiting *et al.* (2007).

2. VEC models with Markov-switching heteroscedasticity

An n-variate VAR(k) model with Markov-switching conditional covariance matrix can be written in its VEC (henceforth VEC-MSH) form as:

$$\Delta x_t = \widetilde{\Pi} x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi D_t + \varepsilon_t, t = 1, 2, \dots, T, \tag{1}$$

$$\varepsilon_t | \psi_{t-1}, S_t, \theta \sim N(0, \Sigma_t), \tag{2}$$

where x_t is an n-variate random vector, $\{\varepsilon_t\}$ is a vector white noise with some unconditional covariance matrix Σ, matrix D_t comprises deterministic variables (such as the constant, trend and seasonal dummies), $\widetilde{\Pi}$, Γ_i and Φ are real-valued matrices of parameters, all collected in θ, and ψ_{t-1} denotes the past of the process $\{x_t\}$ up to time $t-1$. The matrix $\widetilde{\Pi}$ decomposes as $\widetilde{\Pi} = \alpha\tilde{\beta}'$, with α $(n \times r)$ storing the adjustment coefficients and $\tilde{\beta}$ $(m \times r,\, m \geq n)$ pertaining to the cointegration relationships, once they exist (then $r < n$ is their number). Note that $m > n$ only under the deterministic components restricted to the cointegration relationships. The initial conditions x_{-k+1}, x_{-k+2}, …, x_0 are assumed to be known and set as pre-sample observations; see Wróblewska and Pajor (2019) and the references therein. Finally, $\{S_t\}$, where $S_t \in \{1, 2\}$, forms a two-state homogenous and ergodic Markov chain with the (time-invariant) transition probabilities $p_{ij} \equiv \Pr(S_t = j | S_{t-1} = i, \theta)$, $(i = 1, 2)$, with p_{11} and p_{22} also stored in θ. This latent process governs the switches between the two regimes, each featured by 'its own' conditional (given ψ_{t-1} and θ) covariance matrix $\Sigma_t \equiv \Sigma_{S_t}$ of the error term ε_t. Note that in our setting we restrict the regime changes only to the volatility, thereby restricting the other parameters of the VEC structure to be time-invariant, assuming that possible long-term relationships and short-term adjustments hold constant over the entire sample.

3. Bayesian model specification, estimation and prediction

The methodology of Bayesian Markov-switching VEC models has been developed by Jochmann and Koop (2015) and we follow their approach to a large extent, with minor modifications so as to tailor our framework to the one presented in Wróblewska and Pajor (2019).[2] As Bayesian modelling requires specification of the prior distributions for model parameters, we adopt their structure for the VEC part from the cited paper, while also assuming that Σ_1 and Σ_2 follow the inverse Wishart distribution – the same as the one considered in Wróblewska and Pajor (2019) for Σ in homoscedastic VECs. For p_{11} and p_{22} in all the VEC-MSH models we impose uniform priors.

Bayesian estimation of the models at hand necessitates a use of MCMC methods, including the Gibbs sampler (in all the models) and the Forward-Filtering-Backward-Sampling scheme (developed by Carter and Kohn, 1994) for sampling latent Markov chain's state variables; see also Jochmann and Koop (2015). Additionally, to handle the label switching, a problem inherent to mixture models, we use the permutation sampler designed by Frühwirth-Schnatter (2006). Although requiring additional simulations at each MCMC step, prediction within the VEC and VEC-MSH models is quite straightforward, owing to a sequential structure of the likelihood.

