Central European Journal of Economic Modelling and Econometrics

A Long-Run Relationship between Daily Prices on Two Markets: The Bayesian VAR(2)–MSF-SBEKK Model

Krzysztof Osiewalski*, Jacek Osiewalski†

Submitted: 9.5.2013, Accepted: 5.7.2013

Abstract

We develop a fully Bayesian framework for analysis and comparison of two competing approaches to modelling daily prices on different markets. The first approach, prevailing in financial econometrics, amounts to assuming that logarithms of prices behave like a multivariate random walk; this approach describes logarithmic returns most often by the VAR(1) model with MGARCH (or sometimes MSV) disturbances. In the second approach, considered here, it is assumed that daily price levels are linked together and, thus, the error correction term is added to the usual VAR(1)–MGARCH or VAR(1)–MSV model for logarithmic returns, leading to a reduced rank VAR(2) specification for logarithms of prices. The model proposed in the paper uses a hybrid MSV-MGARCH structure for VAR(2) disturbances. In order to keep cointegration modelling as simple as possible, we restrict to the case of two prices representing two different markets.

The aim of the paper is to show how to check if a long-run relationship between daily prices exists and whether taking it into account influences our inference on volatility and short-run relations between returns on different markets. In the empirical example the daily values of the S&P500 index and the WTI oil price in the period 19.12.2005 – 30.09.2011 are jointly modelled. It is shown that, although the logarithms of the values of S&P500 and WTI oil price seem to be cointegrated, neglecting the error correction term leads to practically the same conclusions on volatility and conditional correlation as keeping it in the model.

Keywords: Bayesian econometrics, vector error correction model, hybrid MGARCH-MSV processes, financial markets, commodity markets

JEL Classification: C11, C32, C51, C58

*e-mail: krzysztof@osiewalski.eu

†Cracow University of Economics, e-mail: eeosiewa@cyf-kr.edu.pl

1 Introduction

In financial econometrics it is usually assumed that vector autoregression (VAR) is the basic model structure for logarithmic return rates with the disturbances following one of multivariate volatility processes. This approach amounts to describing logarithms of prices by (some generalization of) a multivariate random walk, so that price levels are not linked. In order to capture possible long-run relationships among price levels, we consider cointegration of logarithms of prices and introduce the error correction mechanism (ECM) into the multivariate model for return rates. Long-term relationships have already been introduced into volatility models by e.g. Chang C., Lai J. and Chuang I. (2010), Ji Q. and Fan Y. (2011), Mahadevan R. and Suardi S. (2011). Before that, both approaches were usually separated, although applied to daily data, e.g. Bekiros S. and Diks C. (2008) filtered the data by GARCH models and then built the ECM specification for the residuals.

It is worth noting that much earlier Osiewalski J. and Pipień M. (2004) included the ECM term in the conditional mean specification of bivariate VAR–MGARCH models with competing MGARCH structures. However, their approach was completely different from the one considered here as they used an obvious cointegration relation without testing it. Also, their long-run relationship involved one more variable, which was assumed exogenous and thus not modelled. Here we are interested in estimating, testing and exploiting potential long-run relations linking these variables, which are jointly modelled and forecasted - without any extra (external) variables, assumed to be exogenous. Pajor A. (2011), illustrating her theory of exogeneity in Bayesian VECM–MSV models, uses the same obvious cointegration relation as did Osiewalski J. and Pipień M. (2004), but in the context of a full trivariate specification. Again, the long-run relationship was assumed and not formally derived within the Bayesian model. Also, the main comparison was between trivariate and bivariate VECM–MSV models with the same ECM term, and not between bivariate models with or without ECM.

In this work we consider a parsimonious model structure for joint analysis of long- and short-run relationships among prices on different markets. In order to keep cointegration modelling as simple as possible, we restrict here to the bivariate case - with two prices representing two markets. To model the conditional covariance matrix we use a simple hybrid MSV-MGARCH structure, namely the MSF-SBEKK one, proposed by Osiewalski J. (2009) and Osiewalski J. and Pajor A. (2009). It shows both parsimony and strengths of each basic class of multivariate volatility models. Due to nonlinearities and latent processes, Bayesian statistics is the natural approach to inference in our proposed models, which also enables us to formally compare competing specifications and, thus, to check which elements of model structure are most important for good fit to the observed data. We propose to conduct Bayesian inference using a hybrid MCMC sampler that enables to efficiently simulate from the posterior distribution. The marginal data density value, which is the main ingredient of Bayesian model comparison, can be approximated with the Newton and Raftery

(1994) harmonic mean estimator (HME), but adjusted according to the suggestion by Lenk (2009). We propose a specific, practical way of computing Lenk's correction in our highly dimensional space of parameters and latent variables.
The aim of the paper is to formally test the long-run relationship between daily prices and to check whether taking it into account (if it is present) influences (improves) our inference on volatility and short-run relations between returns on different markets. In the empirical example the daily values of the S&P500 index and the WTI oil price in the period 19.12.2005 – 30.09.2011 are jointly modelled. It is shown that, although the logarithms of both series are linked and the ECM term is "significant", the model with no ECM term leads to practically the same conclusions on volatility and conditional correlation. The paper is organized as follows. In the next section details of our Bayesian VAR(2)–MSF-SBEKK model for daily prices are presented. Section 3 is devoted to simulation from the posterior distribution and to numerical problems of model comparison. Our empirical example is presented in Section 4. Section 5 contains concluding remarks.

2 Bayesian VAR(2)–MSF-SBEKK model for prices

Let us denote the natural logarithm of the time t price of asset $i \in \{1, \ldots, n\}$ by $0.01\, x_{t,i}$ and the growth rate of the price of asset i at time t by $r_{t,i} = x_{t,i} - x_{t-1,i}$, for any time $t \in \{1, \ldots, T\}$. Then let us denote the n-variate vectors of hundreds of logarithms of prices and logarithmic return rates (in percentage points) by $x_t = (x_{t,1}, \ldots, x_{t,n})'$ and $r_t = (r_{t,1}, \ldots, r_{t,n})'$, respectively. We will model the prices using an n-variate VAR(2) process, which in terms of the r_t series takes the form:

$$r_t = \lambda + \Lambda r_{t-1} + \Pi x_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1}$$

where either we assume $\Pi = 0$ (as usually) or matrix Π is further decomposed as presented in subsection 2.1, and ε_t is described in subsection 2.2.

