

A Bayesian Analysis of Exogeneity in Models with Latent Variables

Anna Pajor*

Submitted: 10.12.2011, Accepted: 24.01.2012

Abstract

This paper presents some new results on exogeneity in models with latent variables. The concept of exogeneity is extended to the class of models with latent variables, in which a subset of parameters and latent variables is of interest. Exogeneity is discussed from the Bayesian point of view. We propose sufficient weak and strong exogeneity conditions in the vector error correction model (VECM) with stochastic volatility (SV) disturbances. Finally, an empirical illustration based on the VECM-SV model for the daily growth rates of two main official Polish exchange rates: USD/PLN and EUR/PLN, as well as EUR/USD from the international Forex market is presented. The exogeneity of the EUR/USD rate is examined. The strong exogeneity hypothesis of the EUR/USD rate is not rejected by the data.

Keywords: exogeneity, Bayesian cuts, latent variables, non-causality, stochastic volatility

JEL Classification: C11, C32.

*Cracow University of Economics, e-mail: eopajor@cyf-kr.edu.pl

Anna Pajor

1 Introduction

In most econometric models some of variables are treated as if they were not random, i.e. the marginal process of these variables is not specified. When such a reduction of the model is admissible (there is no loss of relevant information), these conditioning variables are called "exogenous". As pointed out by Florens, Mouchart, Rolin (1990) "a variable is not exogenous by itself but is exogenous in a particular inference problem". The econometric concepts of exogeneity have been illuminated and formalized in the literature by Engle, Hendry, Richard (1983), Ericsson, Hendry, Mizon (1998), Hendry and Richard (1982, 1983). De Luna and Johansson (2006) proposed a new concept of exogeneity called Kullback-Leibler exogeneity. Recently, Rault (2011) defined "long-run strong-exogeneity" which also may be seen as a new concept of exogeneity (it only emerges in a VAR-ECM model and it is distinct from a known strong exogeneity). The Bayesian concepts of exogeneity were developed by Florens and Mouchart (see, for example, Florens and Mouchart, 1977, 1980, 1982, Florens, Mouchart, Rolin 1990, Mouchart, Russo, Wunsch 2007) , and Osiewalski and Steel (1996). However, the underlying assumption of these analyses is that the researcher is interested in a subset of model parameters, explicitly defined as *parameters of interest*. A problem arises once the researcher takes interest not only in a subset of model parameters, but also in some latent variables, which possess a specific meaning to him. The immediate two questions which arise are: first, can we infer about some parameters and latent variables basing on the conditional model alone (without loss of relevant information)? and second, can the dimensionality of the model with latent variables be reduced? These questions concern the exogeneity problem in models with unobservable variables. In this paper the concepts of exogeneity are then extended to the class of models with latent variables, in which a subset of parameters and latent variables is of interest (if we wish to infer about the parameters only, the latent variables can be integrated out and the Bayesian concepts of exogeneity in terms of the distribution of observable variables can be applied). Exogeneity is discussed from the Bayesian point of view. As an extension of our previous research (see Pajor 2008, 2010), some new results on exogeneity in models with latent variables are presented. We propose sufficient weak and strong exogeneity conditions in the vector error correction model (VECM) disturbances of which follow stochastic volatility (SV) processes. In this paper we use the terminology and notation developed in the non-causality and exogeneity literature, most notably in the works of Florens and Mouchart (1985) and Osiewalski and Steel (1996). Similarly as in these papers, we introduce a Bayesian initial cut, a Bayesian sequential cut, and the concept of non-causality in a model with latent variables. All definitions are formulated in terms of density functions. The symbol $p(\cdot)$ is generically used for densities.

The paper is organized in the following manner. In Section 2 the basic framework and the Bayesian econometric model are set. In Section 3 several types of a Bayesian cut are defined. In Section 4 the concept of exogeneity in models with latent variables is introduced. The concept of exogeneity in the Bayesian VECM-SV model with

a known cointegrating vector is elaborated on in Section 5. Finally, in Section 6 an empirical illustration using daily data of the main official Polish exchange rates is provided.