4. Empirical analysis

The empirical analysis to follow is based on various VEC and VEC-MSH specifications of the so-called small models of monetary policy, and considered separately for the Polish and US economies. Following Primiceri (2005) and Wróblewska and Pajor (2019), the models comprise three variables: inflation rate of consumer prices, unemployment rate and short-term interest rate (WIBOR3M for Poland, and the so-called shadow interest rates, calculated according to Wu and Xia 2015, for US).[3] In both cases we use quarterly data, covering the periods: 1998:Q1–2016:Q4 ($T = 76$ observations, seasonally unadjusted) for Poland, and 1960:Q1–2015:Q4 ($T = 224$ observations, seasonally adjusted) for the US; see Fig. 1. To account for seasonal effects in the data for Poland, zero-mean centered seasonal dummies are introduced in the models (following Wróblewska and Pajor 2019). The dataset for Poland

[2] Details can be provided by the author upon request.

[3] However oversimplified such models may appear from the macroeconomic perspective, we refrain from an otherwise due discussion on the limitations and relevant extensions of their structure (for the open economy of Poland, in particular). The choice of the models at hand is primarily dictated by our intent to extend the research and contribute to the discussion by Wróblewska and Pajor (2019) on predictive performance of heteroscedastic VEC models.

coincides precisely with the one analysed in Wróblewska and Pajor (2019), which enables us to compare the results of both studies. Including here also the US economy is intended to broaden the scope of the analysis, and relates to an influential paper by Primiceri (2005).

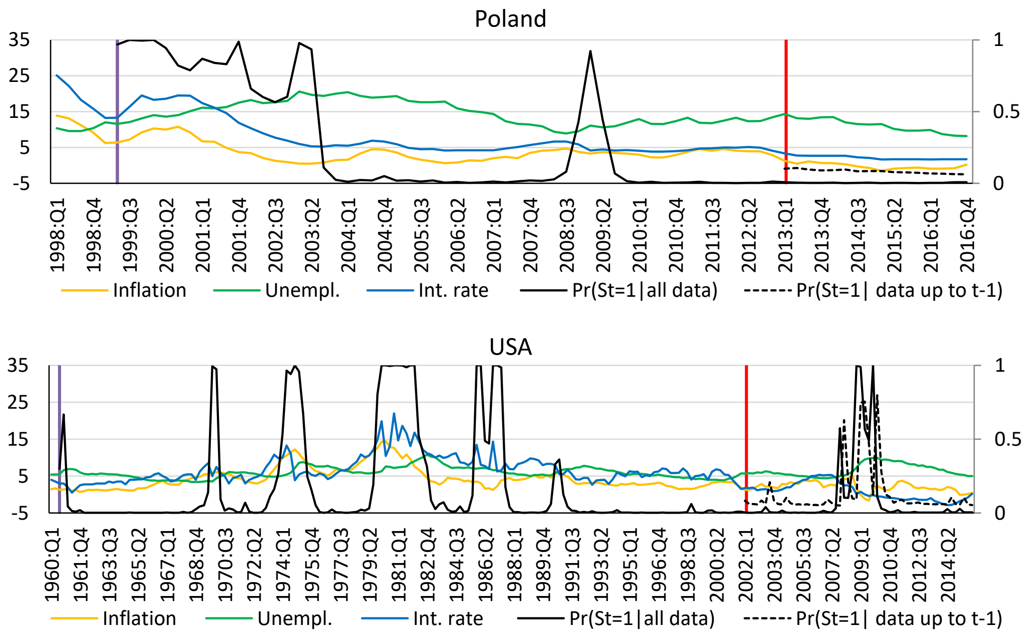


Figure 1. Modelled data (values on the LHS axis) along with the posterior and predictive probabilities of the first state (the RHS axis)

Note: Vertical lines demark the initial conditions (violet line) and the period of predictive performance evaluation.