2.1 Imposing cointegration structure

Assuming that individual processes in x_t are linked through m cointegration relationships ($1 \leqslant m \leqslant n-1$), we can decompose the $n \times n$ matrix Π in (1) as

$$\Pi = ab', \tag{2}$$

with full rank matrices b and a of dimensions $n \times m$. As the data contain information only about the cointegrating space (but not about vectors spanning it), for any nonsingular matrix C the decompositions $\Pi = ab'$ and $\Pi = (aC)\left(b\left(C^{-1}\right)'\right)'$ are equivalent. Therefore we impose the restriction

$$b'b = I_m, \tag{3}$$

which means that the vector b belongs to the Stiefeld manifold $V_{m,n}$, see Strachan R. (2003), Koop G., León-Gonzalez R. and Strachan R. (2009), and Wróblewska J. (2010). Since in this paper we only discuss the bivariate case, we will use the fact that for $n = 2$ introducing the uniform prior on the angle in polar coordinates is equivalent to an uninformative prior on the manifold. That is, having $n = 2$, we assume $m = 1$ and represent b in (3) in terms of its polar coordinates:

$$b = \begin{bmatrix} \sin(\kappa) \\ -\cos(\kappa) \end{bmatrix}, \quad \kappa \in [0, \pi]. \tag{4}$$

In such case, the uniform prior of κ is uninformative on the cointegrating space. Thus, for the bivariate cointegrated series, we can write matrix Π explicitly as

$$\Pi' = \begin{bmatrix} \sin(\kappa) \\ -\cos(\kappa) \end{bmatrix} \begin{bmatrix} a_1 & a_2 \end{bmatrix}, \quad \kappa \in [0, \pi], \quad a_1, a_2 \in \mathbb{R}. \tag{5}$$

2.2 MSF-SBEKK modelling framework for disturbances

In hybrid MSV-MGARCH structures, like the MSF-SBEKK specification proposed by Osiewalski J. (2009) and Osiewalski J. and Pajor A. (2009), the error term can be decomposed as follows:

$$\varepsilon_t = G_t^{\frac{1}{2}} H_t^{\frac{1}{2}} \xi_t \quad (t = 1, \ldots, T), \tag{6}$$

where G_t is the MSV component and H_t is the MGARCH one. The MSF case of the MSV part amounts to assuming

$$G_t = g_t I_n, \quad \ln g_t = \phi \ln g_{t-1} + \zeta_t, \tag{7}$$

i.e. that G_t depends on one AR(1) latent process (with stationarity restriction $|\phi|<1$). In the MSF-SBEKK model the MGARCH part is also represented by a simple structure, i.e. by H_t coming from the scalar BEKK(1,1) model:

$$H_t = (1 - \beta - \gamma)A + \beta \varepsilon_{t-1} \varepsilon_{t-1}' + \gamma H_{t-1}, \tag{8}$$

restricted to the covariance stationarity case ($\beta, \gamma > 0$, $\beta + \gamma < 1$), where A is a positive definite $n \times n$ matrix. We assume that ξ_t and ζ_t are independent and follow a multivariate Normal distribution:

$$\left[\xi_t' \, \zeta_t\right]' \sim iiN^{(n+1)} \left(0_{[(n+1)\times 1]}, \begin{bmatrix} I_n & 0 \\ 0 & \tau^{-1} \end{bmatrix} \right). \tag{9}$$

The initial condition for H_t in (8) is taken as $H_0 = h_0 I_n$ with a scalar parameter $h_0 > 0$ and we assume $g_0 = 1$ to initialize (7).
Although we use two oversimplified structures (viewing from the pespective of

each individual modelling framework they represent), combined together they can outperform other more sophisticated representants of the MSV or MGARCH classes by orders of magnitude in terms of Bayes factors, see Osiewalski J. and Pajor A. (2009). Such hybrid approach enables us to both capture outliers well (due to the presence of a latent process) and include information up to time $t-1$ in the conditional covariance matrix at time t.
However, there is a theoretical problem with using the hybrid MSF-SBEKK structure for the disturbance term in (1) in the context of cointegration analysis. We do not know any stationarity conditions for the stochastic process defined by (6)-(9). It was proposed *ad hoc*, in order to greatly improve model fit at relatively low computational cost. It has worked very well in all applications considered until now, but we cannot say that stationarity conditions imposed on the MSF special case ($|\phi| < 1$) and the SBEKK special case ($\beta + \gamma < 1$) are jointly sufficient for covariance stationarity of the MSF-SBEKK process. Therefore, in the empirical part, we first conduct inference on cointegration within the theoretically well defined framework of the covariance stationary MSF and SBEKK processes for ε_t, considered separately. Only when the hybrid MSF-SBEKK specification is much better in terms of Bayes factors, we will focus on analysis within this framework - more general, but less theoretically developped.

2.3 The prior distributions and Bayesian model

In the Bayesian approach, we treat as random all quantities that are unknown prior to collecting the data. In the presence of latent variables (g_t), we are usually interested in making inference on both the parameter vector, i.e. $\theta = (\lambda' \, (\mathrm{vec}\Lambda)' \, (\mathrm{vech}A)' \, \beta \, \gamma \, h_0 \, \phi \, \tau, a, \kappa)'$ and the latent variables vector: $g = (g_1 \ \ldots \ g_T)'$. The joint density of observations, $x = (x_1' \ \ldots \ x_T')'$, latent variables and parameters, which constitutes the Bayesian model, can be factorized as follows:

$$\begin{aligned} p(x, g, \theta) &= p(\theta) p(x, g|\theta) \\ &= p(\theta) \prod_{t=1}^{T} p(x_t|\psi_{t-1}, g_t, \theta) p(g_t|\psi_{t-1}, \theta), \end{aligned} \tag{10}$$

where ψ_{t-1} denotes the past of both x_t and g_t. It is worth stressing here that the MSF–SBEKK structure is based on two basic conditional independence assumptions, which hold for any value of θ. Firstly, x_t is independent of the past of g_t, given g_t itself and the past of x_t. Secondly, g_t is independent of the past of x_t, given the past of g_t. Thus in (10) we have

$$p(g_t|\psi_{t-1}, \theta) = p(g_t|g_{t-1}, \theta) = g_t^{-1} f_N \left(\ln g_t | \phi \ln g_{t-1}, \tau^{-1} \right), \tag{11}$$

which is a univariate log-normal density, and, for $\mu_t = \lambda + \Lambda r_{t-1} + (I + \Pi) x_{t-1}$,

$$p(x_t|\psi_{t-1}, g_t, \theta) = f_N^n \left(x_t | \mu_t, g_t H_t \right), \tag{12}$$

which is a multivariate Normal density function.
In order to complete the Bayesian model, let us specify the prior distribution of the parameter vector θ. We assume a proper joint prior and subjectively set the marginal distributions of interest, which reflect our weak knowledge about the model parameters (see Osiewalski J. and Pajor A. (2009)). Let us assume that:

$$p(\theta) = p(\lambda)p(\text{vec}\Lambda)p(A^{-1})p(\beta,\gamma)p(h_0)p(\phi)p(\tau)p(a)p(\kappa), \tag{13}$$

which means prior independence among blocks of parameters. Furthermore we take:

$p(\lambda) = f_N^n(\lambda|0, \frac{1}{10}I_n)$ – the n-variate Normal density with mean 0 and covariance matrix $\frac{1}{10}I_n$,

$p(\text{vec}\Lambda) \propto f_N^{n^2}(\text{vec}\Lambda|0, I_{n^2})\mathbf{1}_{\{M:\, \rho(M)<1\}}(\Lambda)$ – a multivariate Normal distribution truncated by the restriction that all eigenvalues of Λ lie inside the unit circle, where $\rho(M)$ is the spectral radius of matrix M and $\mathbf{1}_S(x)$ is the indicator function of the set S: $\mathbf{1}_S(x) = \begin{cases} 1, & x \in S \\ 0, & x \notin S \end{cases}$,

$p(A^{-1}) = f_{Wishart}(A^{-1}|I_n, n+2)$ – the Wishart distribution with mean I_n and $n+2$ degrees of freedom,

$p(\beta,\gamma) \propto \mathbf{1}_{\{(x,y)\in[0,1]^2:\, x+y<1\}}(\beta,\gamma)$ – the uniform distribution over the unit simplex,

$p(h_0) = f_{Exp}(h_0|1)$ – the Exponential distribution with mean 1,

$p(\phi) \propto f_N(\phi|0,100)\mathbf{1}_{\{x:\, |x|<1\}}(\phi)$,

$p(\tau) = f_{Exp}(\tau|200)$ – the Exponential distribution with mean 200,

$p(a) = f_N^n(a|0, \frac{1}{100})$,

$p(\kappa) = \frac{1}{\pi}\mathbf{1}_{[0,\pi]}(\kappa)$.

We can write the joint density function that represents our bivariate Bayesian reduced rank VAR(2)–MSF-SBEKK model as:

$$\begin{aligned} p(x,g,\theta) \;=\; & p(a)p(\kappa)p(\lambda)p(\text{vec}\Lambda)p(A^{-1})p(\beta,\gamma)p(h_0)p(\phi)p(\tau)\cdot \\ & \cdot \prod_{t=1}^{T} g_t^{-1} f_N\left(\ln g_t|\phi \ln g_{t-1}, \tau^{-1}\right) f_N^n\left(x_t|\mu_t, g_t H_t\right). \end{aligned} \tag{14}$$

The posterior distribution of all unobservable quantities (i.e. latent variables and parameters) is characterised by the conditional density function $p(\theta, g|x)$, which is proportional to $p(\theta, g, x)$ in (14).

3 MCMC for inference and model comparison

3.1 Sampling from the posterior

The joint posterior distribution, represented by the density function $p(\theta, g|x)$, is highly dimensional and too complicated to obtain any analytical results. In this case, Monte Carlo methods must be applied in order to generate a (pseudo) random sample from the posterior distribution and to obtain estimates of posterior characteristics. Following Osiewalski J. (2009) and Osiewalski J. and Pajor A. (2009), we use a hybrid Markov Chain Monte Carlo method: the Gibbs sampler with Metropolis and Hastings (MH) steps. The sampler sequentially draws from the conditional posterior distributions that result from (14) under a natural block partition of all unknown quantities in the model.
For the parameters λ, Λ, A, β, γ, h_0, ϕ, τ and the latent process $\{g_t\}_{t=1,\ldots,T}$ we use the same sampling scheme as Osiewalski K. and Osiewalski J. (2012). The remaining conditional posterior distributions, used to complete the algorithm, are listed below.

1. The bivariate adjustment parameter a has the following density:

$$p(a|\lambda, \Lambda, A, \beta, \gamma, h_0, \phi, \tau, g, \kappa, x) \propto p(a) \prod_{t=1}^{T} f_N^n \left(x_t|\mu_t, g_t H_t\right). \tag{15}$$

   Similarly as for VAR parameters, we cannot directly sample from this conditional posterior distribution (as in pure MSV models) due to the presence of the MGARCH (SBEKK) structure. Thus, the MH step is implemented. The choice of the proposal distribution is arbitrary – it is set to be a bivariate Normal distribution centered at the previous state of the Markov chain (Random Walk MH). The covariance matrix of the proposal distribution is set to the "posterior" covariance matrix (multiplied by a factor of 2) obtained from initial cycles, which were performed to calibrate the sampling mechanism. The resulting acceptance rate oscillated between 1 and 13 percent in the empirical example presented in the next section.

2. For the angle κ we have:

$$p(\kappa|\lambda, \Lambda, A, \beta, \gamma, h_0, \phi, \tau, g, a, x) \propto p(\kappa) \prod_{t=1}^{T} f_N \left(x_t|\mu_t, g_t H_t\right). \tag{16}$$

   Here again an MH step enables sampling from this conditional distribution. We sample candidate states from the univariate Normal distribution, but truncated to the $[0, \pi]$ interval. The parameters of the candidate generating distribution are: previous state for the mean (Random Walk MH) and initial "posterior" variance multiplied by 3. The acceptance ratio in the empirical example oscillated around 6%.

For such an algorithm as constructed above, the convergence is monitored via standardised CUMSUM plots; see Yu B. and Mykland P. (1998) and Pajor A. (2003). The chain length was set to 2.1 million states, in which the first 1.6 million were considered as the *burn-in* period. The sampling speed was approximately 30% slower than in the MSF-SBEKK model with no long run adjustments. Such a big difference resulted from the requirement of re-computing all H_t, $t = 1, \ldots, T$ twice more in each chain state (in addition to four more times, i.e. for λ, Λ, A and jointly (β, γ)). As for the SV parameters ϕ and τ the formulas do not involve matrix calculation of H_t, their drawing is far faster. All empirical results in the following section are based on the last 500,000 MCMC states, treated as a sample from the posterior distribution.