2 Bayesian econometric model

Let the triple (Ω, \mathcal{J}, P) be a probability space, where Ω is a set, \mathcal{J} is a σ -field of subsets of Ω , and P is a probability measure on \mathcal{J} . Usually $\Omega = \mathbb{R}^n$ and $\mathcal{J} = \mathcal{B}(\mathbb{R}^n)$, with the latter symbol denoting the Borel σ -field of \mathbb{R}^n . As in Florens, Mouchart, Rolin (1990), we start from a statistical experiment defined as $\{(\Omega, \mathcal{J}, P^\theta : \theta \in \Theta)\}$, where (Ω, \mathcal{J}) is a measurable space (the sample space) and $\{P^\theta : \theta \in \Theta\}$ is a family of probability measures on the sample space indexed by a vector of parameters, θ , belonging to the parameter space Θ . A Bayesian experiment is defined by the following probability space: $\{(\Theta \times \Omega, \mathcal{J}_\Theta \vee \mathcal{J}), \Pi\}$, where $(\Theta, \mathcal{J}_\Theta)$ is the parameter space with a probability measure μ , (Ω, \mathcal{J}) is the sample space, and $\Pi = \mu \otimes P^\theta$ is a probability measure on the product space $(\Theta \times \Omega, \mathcal{J}_\Theta \vee \mathcal{J})$ defined as follows:

$$\Pi(E, X) = \int_E P^\theta(X) \mu(d\theta), \quad E \in \mathcal{J}_\Theta, X \in \mathcal{J}.$$

Let $\{y_t\}$ be a stochastic process defined on $\{(\Omega, \mathcal{J}), P^\theta, \theta \in \Theta\}$, with values in \mathbb{R}^n . We assume that y_t is an observable random vector at time t and that we possess observations on y_t for $t = 1, \dots, T$. Let $\{h_t\}$ be an unobservable stochastic process defined on $\{(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)), P^\theta, \theta \in \Theta\}$, with values in \mathbb{R}^m . For the sake of simplicity we assume the same parameter space Θ . In other words, the dimensionality of Θ is extended to include all parameters.

Further, let Y_t^s denote the $n \times t - s + 1$ matrix $Y_t^s = [y_s \dots y_t]$, conformably H_t^s denote the $m \times t - s + 1$ matrix $H_t^s = [h_s \dots h_t]$ for $s \leq t$, and let Y_0 and H_0 represent the matrices of the initial conditions related to $\{y_t\}$ and $\{h_t\}$ processes, respectively. Let us denote the history of the stochastic process $\{y_t\}$ up to time $(t - 1)$ by $Y_{t-1} = [Y_0 y_1 y_2 \dots y_{t-1}]$, and the history of the stochastic process $\{h_t\}$ by $H_{t-1} = [H_0 h_1 h_2 \dots h_{t-1}]$. We assume that the process which generates the observations is continuous with respect to some appropriate measures (for example the Lebesgue measure), and it is characterized by the joint data density function:

$$p(Y_T^1 | \theta, Y_0, H_0) = \int_H p(Y_T^1, H_T^1 | \theta, Y_0, H_0) dH_T^1, \quad (1)$$

where $p(Y_T^1, H_T^1 | \theta, Y_0, H_0) = p(Y_T^1 | H_T^1, \theta, Y_0, H_0) p(H_T^1 | \theta, Y_0, H_0)$.

The Bayesian model is characterized by the joint probability density function, which can be written as the product of three densities:

$$p(Y_T^1, H_T^1, \theta | M_0) = p(Y_T^1 | H_T^1, \theta, M_0) p(H_T^1 | \theta, M_0) p(\theta | M_0), \quad (2)$$

Anna Pajor

where $p(Y_T^1|H_T^1, \theta, M_0)$ is the conditional density of Y_T^1 given H_T^1 and $\theta \in \Theta$, $p(H_T^1|\theta, M_0)$ is the density of the latent variables conditional on θ , $p(\theta|M_0)$ is the prior density function, for the sake of simplicity the initial conditions $M_0 = [Y_0 H_0]$ are trivial (i.e. M_0 is not random).

3 Bayesian cuts

Let us now focus on the concept of Bayesian initial and global cuts (similar to Osiewalski and Steel, 1996). We partition y_t into $y_t = (z_t', w_t')'$, $z_t \in \mathbb{R}^q$, $w_t \in \mathbb{R}^p$, $p + q = n$, and Y_t^s , Y_t , H_t^s into $Y_t^s = \begin{bmatrix} Z_t^s \\ W_t^s \end{bmatrix}$, $Y_t = \begin{bmatrix} Z_t \\ W_t \end{bmatrix}$, $H_t^s = \begin{bmatrix} H_t^{s,z} \\ H_t^{s,w} \end{bmatrix}$, respectively.