In each of the models under study we assume that the order of the underlying VAR process equals $k = 2$ (see Primiceri, 2005, Wróblewska and Pajor, 2019). We consider two alternative specifications of the constant term in Eq. (1) (either an unrestricted constant, conventionally denoted as $d = 3$, or a constant restricted to the cointegration relationships: $d = 4$), and three different numbers of cointegration relations ($r \in \{0, 1, 2\}$), and the case of a stationary VAR system (i.e. $r = n = 3$). Models with given d and r are labelled as VEC(d, r) and VEC(d, r)-MSH. Some of all possible specifications are omitted from further analysis, either due to some numerical problems encountered in estimation (VEC(3, 2)-MSH, VEC(3, 3)-MSH for both the Polish and US datasets, and VEC(4, 1)-MSH and VEC(4, 2)-MSH for US), or due to their absence in Wróblewska and Pajor (2019) (VEC(4, 0) for Poland) or their methodological irrelevance (VEC(4, 3 = n), with and without the MSH structure). Eventually, in the US case only three VEC-MSH are considered (see Table 2). For the Polish data we additionally include two specifications of the VEC-MSF-SBEKK family – the ones that in Wróblewska and Pajor (2019) emerged the best (full VEC-MSF-SBEKK) and the worst (VEC-SBEKK). To address the problem of label switching in the VEC-MSH models we impose an identification restriction: $Var(Int.rate_t | S_t = 1, \psi_{t-1}, \theta) > Var(Int.rate_t | S_t = 2, \psi_{t-1}, \theta)$, so that the first regime features a higher volatility of the interest rates, although we note that the results are ‘robust’ to the choice of the underlying variable. Moreover, estimation of the Markov-switching models with $r \in \{1, 2\}$ for Poland required a strong tightening of the priors of the adjustment coefficients and the elements of the matrix B (see Wróblewska and Pajor, 2019), with their standard deviations reduced from 1 in all the other models to ca. 0.032, although the latter modification does not affect the prior for the cointegration space.

The predictive performance (in the sense of density forecasts) of the models under consideration is evaluated *via* series of *ex-post* one-quarter-ahead density predictions, based on a sequence of expanding (recursive) samples, with each model being reestimated upon the arrival of each new observation. For the sake of the experiment we spare the final $N = 16$ and $N = 56$ observations in the case of Poland and US, respectively, so that the forecasting periods cover 2013:Q1–2016:Q4 (Poland), and 2002:Q1–2015:Q4 (US). As can be inferred from Fig. 1, throughout the entire prediction period for Poland the second regime (of low volatility) prevails unequivocally, as opposed to the US data, where the corresponding period witnesses some regime changes. Therefore, it is even more interesting to examine and compare the predictive abilities of Markov-switching models in these two markedly distinct settings.

Each of the predictive densities is based upon 200 000 MCMC posterior draws, preceded by either 400 000 burn-in passes – for the first of N forecasts – or 10 000 cycles for the subsequent $N - 1$ predictions, with the sampler each time initiated at the final draw of the previous run. Density forecasts are evaluated by means of the (decimal) log predictive score (LPS; the higher the value, the better), with the difference between LPS’s for two alternative models defining the log predictive Bayes factor (LPBF). Their values cumulated over the entire *ex-post* forecasting period are denoted as CLPS and CLPBF, respectively, and presented in Tables 1 and 2.

As can be inferred from Tables 1 and 2, the Markov-switching models prove at least marginally (in the case of Poland) or substantially better (US) in terms of LPS. Although for the Polish dataset (displaying apparent ‘tranquillity’ throughout the prediction period) the difference in the predictive power of VEC and VEC-MSH models is negligible, it is still systematic, holding for all variants of d and r, which is most probably attributable to an overall higher flexibility of mixture models. Nevertheless, specifications featuring only discrete (rather than continuously-valued) volatility changes are still outperformed by the VEC-MSF-SBEKK and even VEC-SBEKK models. For the US data, the VEC-MSH specifications gain far more advantage over homoscedastic models. As indicated by Fig. 2, this superiority of Markov-switching models hinges directly upon evident occurrences of volatility breaks over the prediction period, which is quite intuitive.