3.2 Model comparison via HME with Lenk's correction

Bayesian model comparison enables us to formally assess the relative explanatory power of K competing specifications. The marginal data density of model M_k $(k = 1, \ldots, K)$, with all the parameters and latent variables collected in $\omega_{(k)} \in \Omega_k$, is defined as:

$$p\left(x|M_k\right) = \int_{\Omega_k} p\left(x, \omega_{(k)}|M_k\right)\, d\omega_{(k)} = \int_{\Omega_k} p\left(x|\omega_{(k)}, M_k\right) p\left(\omega_{(k)}|M_k\right)\, d\omega_{(k)}. \quad (17)$$

Within the MCMC sampler, under a proper prior, the value of $p(x|M_k)$ for the observed data can be approximated using the Newton and Raftery (1994) harmonic mean estimator:

$$p\left(x|M_k\right) \approx \left[\frac{1}{N}\sum_{q=S+1}^{S+N} \frac{1}{p\left(x|\omega_{(k)}^{(q)}, M_k\right)}\right]^{-1}, \quad (18)$$

where S is the number of the burn-in cycles, N – the number of drawings from the posterior distribution, q – the index of a single Markov chain state $(q = 1, \ldots, S + N)$, and $\omega_{(k)}^{(q)}$ – the draw of $\omega_{(k)}$ from the q-th cycle. Although heavily used, the HME suffers from both theoretical and empirical weaknesses. As it is widely known it has the infinite asymptotic variance property. It was also shown by Lenk (2009) that in practice the HM estimator may have a "pseudo bias" resulting from the difference in the prior support and the posterior *numerical* support. In (17) we are integrating over the parameter space, i.e. the whole support Ω_k of the prior and posterior distributions. Thus, in order to obtain the correct value of the HME, the MCMC chain should visit all of the regions of the parameter space – which often does not happen, as the posterior *numerical* support A_k (i.e., the region visited by the MCMC sampler) is usually significantly smaller. To overcome this problem, Lenk (2009) splits Ω_k into

A_k and its compliment A'_k, which leads to:

$$\begin{aligned} p\left(x|M_k\right) &= \int_{A_k} p\left(x|\omega_{(k)}, M_k\right) p\left(\omega_{(k)}|M_k\right)\, d\omega_{(k)} + \\ &+ \int_{A'_k} p\left(x|\omega_{(k)}, M_k\right) p\left(\omega_{(k)}|M_k\right)\, d\omega_{(k)}. \end{aligned} \tag{19}$$

We can then write the first integral in the sum as

$$\begin{aligned} &\int_{A_k} p\left(x|\omega_{(k)}, M_k\right) p\left(\omega_{(k)}|M_k\right)\, d\omega_{(k)} \\ &= Pr\left(A_k|M_k\right) \int_{A_k} p\left(x|\omega_{(k)}, M_k\right) \frac{p\left(\omega_{(k)}|M_k\right)}{Pr\left(A_k|M_k\right)} d\omega_{(k)} \\ &= Pr\left(A_k|M_k\right) \int_{A_k} p\left(x|\omega_{(k)}, M_k\right) p\left(\omega_{(k)}|A_k, M_k\right)\, d\omega_{(k)} \end{aligned} \tag{20}$$

and the second as

$$\int_{A'_k} p\left(x|M_k\right) p\left(\omega_{(k)}|x, M_k\right)\, d\omega_{(k)} = p\left(x|M_k\right) Pr\left(A'_k|x, M_k\right). \tag{21}$$

If the MCMC sampler converges to the posterior distribution and the generated sample is large enough to represent this distribution sufficiently well, the numerical support (A_k) contains practically all posterior probability mass. So, for A'_k it can be assumed that its posterior probability and, thus, the second integral in the sum, are negligible. Finally, we can then approximate the marginal data density as

$$p\left(x|M_k\right) \approx Pr\left(A_k|M_k\right) \int_{A_k} p\left(x|\omega_{(k)}, M_k\right) p\left(\omega_{(k)}|A_k, M_k\right)\, d\omega_{(k)}. \tag{22}$$

Since the MCMC chain visits only A_k, the MHE is calculated as if A_k constituted the whole parameter space, i.e. it is the integral in (22) that is numerically approximated by the HME.

The question arises about the practical choice of set A_k, which prior probability has to be calculated. Lenk (2009) suggests to take a subset for which the likelihood function does not drop below a certain level:

$$A_k = \left\{\omega_{(k)} \in \Omega_{(k)} :\ p\left(x|\omega_{(k)}, M_k\right) \geqslant \underline{L}\right\}, \tag{23}$$

where $\underline{L} = \min_{q\in\{S+1,\ldots,S+N\}} p\left(x|\omega_{(k)}^{(q)}, M_k\right)$. Additionally, the Monte Carlo – Importance Sampling (MC-IS) computational strategy is advocated to calculate the prior probability of A_k. There are two objections towards this choice of A_k. First, it can be computationally too demanding as the condition defining A_k binds all the parameters and latent variables (about 1500 in the case of our models and data).

Second, since the sampler output constitutes a set of measure zero in the parameter space, the natural candidate for A_k is the set containing, and bounded by, the convex hull of sampled points. As the relation of (23) to this convex hull is unclear, in this paper we use the most direct approach, which is to assume A_k to be the intersection of the parameter space and the cube limited by the range of the sampler output:

$$A_k = \Omega_k \cap \bigtimes_{l=1}^{l_k} \left[\min_{q=1,\ldots,N} \left(\omega_{l,(k)}^{(q)} \right), \max_{q=1,\ldots,N} \left(\omega_{l,(k)}^{(q)} \right) \right], \tag{24}$$

where l_k denotes the parameter space dimension in the k-th specification (including the number of latent variables). We then directly calculate by Monte Carlo (or analytically whenever possible) the prior probability of the set A_k. Although the dimension of the parameter space is large, we can benefit here from assumed prior independence among blocks of parameters in (13). We can draw each block separately and thus make our sampling from the prior distribution more efficient. Note that our choice of A_k contains the convex hull of the sampler output, thus our version gives an upper bound of Lenk's correction. An example, related to the SBEKK parameters (β, γ), is given in Figure 1, where the grey area shows the parameter space, the black points represent the sampler output from the last 200 000 MCMC draws and the dark grey area shows our A_k based on all MCMC draws from the posterior.