Definition 1 *Let there exist some reparameterization (one-to-one transformation) of θ in θ_1 and θ_2 , as well as a partition of H_t^1 into $H_t^{1,w}$ and $H_t^{1,z}$. Then the pair $\left[(\theta_1; H_t^{1,w}; W_t^1), (\theta_2; H_t^{1,z}; Z_t^1) \right]$ operates a Bayesian initial cut for a given sample period $\{1, \dots, T\}$ if and only if:*

$$(i) \quad p(\theta|M_0) = p(\theta_1|M_0)p(\theta_2|M_0), \quad (3)$$

$$(ii) \quad \forall_{t \in \{1, \dots, T\}} p(H_t^1|\theta, M_0) = p(H_t^{1,w}|\theta_1, M_0)p(H_t^{1,z}|\theta_2, M_0), \quad (4)$$

$$(iii) \quad \forall_{t \in \{1, \dots, T\}} p(W_t^1|Z_t^1, \theta, H_t^1, M_0) = p(W_t^1|Z_t^1, H_t^{1,w}, \theta_1, M_0), \quad (5)$$

$$(iv) \quad \forall_{t \in \{1, \dots, T\}} p(Z_t^1|H_t^1, \theta, M_0) = p(Z_t^1|H_t^{1,z}, \theta_2, M_0). \quad (6)$$

Definition 2 *The pair $\left[(\theta_1; H_t^{1,w}; W_t^1), (\theta_2; H_t^{1,z}; Z_t^1) \right]$ operates a Bayesian global cut if and only if $\left[(\theta_1; H_t^{1,w}; W_t^1), (\theta_2; H_t^{1,z}; Z_t^1) \right]$ operates a Bayesian initial cut for $t = T$.*

The first condition of definition 1 says that vectors θ_1 and θ_2 are a priori independent. The second one shows that the density $p(H_t^1|\theta, M_0)$ can be factorized into a product such that one factor, $p(H_t^{1,w}|\theta_1, M_0)$, does not depend on θ_2 , and the other factor, $p(H_t^{1,z}|\theta_2, M_0)$, does not depend on θ_1 . Note that both conditions (3) and (4) are equivalent to a single condition:

$$\forall_{t \in \{1, \dots, T\}} p(H_t^1, \theta|M_0) = p(H_t^{1,w}, \theta_1|M_0)p(H_t^{1,z}, \theta_2|M_0).$$

When we treat the latent variables as if they were additional parameters of the model, we can say that $\left(\text{vec}\left(H_t^{1,w}\right)', \theta_1'\right)'$ and $\left(\text{vec}\left(H_t^{1,z}\right)', \theta_2'\right)'$ are a priori independent. The last two conditions in definition 1 imply a factorization of the joint conditional density:

$$p\left(W_t^1, Z_t^1 | \theta, H_t^1, M_0\right) = p\left(W_t^1 | Z_t^1, H_t^{1,w}, \theta_1, M_0\right) p\left(Z_t^1 | H_t^{1,z}, \theta_2, M_0\right).$$

The conditional distribution of W_t^1 given Z_t^1, θ, H_t^1 and M_0 does not depend on $H_t^{1,z}$ and θ_2 , whereas the conditional distribution of Z_t^1 given H_t^1, θ , and M_0 does not depend on $H_t^{1,w}$ and θ_1 . Hence, no information about $H_t^{1,w}$ and θ_1 can be derived from the conditional distribution of Z_t^1 (the marginal model).