Table 1. Cumulated log predictive scores (CLPS) and cumulated log predictive Bayes factors (CLPBF) in favour of the best model: Poland. Results for VEC-MSF-SBEKK and VEC-SBEKK come from Wróblewska and Pajor (2019)

Ranking (*i*)	*d*	*r*	Model	$CLPS_i$	$CLPBF_{1i}$
1	4	2	VEC-MSF-SBEKK	-5.471	0
2	3	1	VEC-SBEKK	-11.287	5.816
3	4	1	VEC-MSH	-12.314	6.843
4	4	2	VEC-MSH	-12.321	6.850
5	3	0	VEC-MSH	-12.442	6.971
6	4	0	VEC-MSH	-12.452	6.981
7	3	1	VEC-MSH	-12.491	7.020
8	4	2	VEC	-12.716	7.245
9	4	1	VEC	-12.797	7.326
10	3	2	VEC	-13.048	7.577
11	3	1	VEC	-13.093	7.622
12	3	0	VEC	-13.377	7.906
13	3	3	VEC	-13.891	8.420

Table 2. Cumulated log predictive scores (CLPS) and cumulated log predictive Bayes factors (CLPBF) in favour of the best model: US

Ranking (*i*)	*d*	*r*	Model	$CLPS_i$	$CLPBF_{1i}$
1	3	1	VEC-MSH	-71.590	0
2	3	0	VEC-MSH	-71.997	0.407
3	4	0	VEC-MSH	-72.066	0.476
4	4	2	VEC	-82.768	11.178
5	4	1	VEC	-82.875	11.285
6	3	2	VEC	-82.887	11.297
7	3	1	VEC	-82.974	11.384
8	4	0	VEC	-83.067	11.477
9	3	0	VEC	-83.180	11.590
10	3	3	VEC	-83.569	11.979

Finally, we examine the calibration of density forecasts *via* PIT histograms, displayed in Fig. 3 and 4, along with 95% (non-Bayesian) confidence bands constructed as in Wróblewska and Pajor (2019) around the value of 0.2 (representing, in our setting, the ideal case of PIT uniformity). Overall, although none of the models considered in these figures proves ideal, it appears the introducing Markovian breaks in the conditional covariance matrix of VEC models may somewhat improve (more or less visibly) predictive densities calibration.

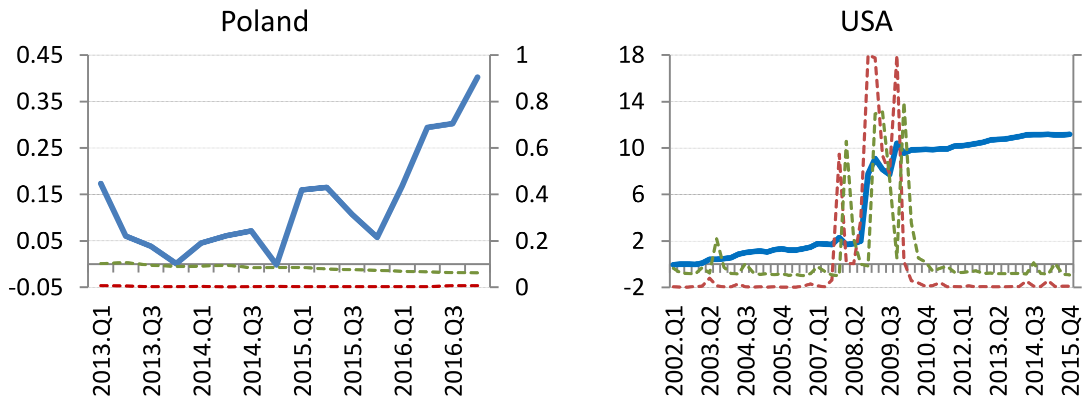


Figure 2. Cumulative log predictive Bayes factors in favour of the best VEC-MSH against the best VEC model (solid line; the LHS axis), along with the posterior (red dashed line) and predictive (green dashed line) probabilities of the first state (the RHS axis)