4 Relationships between the U.S. stock market and the price of oil

4.1 Data description

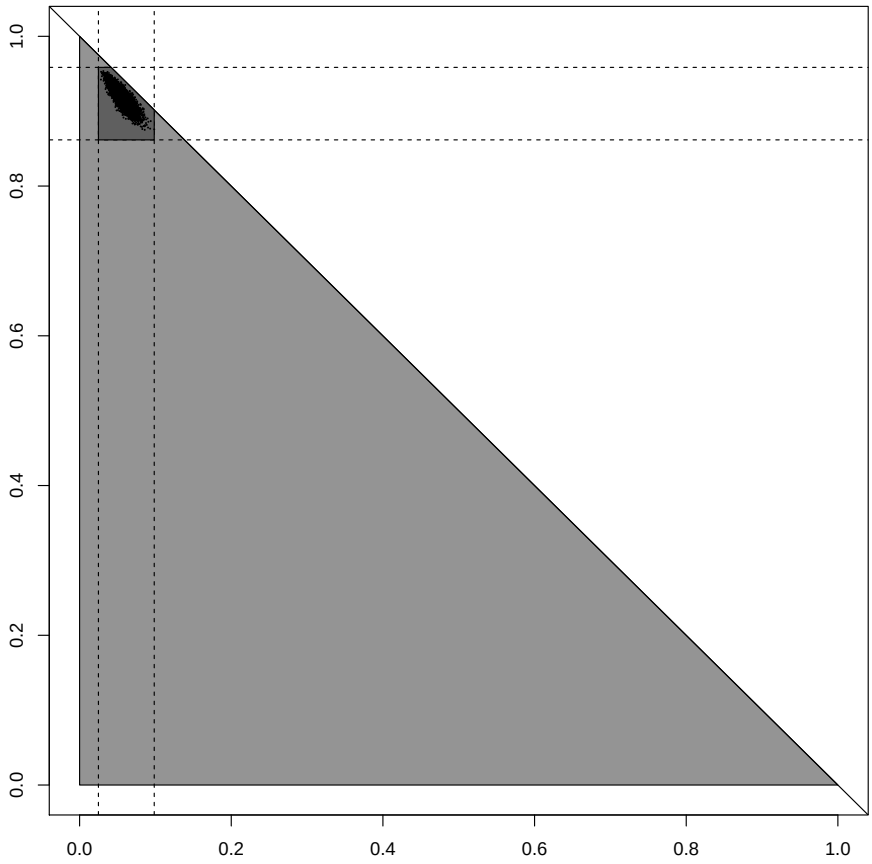
Daily observations from the period 21.12.2005 – 30.09.2011 and $n = 2$ different markets are considered:

the U.S. stock market is represented by the S&P500 index $(x_{t,1})$,

the crude oil market is represented by the WTI spot price in USD per barrel $(x_{t,2})$.

As one of the the objects of the study is a long term relationship between oil prices and the U.S. economy, the WTI oil price was chosen. Although the Brent spot price has recently been chosen as the reference price by the U.S. Energy Information Administration (2013), this decision is motivated by the fact that the Brent price better reflects the global oil demand and economic growth. However, when focusing on the U.S. only, choosing WTI seems the natural choice.
Both markets function in the same country, however there were 5 days when only one of the markets was in operation. As Osiewalski K. and Osiewalski J. (2012) suggest, the missing data points were linearly interpolated to avoid potential consequences

Figure 1: An example of the choice of A_k



of introducing artificial jumps into the series. All the days when none of the assets was valued were removed from the data set (38 in total). After these adjustments, $T = 1455$ days were taken into the analysis. There are at least two important features

Table 1: Descriptive statistics of analyzed growth rates (r_t)

	min	max	mean	sd	skewness	kurtosis	correlation
S&P500	-9.47	10.957	-0.007	1.546	-0.294	8.364	0.329
OIL	-12.827	16.414	0.024	2.661	0.072	4.606	

of the empirical distribution of r_t. The first one is a very high kurtosis of S&P500, resulting mainly from abnormal volatility on the stock market at the end of 2008 (see Figure 3). The second one is positive correlation (about 0.33) between logarithmic

returns on both markets. The empirical correlation coefficient for levels (x_t) is even higher (0.386). This positive correlation in the logarithms of price levels is visible in Figure 2 – especially during the *sub-prime* crisis, when both assets suffered from a downward pressure resulting from economic perturbations.

Figure 2: Price data – hundreds of logarithms

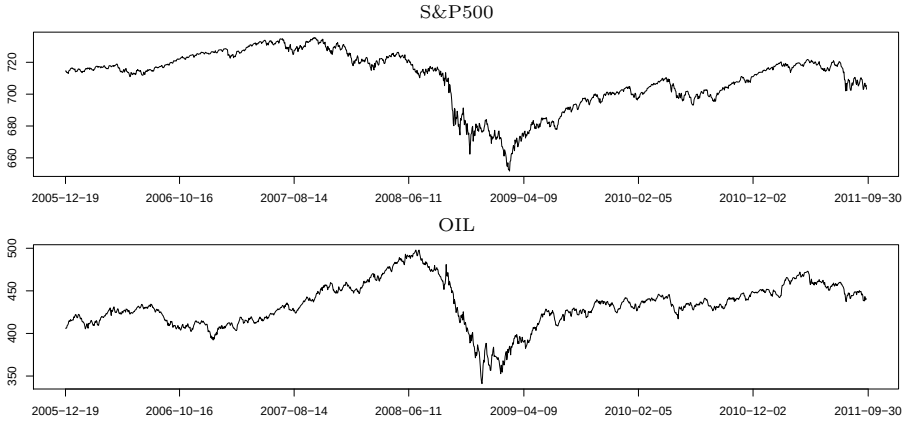
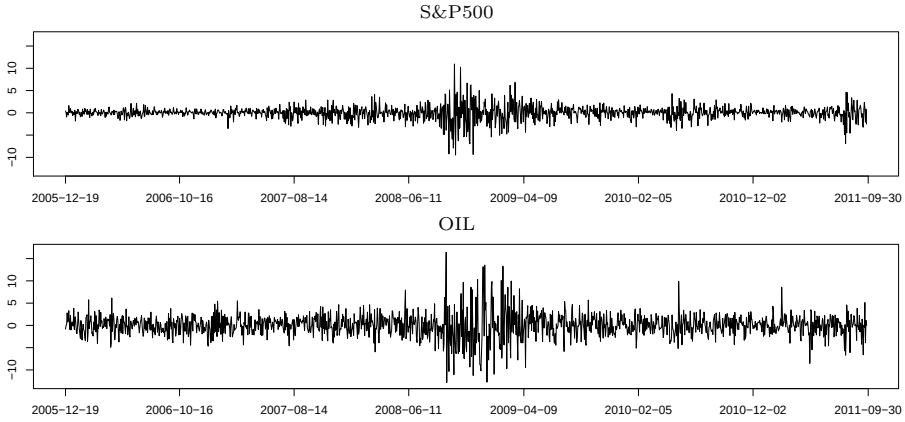