An immediate consequence of a Bayesian global cut is that vectors $\left(\text{vec}\left(H_T^{1,w}\right)', \theta_1'\right)'$ and $\left(\text{vec}\left(H_T^{1,z}\right)', \theta_2'\right)'$ are a posteriori independent. The following theorem holds:

Theorem 3 *If the pair $\left[\left(\theta_1; H_t^{1,w}; W_t^1\right), \left(\theta_2; H_t^{1,z}; Z_t^1\right)\right]$ operates a Bayesian global cut, then $\left(\text{vec}\left(H_T^{1,w}\right)', \theta_1'\right)'$ and $\left(\text{vec}\left(H_T^{1,z}\right)', \theta_2'\right)'$ are a posteriori independent.*

Note that in a Bayesian global cut, for inference on $\left(\text{vec}\left(H_T^{1,w}\right)', \theta_1'\right)'$ (or any function of $\left(\text{vec}\left(H_T^{1,w}\right)', \theta_1'\right)'$) we have to neither specify the marginal model with $\left(\text{vec}\left(H_T^{1,z}\right)', \theta_2'\right)'$ nor the prior distribution of $\left(\text{vec}\left(H_T^{1,z}\right)', \theta_2'\right)'$. The posterior inference on any measurable function of $\left(\text{vec}\left(H_T^{1,w}\right)', \theta_1'\right)'$ can be made on the basis of the conditional model (defined by $p\left(W_t^1 | Z_t^1, H_t^{1,w}, \theta_1, M_0\right)$ and $p\left(H_t^{1,w}, \theta_1 | M_0\right)$) and the prior distribution of θ_1 . Note that in the initial cut $\left(\text{vec}\left(H_t^{1,w}\right)', \theta_1'\right)'$ and $\left(\text{vec}\left(H_t^{1,z}\right)', \theta_2'\right)'$ are a posteriori independent for each $t \in \{1, \dots, T\}$, i.e.:

$$\left(\text{vec}\left(H_t^{1,w}\right)', \theta_1'\right)' \perp \left(\text{vec}\left(H_t^{1,z}\right)', \theta_2'\right)' | M_0, Y_t^1, \text{ for } t \in \{1, \dots, T\}.$$

Thus the Bayesian initial cut allows a complete separation of the inference on $\left(\text{vec}\left(H_t^{1,w}\right)', \theta_1'\right)'$ and $\left(\text{vec}\left(H_t^{1,z}\right)', \theta_2'\right)'$. As follows from theorem 3.5 in Florens and Mouchart (1977) and implication 3.2.5 in Florens, Mouchart, Rolin (1990), pp.

Anna Pajor

101, in the Bayesian global cut we obtain that

$$p(Y_T^1, \theta | M_0) = p(W_T^1 | Z_T^1, \theta_1, M_0) p(\theta_1 | M_0) p(Z_T^1 | \theta_2, M_0) p(\theta_2 | M_0). \quad (7)$$

Thus the joint data density $p(Y_T^1 | \theta, M_0)$ can be factorized as in the classical approach:

$$p(Y_T^1 | \theta, M_0) = p(W_T^1 | Z_T^1, \theta_1, M_0) p(Z_T^1 | \theta_2, M_0). \quad (8)$$

Now we consider the model from the point of view of sequential analysis, so we sequentially factorize the joint data and latent variables density as:

$$\begin{aligned} p(Y_T^1, H_T^1 | \theta, Y_0, H_0) &= \prod_{t=1}^T p(y_t, h_t | Y_{t-1}, H_{t-1}, \theta) \\ &= \prod_{t=1}^T p(y_t | h_t, Y_{t-1}, H_{t-1}, \theta) p(h_t | Y_{t-1}, H_{t-1}, \theta). \end{aligned} \quad (9)$$

Since a dynamic framework is under consideration here, it is natural to define Bayesian one-shot and sequential cuts. The following definitions are adapted from Osiewalski and Steel (1996) and Florens, Mouchart, Rolin (1990).

Definition 4 Let $(\theta'_1, \theta'_2)'$ be a reparameterization of θ . For a given $t \in \mathbb{Z}$ the vector h_t is partitioned into $(h_t^{w'}, h_t^{z'})'$. The pair $[(\theta_1; h_t^w; w_t), (\theta_2; h_t^z; z_t)]$ operates a Bayesian one-shot cut if and only if:

$$(i) \quad p(\theta | M_0) = p(\theta_1 | M_0) p(\theta_2 | M_0), \quad (10)$$

$$(ii) \quad p(h_t | Y_{t-1}^1, H_{t-1}, \theta, M_0) = p(h_t^w | H_{t-1}^{1,w}, \theta_1, M_0) p(h_t^z | H_{t-1}^{1,z}, \theta_2, M_0), \quad (11)$$

$$(iii) \quad p(w_t | z_t, H_t^1, Y_{t-1}^1, \theta, M_0) = p(w_t | z_t, \theta_1, H_t^{1,w}, Y_{t-1}^1, M_0), \quad (12)$$

$$(iv) \quad p(z_t | H_t^1, Y_{t-1}^1, \theta, M_0) = p(z_t | \theta_2, H_t^{1,z}, Y_{t-1}^1, M_0). \quad (13)$$