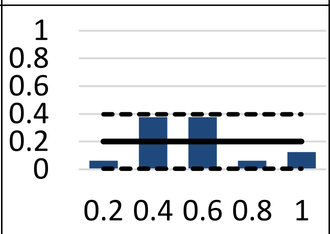
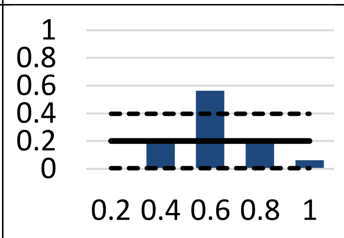
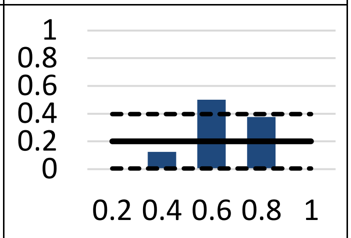
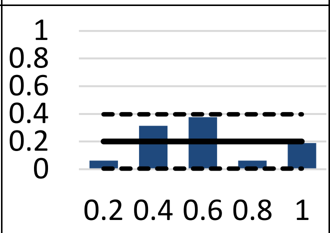
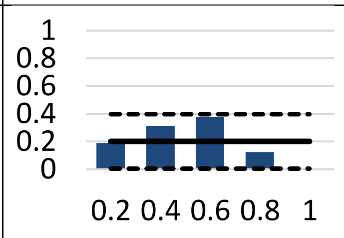
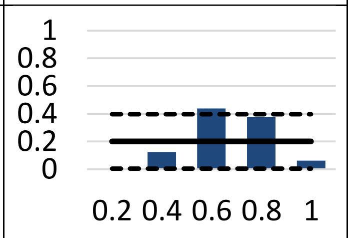
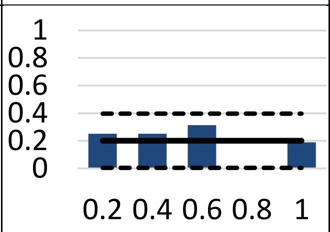
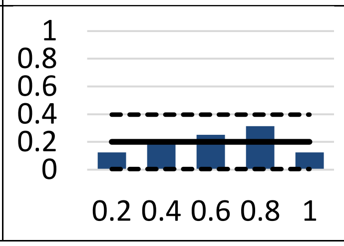
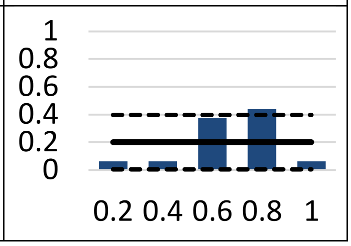
Model	Inflation	Unemployment	Interest rate
VEC(4,2)			
VEC(4,1)-MSH			
VEC(4,2)-MSF-SBEKK			

Figure 3. PIT histograms in the best VEC, VEC-MSH and VEC-MSF-SBEKK models for Poland