Figure 3: Percentage return rates



4.2 Empirical results

4.2.1 Model comparison and the long-run relationship

Let us focus on model comparison first. We compare two special cases of the conditional mean specification in (1): M_1 – VAR(1) for r_t ($\Pi \equiv 0$, no cointegration between logarithms of prices) and M_2 – the cointegrated VAR(2) for x_t (i.e., Π as in (5)). For both M_1 and M_2 we consider three different assumptions about the volatility structure for ε_t: the covariance stationary MSF process ($\beta = \gamma = 0$, $|\phi| < 1$), the covariance stationary SBEKK process ($g_t \equiv 1$, $\beta + \gamma < 1$) and the general MSF-SBEKK process in (6)-(9) with $|\phi| < 1$ and $\beta + \gamma < 1$. In total there are six specifications to be formally compared. The results of Lenk's adjustment and

Table 2: Lenk's correction and harmonic mean estimates

	MSF		SBEKK		MSF-SBEKK	
	M_1	M_2	M_1	M_2	M_1	M_2
	$\log_{10} Pr(A_k\|M_k)$					
λ	−0.824	−0.004	−0.83	−0.092	−0.918	−0.039
Λ	−3.301	−3.241	−3.793	−3.749	−3.620	−3.720
A	−3.324	−3.367	−1.953	−1.976	−0.479	−0.519
(β, γ)	–	–	−2.555	−2.555	−2.505	−2.458
h_0	–	–	−0.036	−0.024	−0.009	−0.015
φ	−1.750	−1.749	–	–	−0.196	−0.218
τ	−0.372	−0.375	–	–	−1.089	−1.105
a	–	−2.217	–	−2.232	–	−2.241
κ	–	−0.301	–	−0.301	–	−0.301
g	−351.2	−352.2	–	–	−0.764	−0.432
corrections in total						
$\log_{10} Pr(A_k\|M_k)$	−360.771	−363.454	−9.167	−10.929	−9.58	−11.048
Newton and Raftery HM estimates						
$\log_{10} p_{NR}(x\|M_k)$	−2339	−2336.9	−2975.8	−2973.6	−2275.9	−2273.3
final marginal data density estimate						
$\log_{10} p(x\|M_k)$	−2699.8	−2700.4	−2985	−2984.5	−2285.5	−2284.3
	$\log_{10}(BF_{M_1:M_2}) \approx -0.6$		$\log_{10}(BF_{M_1:M_2}) \approx 0.5$		$\log_{10}(BF_{M_1:M_2}) \approx 1.2$	

the HME values for the marginal data density are presented in Table 2. The number of draws at each coordinate varies from 10^4 to 10^7, but each time the estimator's stability was ensured. Surprisingly, when the HME values are properly adjusted, there is no evidence of cointegration in either the pure MSF or SBEKK framework. However, these two simple specifications are hundreds of orders of magnitude less probable a posteriori. So neither MSF nor SBEKK can compete with the hybrid MSF-SBEKK specification, where the crude HM estimate of the Bayes factor is $10^{2.6}$ in favor of M_2. The correction proposed by Lenk (2009) diminishes it to $10^{1.2}$, still indicating data support for the presence of the ECM term in the bivariate model for

growth (return) rates of S&P500 and the WTI oil price.
Remark that the adjustment proposed by Lenk (2009) is relatively small (in absolute value) for these quantities, for which the marginal posterior distribution is dispersed and similar to the prior distribution. This is particularly valid not only for h_0 (almost unchanged by the data in both models) and matrix A, but also for 1455 latent variables g_t in the MSF-SBEKK specification. It is not surprising as there is only one vector observation per one latent variable – so the marginal prior distribution of g_t cannot be strongly modified by the data if there is no conflict between the data and the prior information. Such a conflict is present (for the financial crisis period) in the case of the pure MSF process, as the posterior distributions of some latent variables are located in the tails of the prior distributions and, thus, Lenk's correction is huge in this case.
The posterior means and standard deviations of the original VECM parameters (in brackets) in M_2 are as follows

MSF

$$E(\kappa|x) = \underset{(0.236)}{\mathbf{0.679}}, \quad E(a|x) = \left[\underset{(1.20\cdot 10^{-3})}{4.58\cdot 10^{-4}} \quad \underset{(2.52\cdot 10^{-3})}{2.39\cdot 10^{-3}}\right]'$$

SBEKK

$$E(\kappa|x) = \underset{(0.143)}{\mathbf{0.560}}, \quad E(a|x) = \left[\underset{(1.17\cdot 10^{-3})}{4.449\cdot 10^{-4}} \quad \underset{(3.03\cdot 10^{-3})}{4.85\cdot 10^{-3}}\right]'$$

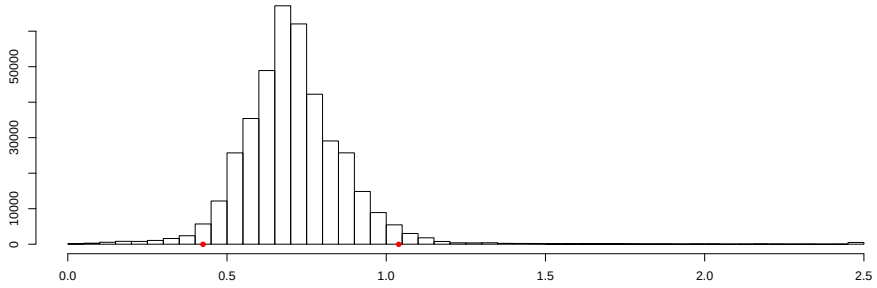
MSF-SBEKK

$$E(\kappa|x) = \underset{(0.104)}{\mathbf{0.610}}, \quad E(a|x) = \left[\underset{(1.22\cdot 10^{-3})}{1.47\cdot 10^{-4}} \quad \underset{(2.89\cdot 10^{-3})}{\mathbf{5.04}\cdot 10^{-3}}\right]',$$