Definition 5 Let $(\theta'_1, \theta'_2)'$ be a reparameterization of θ . For each $t \in \mathbb{Z}$ the vector h_t is partitioned into $(h_t^{w'}, h_t^{z'})'$. The pair $[(\theta_1; h_T^w; w_T), (\theta_2; h_T^z; z_T)]$ operates a Bayesian sequential cut if and only if the pair $[(\theta_1; h_t^w; w_t), (\theta_2; h_t^z; z_t)]$ operates a Bayesian one-shot cut for all t .

A Bayesian sequential cut implies a total separation of information about the parameters and latent variables between the conditional and marginal models and leads to posterior independence.

Theorem 6 If the pair $[(\theta_1; h_t^w; w_t), (\theta_2; h_t^z; z_t)]$ operates a Bayesian sequential cut, then $(h_1^{w'}, \dots, h_T^{w'}, \theta'_1)'$ and $(h_1^{z'}, \dots, h_T^{z'}, \theta'_2)'$ are a posterior independent.

Theorem 6 indicates that there is no need to use the full probability density function $p(Y_T^1, H_T^1 | \theta, M_0) = \prod_{t=1}^T p(y_t, h_t | Y_{t-1}, H_{t-1}, \theta)$, or the joint prior density $p(\theta | M_0)$ for the purpose of inference on any function of $(h_1^{w'}, \dots, h_T^{w'}, \theta_1)'$. We can infer about $(h_1^{w'}, \dots, h_T^{w'}, \theta_1)'$ basing only on the conditional model and on the prior distribution of θ_1 . In other words, the posterior density of the parameters and latent variables of interest derived from the conditional model "coincides with" the one derived from the complete model. Obviously, the Bayesian sequential cut is not necessary for it. Finally, we define the concept of non-causality (analogous to Granger noncausality, see Granger 1969, Engle, Hendry, Richard 1983).

Definition 7 We say that w_t does not cause z_t given $H_t^{1,z}$ and θ_2 if and only if

$$p(z_t | H_t^{1,z}, Y_{t-1}^1, \theta_2, M_0) = p(z_t | H_t^{1,z}, Z_{t-1}^1, \theta_2, M_0). \quad (14)$$

Note that when w_t does not cause z_t given $H_t^{1,z}$ and θ_2 , the density $p(z_t | H_t^{1,z}, Y_{t-1}^1, \theta_2, M_0)$ does not depend on W_{t-1}^1 . Thus the past of w_t does not influence the conditional distribution of z_t given the past of z_t , $H_t^{1,z}$ and θ_2 . This concept of non-causality is very similar to that proposed by Osiewalski and Steel (1996). Here the conditioning is extended to selected latent variables (which in Bayesian model can be treated as additional parameters).

Now we consider the relationship between initial and sequential cuts. It is interesting to know the conditions under which an initial cut is also a sequential cut. We can prove the following theorem, which shows that the Bayesian sequential cut and non-causality imply the Bayesian initial cut.

Theorem 8 If

(1) the pair $[(\theta_1; h_t^w; w_t), (\theta_2; h_t^z; z_t)]$ operates a Bayesian sequential cut over the sample period $\{1, \dots, T\}$,

(2) $\forall_{t \in \{1, \dots, T\}}$ w_t does not cause z_t given $H_t^{1,z}$ and θ_2 ,

then there exists a Bayesian initial cut for the sample period $\{1, \dots, T\}$.

Theorem 8 shows the relationship between sequential and initial cuts. The Bayesian sequential cut and non-causality, given $H_t^{1,z}$ and θ_2 , imply the Bayesian initial cut. The basic idea of the proof is that the non-causality condition ensures that the factorization of the data density implied by the Bayesian sequential cut coincides with that from the initial cut.

Anna Pajor

4 Exogeneity

Following the works of Engle, Hendry, Richard (1983), and Osiewalski and Steel (1996) we use the term "weak exogeneity" for the concept that validates inference based solely on the conditional model without taking the marginal model into consideration. In other words, z_t is weakly exogenous for the parameters and latent variables of interest, if the inference based on the complete process $\{y_t\}$ is the same as the one based on the conditional process $\{w_t|z_t\}$.