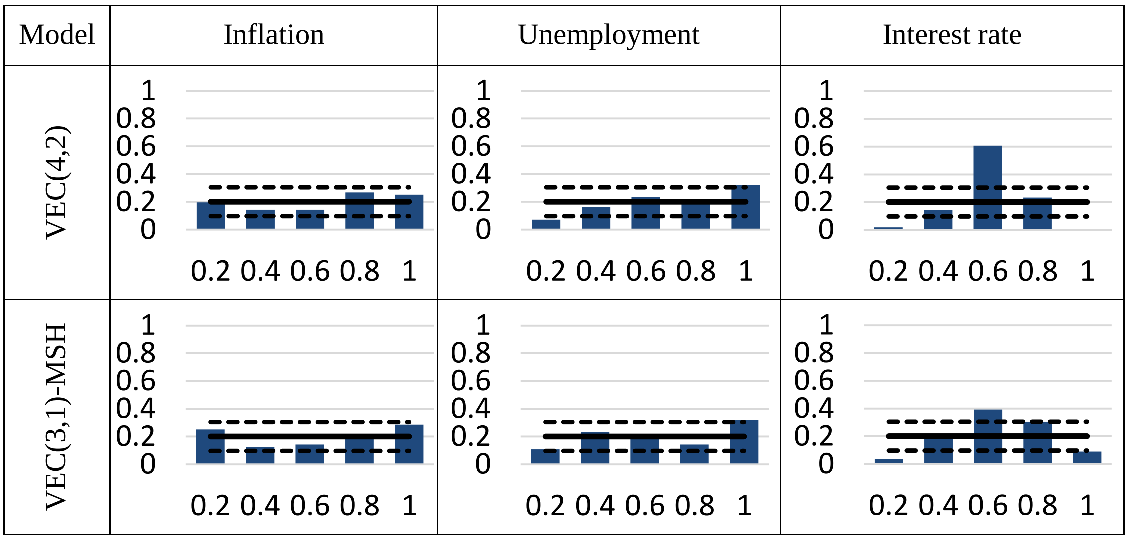


Figure 4. PIT histograms in the best VEC and VEC-MSH models for US

5. Conclusions

In the paper we examined and compared probabilistic predictive performance of Bayesian homoscedastic VEC models with their extensions allowing for two-state Markov-switching heteroscedasticity. To this end, the log predictive score (LPS) and Bayes factors, as well as Probability Integral Transform were employed, which are typically used for such assessments.

In general, the results of our two empirical studies based on data representing, separately, the Polish and US economies indicate that allowing for Markovian shifts in conditional covariance matrix of VEC models provide at least as good density forecasts as the ones obtained within 'standard' VEC structures, with the VEC-MSH outperforming the latter in the presence of volatility shifts occurring over the prediction period. Nevertheless, the comparison with the results presented by Wróblewska and Pajor (2019) implies that enabling Markovian dynamics in conditional volatility may still prove empirically insufficient, thereby necessitating a use of more sophisticated specifications like MSF-SBEKK.

Acknowledgements

This research is supported through a grant from the National Science Center (NCN, Poland) under decision no. UMO-2018/31/B/HS4/00730. The author is greatly indebted to Justyna Wróblewska for her support and help with developing relevant codes in GAUSS.

References

Carter, C.K., Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, *81*(3), 541-553.

Clarida, R.H., Sarno, L., Taylor, M.P., Valente, G. (2003). The out-of-sample success of term structure models as exchange rate predictors: a step beyond. *Journal of International Economics*, *60*, 61-83.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Series in Statistics, Springer.

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

Gneiting, T., Balabdaoui, F., Raftery, A.E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*, 243-268.

Gneiting, T., Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*, 359-378.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*, 357-384.

Jochmann, M., Koop, G. (2015). Regime-switching cointegration. *Studies in Nonlinear Dynamics & Econometrics*, *19*(1), 35-48.

Osiewalski, J. (2009). New hybrid models of multivariate volatility (a Bayesian perspective). *Przegląd Statystyczny* (*Statistical Review*), *56*(1), 15-22.

Osiewalski, J., Pajor, A. (2009). Bayesian analysis for hybrid MSF–SBEKK models of multivariate volatility. *Central European Journal of Economic Modelling and Econometrics*, *1*(2). 179-202.

Primiceri, G.E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, *72*, 821-852.

Psaradakis, Z., Spagnolo, F. (2005). Forecast performance of nonlinear error-correction models with multiple regimes. *Journal of Forecasting*, *24*, 119-138.

Sarno, L., Thornton, D.L., Valente, G. (2005). Federal Funds Rate Prediction. *Journal of Money, Credit, and Banking*, *37*(3), 449-471.

Wróblewska, J., Pajor, A. (2019). One-period joint forecasts of Polish inflation, unemployment and interest rate using Bayesian VEC-MSF models. *Central European Journal of Economic Modelling and Econometrics*, *11*, 23-45.

Wu, C.J., Xia, F.D. (2015). Measuring the macroeconomic impact of monetary policy at the zero lower bound. Chicago Booth Research Paper No. 13-77. Available at: http://dx.doi.org/10.2139/ssrn.2321323.