When zero is not between the 0.05 and 0.95 posterior quantiles, the posterior mean is marked with bold. The conclusion, based on the Bayes factors, that the pure MSF and SBEKK specifications do not need the ECM term, is supported by the marginal posterior distributions of a_1 and a_2; both distributions are located very close to zero. The situation is completely different for the hybrid MSF-SBEKK case (which, in view of model comparison, is the only reasonable specification); here zero is in the left tail of the posterior distribution of a_2. This indicates the relevance of the ECM term in the equation describing the growth rate of the oil price. From now on we will only consider the hybrid MSF-SBEKK specification as the one describing data much better than its special cases (MSF and SBEKK).
Since the hypothesis that a_1=0 cannot be rejected by a Lindley type test at any reasonable posterior probability level, it is tempting to treat the S&P500 index as the exogenous variable (for inference on long-run parameters) in our bivariate system, although formal exogeneity testing in models with latent variables is quite complicated – see Pajor (2011). Additional complication for inference on exogenity is the presence of the GARCH structure in the model. However, it is convenient for the purpose of

interpretation to write our long term relationship between logarithms of daily prices in the form: $x_{t,2} = (\tan \kappa)\, x_{t,1} + \nu_t$, where the tangent of κ can be treated as the long-run elasticity of the WTI oil price with respect to S&P500. Looking at the posterior

Figure 4: Posterior distribution of the tangent of κ in the MSF-SBEKK model with ECM

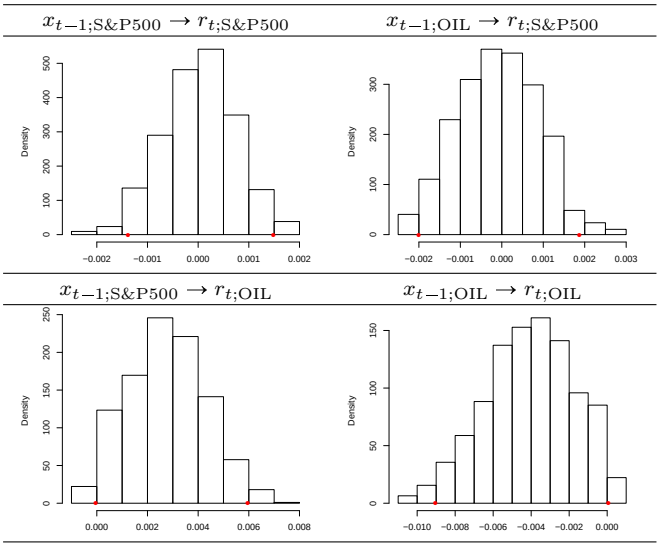


distribution of the tangent function of the angle κ (Figure 4) we see that it contains 1 in the 95% HPD interval, but not in the 90% one – the value 1 is the 0.96375 quantile. The posterior odds ratio $\frac{p(\kappa<1|r)}{p(\kappa>1|r)} \approx 0.996$ suggests that the angle κ is smaller than $\frac{\pi}{4}$. This means that the long-run elasticity of the WTI oil price with respect to S&P500 is smaller than 1; its posterior median is 0.698 with 0.174 as the posterior interquartile range (we do not report moments as the posterior distribution has a very fat tail and some κ values near $\frac{\pi}{2}$ were drawn). In view of the posterior results for our best model (MSF-SBEKK with ECM), 1% increase in the stock market index corresponds in long-run to about 0.7% increase in the WTI oil price.
Finally let us present (in Figure 5) the marginal posterior distributions of the elements of matrix Π, resulting from the posterior distribution of a and κ according to formula (5) defining M_2.

4.2.2 Short-run parameters and volatility analysis in the hybrid model

The only remarkable difference in the posterior distributions obtained in M_1 and M_2 can be seen for the λ_2 parameter, which in M_2 goes together with – and thus is modified by the presence of – the significant error correction term $a_2 b' x_{t-1}$ (with parameters presented in the second row of Figure 5). The posterior means (and standard deviations in brackets) of Λ in M_1 and M_2 are very close; in both models they indicate significant impact of both assets' lagged returns on the current stock market return. However, there is no significant impact of lagged returns on the current growth rate of the oil price.

Figure 5: Posterior histograms of the elements of Π in the MSF-SBEKK model with ECM



$$M_1 \qquad\qquad M_2$$

$$E(\lambda|x) = \begin{bmatrix} 0.086 & 0.087 \\ (0.022) & (0.049) \end{bmatrix}' \qquad E(\lambda|x) = \begin{bmatrix} 0.088 & -0.120 \\ (0.122) & (0.296) \end{bmatrix}'$$

$$E(\Lambda|x) = \begin{bmatrix} \mathbf{-0.058} & -0.012 \\ (0.025) & (0.011) \\ \mathbf{0.153} & -0.040 \\ (0.047) & (0.026) \end{bmatrix} \qquad E(\Lambda|x) = \begin{bmatrix} \mathbf{-0.060} & -0.012 \\ (0.026) & (0.011) \\ \mathbf{0.149} & -0.038 \\ (0.047) & (0.027) \end{bmatrix}$$

The almost identical posterior moments (in M_1 and M_2) of the parameters describing the latent process g_t and the SBEKK matrix H_t indicate that any analysis of volatility, risk or contagion would lead to very similar conclusions, no matter what we assumed about the long-run behaviour of the stock index and the oil price. For the empirical finance aspects of our research it would be irrelevant whether we (properly) took the error correction mechanism into account or assumed that logarithms of prices (x_t) behave like a bivariate random walk.