Definition 9 (*weak exogeneity*) We say that z_t is weakly exogenous over the sample period $\{1, \dots, T\}$ for a measurable function of H_T^1 and θ (i.e. $f(H_T^1, \theta)$) if and only if there exists a reparameterization of θ in θ_1 and θ_2 , as well as a partition of H_T^1 into $H_T^{1,w}$ and $H_T^{1,z}$ such that:

(i) a Bayesian one-shot cut is obtained for each $t \in \{1, \dots, T\}$ (with $(\theta'_1, \theta'_2)'$ and $(h_t^{w'}, h_t^{z'})'$),

(ii) $f(H_T^1, \theta) = f_0(H_T^{1,w}, \theta_1)$.

Condition (i) gives the separation of information between the conditional and marginal models, whereas condition (ii) ensures that we can infer about some function of the parameters and latent variables from the conditional model only.

It is worth noticing that the Bayesian sequential cut (see definition 4) implies weak exogeneity of z_t for $f_0(H_T^{1,w}, \theta_1)$.

For predictive inference on w_t given z_t , weak exogeneity is not sufficient. For the purpose of conditional forecasting we need to define predictive and strong exogeneity.

Definition 10 (*predictive exogeneity*) We say that z_t is predictively exogenous over the forecasting period $\{T+1, \dots, T+s\}$ if and only if there exists a reparameterization of θ in θ_1 and θ_2 , as well as a partition of H_{T+s}^1 into $H_{T+s}^{1,w}$ and $H_{T+s}^{1,z}$ such that:

(i) $p(H_T^1, \theta | Y_T^1, M_0) = p(H_T^{1,w}, \theta_1 | Y_T^1, M_0) p(H_T^{1,z}, \theta_2 | Y_T^1, M_0)$,

(ii) $p(H_{T+s}^{T+1} | H_T^1, \theta, Y_T^1, M_0) =$
 $p(H_{T+s}^{T+1,w} | H_T^{1,w}, \theta_1, Y_T^1, M_0) p(H_{T+s}^{T+1,z} | H_T^{1,z}, \theta_2, Y_T^1, M_0)$,

(iii) $p(W_{T+s}^{T+1} | Z_{T+s}^{T+1}, H_{T+s}^{T+1}, H_T^1, \theta, Y_T^1, M_0) =$
 $p(W_{T+s}^{T+1} | Z_{T+s}^{T+1}, H_{T+s}^{T+1,w}, H_T^{1,w}, \theta_1, Y_T^1, M_0)$,

(iv) $p(Z_{T+s}^{T+1} | H_{T+s}^{T+1}, H_T^1, \theta, Y_T^1, M_0) = p(Z_{T+s}^{T+1} | H_{T+s}^{T+1,z}, H_T^{1,z}, \theta_2, Y_T^1, M_0)$.

Treating Y_T^1 and M_0 as initial conditions for prediction, we can say that vectors $\left(\text{vec}\left(H_T^{1,w}\right)', \theta_1'\right)'$ and $\left(\text{vec}\left(H_T^{1,z}\right)', \theta_2'\right)'$ are a priori independent. The second condition says that the predictive distribution of H_{T+s}^{T+1} is factorized into a product such that the predictive distribution of $H_{T+s}^{T+1,w}$ does not depend on θ_2 and $H_T^{1,z}$, whereas the predictive distribution of $H_{T+s}^{T+1,z}$ does not depend on θ_1 and $H_T^{1,w}$. The last two conditions in definition 10 imply a factorization of the joint predictive density for W_{T+s}^{T+1} and Z_{T+s}^{T+1} :

$$\begin{aligned} p\left(W_{T+s}^{T+1}, Z_{T+s}^{T+1} | H_{T+s}^{T+1}, H_T^1, \theta, Y_T^1, M_0\right) = \\ p\left(W_{T+s}^{T+1} | Z_{T+s}^{T+1}, H_{T+s}^{T+1,w}, H_T^{1,w}, \theta_1, Y_T^1, M_0\right) \times \\ \times p\left(Z_{T+s}^{T+1} | H_{T+s}^{T+1,z}, H_T^{1,z}, \theta_2, Y_T^1, M_0\right). \end{aligned}$$