In fact, both M_1 and M_2 lead to the same conclusions as regards the conditional standard deviation $\sigma_{t,i} = D\left(x_{t,i}|\psi_{t-1}, \theta, g_t\right)$, which is the obvious volatility measure, and the conditional correlation coefficient $\rho_{t,i-j} = Corr\left(x_{t,i}, x_{t,j}|\psi_{t-1}, \theta, g_t\right)$ for $i, j \in \{1, 2\}$, which measures instantaneous relations between returns on both markets. In Table 3 we report time averages of the posterior means (and standard deviations, in brackets) of these volatility and correlation measures, as well as the empirical correlation coefficients between posterior means in M_1 and M_2. The results clearly indicate that inference on individual market volatility and between markets

correlation is identical in both models. Thus the relationship between logarithms of price levels, although present, is not important for this kind of inference.

$$E(A|r) = \begin{matrix} M_1 \\ \left[\begin{array}{cc} \underset{(0.146)}{\mathbf{0.595}} & \underset{(0.217)}{0.345} \\ & \underset{(0.848)}{\mathbf{3.577}} \end{array}\right] \end{matrix} \qquad E(A|x) = \begin{matrix} M_2 \\ \left[\begin{array}{cc} \underset{(0.144)}{\mathbf{0.596}} & \underset{(0.216)}{0.349} \\ & \underset{(0.833)}{\mathbf{3.546}} \end{array}\right] \end{matrix}$$

$$E(\beta|x) = \underset{(0.008)}{\mathbf{0.064}} \qquad E(\beta|x) = \underset{(0.008)}{\mathbf{0.064}}$$

$$E(\gamma|x) = \underset{(0.010)}{\mathbf{0.919}} \qquad E(\gamma|x) = \underset{(0.010)}{\mathbf{0.919}}$$

$$E(h_0|x) = \underset{(0.403)}{0.752} \qquad E(h_0|x) = \underset{(0.390)}{0.760}$$

$$E(\phi|x) = \underset{(0.147)}{\mathbf{0.227}} \qquad E(\phi|x) = \underset{(0.142)}{\mathbf{0.216}}$$

$$E(\tau^{-1}|x) = \underset{(0.061)}{\mathbf{0.317}} \qquad E(\tau^{-1}|x) = \underset{(0.060)}{\mathbf{0.318}}$$

Table 3: Comparison of volatility and conditional correlation estimates in the hybrid model with (M_2) or without (M_1) the ECM term

model	$\sigma_{\text{S\&P500}}$	σ_{OIL}	$\rho_{\text{S\&P500-OIL}}$
M_1	1.246 (0.323)	2.371 (0.616)	0.257 (0.030)
M_2	1.246 (0.323)	2.375 (0.615)	0.258 (0.030)
correlation	0.99996	0.9998	0.9998

5 Concluding remarks

In this paper we have developped a fully Bayesian bivariate VAR(2)–MSF-SBEKK framework for inference about a long-run relationship between logarithms of daily prices on two different markets. This modelling framework enables us to formally test the empirical relevance of the error correction term in the conditional mean specification as well as to compare the posterior results on returns volatility and conditional correlation in two competing models: with and without the ECM term. Our empirical example suggests that the long-run relationship between price levels, even when present, can be irrelevant for usual inference on (short-run) dynamics of logarithmic return rates. The posterior results on main quantities of interest can be the same in the reduced rank VAR(2)–MSF-SBEKK model for log-levels and in the VAR(1)–MSF-SBEKK model for logarithmic returns.

Acknowledgements

We would like to kindly thank Łukasz Kwiatkowski for bringing Lenk's (2009) paper to the seminar at Cracow University of Economics and sharing with us its important message. Useful comments by a referee are highly appreciated.

References

[1] Bekiros S. and Diks C. (2008), The relationship between crude oil spot and futures prices: Cointegration, linear and nonlinear causality, *Energy Economics* 30, 2673—2685.

[2] Chang C., Lai J. and Chuang I. (2010), Futures hedging effectiveness under the segmentation of bear/bull energy markets, *Energy Economics* 32, 442—449.

[3] Ji Q. and Fan Y. (2011), A dynamic hedging approach for refineries in multiproduct oil markets, *Energy* 36, 881–887.

[4] Koop G., León-Gonzalez R. and Strachan R. (2009), Efficient posterior simulation for cointegrated models with priors on the cointegration space, *Econometric Reviews* 29, 224–242.

[5] Lenk P. (2009), Simulation pseudo-bias correction to the harmonic mean estimator of integrated likelihoods, *Journal of Computational and Graphical Statistics* 18, 941–960.

[6] Mahadevan R. and Suardi S. (2011), The effects of uncertainty dynamics on exports, imports and productivity growth, *Journal of Asian Economics* 22, 174—188.

[7] Newton M. and Raftery A. (1994), Approximate Bayesian inference by Weighted Likelihood Bootstrap (with discussion), *Journal of the Royal Statistical Society, series B* 56, 3–48.

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

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

[10] Osiewalski J. and Pipień M. (2004), Bayesian comparison of bivariate GARCH processes. The role of the conditional mean specification. Chapter 7 in: *New Directions in Macromodelling* (ed. by A.Welfe), Elsevier, Amsterdam, pp.173-196.

[11] Osiewalski K. and Osiewalski J. (2012), Missing observations in daily returns – Bayesian inference within the MSF-SBEKK model. *Central European Journal of Economic Modelling and Econometrics* 4, 169–197.

[12] Pajor A. (2003), *Procesy zmienności stochastycznej w bayesowskiej analizie finansowych szeregów czasowych (Stochastic Variance Processes in Bayesian analysis of Financial Time Series)*, Monografie: Prace Doktorskie Nr 2, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.

[13] Pajor A. (2011), A Bayesian analysis of exogeneity in models with latent variables. *Central European Journal of Economic Modelling and Econometrics* 4, 49–73.

[14] Strachan R. (2003), Valid Bayesian estimation of the cointegrating error correction model, *Journal of Business & Economic Statistics*, vol. 21, p. 185–195.

[15] U.S. Energy Information Administration, *Annual Energy Outlook* 2013 – Early Release Overview. Available at `http://www.eia.gov/forecasts/aeo/er/pdf/0383er%282013%29.pdf`

[16] Wróblewska J. (2010), *Modele i metody bayesowskiej analizy kointegracji (Models and methods of Bayesian cointegration analysis)*, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie (Cracow University of Economics Press), Kraków.

[17] Yu B. and Mykland P. (1998), Looking at Markov samplers through cusum path plots: a simple diagnostic idea, *Statistics and Computing* 8, 275–286.