Predictive exogeneity permits conditional Bayesian forecasting of W_{T+s}^{T+1} and $H_{T+s}^{T+1,w}$ from the conditional model only, and treating Z_{T+s}^{T+1} as if it were fixed (without loss of information). We have:

$$\begin{aligned} p\left(Y_{T+s}^{T+1}, H_{T+s}^{T+1} | Y_T^1, M_0\right) = \\ \int_{\Theta \times H} \left[p\left(W_{T+s}^{T+1} | Z_{T+s}^{T+1}, H_{T+s}^{T+1,w}, H_T^1, \theta, Y_T^1, M_0\right) \times p\left(Z_{T+s}^{T+1} | H_{T+s}^{T+1,z}, H_T^1, \theta, Y_T^1, M_0\right) \times \right. \\ \left. \times p\left(H_{T+s}^{T+1} | H_T^1, \theta, Y_T^1, M_0\right) \times p\left(H_T^1, \theta | Y_T^1, M_0\right) \right] d\theta dH_T^1. \end{aligned}$$

Under predictive exogeneity:

$$\begin{aligned} p\left(Y_{T+s}^{T+1}, H_{T+s}^{T+1} | Y_T^1, M_0\right) = \\ \int_{\Theta \times H} \left[p\left(W_{T+s}^{T+1} | Z_{T+s}^{T+1}, H_{T+s}^{T+1,w}, H_T^{1,w}, \theta_1, Y_T^1, M_0\right) \times \right. \\ \times p\left(Z_{T+s}^{T+1} | H_{T+s}^{T+1,z}, H_T^{1,z}, \theta_2, Y_T^1, M_0\right) \times \\ \times p\left(H_{T+s}^{T+1,w} | H_T^{1,w}, \theta_1, Y_T^1, M_0\right) \times p\left(H_{T+s}^{T+1,z} | H_T^{1,z}, \theta_2, Y_T^1, M_0\right) \times \\ \left. \times p\left(H_T^{1,w}, \theta_1 | Y_T^1, M_0\right) \times p\left(H_T^{1,z}, \theta_2 | Y_T^1, M_0\right) \right] d\theta dH_T^1, \end{aligned}$$

thus

$$\begin{aligned} p\left(Y_{T+s}^{T+1}, H_{T+s}^{T+1} | Y_T^1, M_0\right) = \\ \int_{\Theta \times H} \left[p\left(W_{T+s}^{T+1} | Z_{T+s}^{T+1}, H_{T+s}^{T+1,w}, H_T^{1,w}, \theta_1, Y_T^1, M_0\right) \times \right. \\ \times p\left(H_{T+s}^{T+1,w} | H_T^{1,w}, \theta_1, Y_T^1, M_0\right) \times p\left(H_T^{1,w}, \theta_1 | Y_T^1, M_0\right) \times \\ \left. \times p\left(Z_{T+s}^{T+1}, H_{T+s}^{T+1,z}, H_T^{1,z}, \theta_2 | Y_T^1, M_0\right) \right] d\theta dH_T^1, \end{aligned}$$

Anna Pajor

consequently

$$\begin{aligned}
 p(Y_{T+s}^{T+1}, H_{T+s}^{T+1} | Y_T^1, M_0) = & \\
 \int_{\Theta \times H} & \left[p(W_{T+s}^{T+1} | Z_{T+s}^{T+1}, H_{T+s}^{T+1,w}, H_T^{1,w}, \theta_1, Y_T^1, M_0) \times \right. \\
 & \times p(H_{T+s}^{T+1,w} | H_T^{1,w}, \theta_1, Y_T^1, M_0) \times p(H_T^{1,w}, \theta_1 | Y_T^1, M_0) \left. \right] d\theta dH_T^{1,w} \times \\
 & \times p(Z_{T+s}^{T+1}, H_{T+s}^{T+1,z} | Y_T^1, M_0)
 \end{aligned}$$

and

$$\begin{aligned}
 p(W_{T+s}^{T+1}, H_{T+s}^{T+1,w} | Z_{T+s}^{T+1}, H_{T+s}^{T+1,z}, Y_T^1, M_0) = & \frac{p(Y_{T+s}^{T+1}, H_{T+s}^{T+1} | Y_T^1, M_0)}{p(Z_{T+s}^{T+1}, H_{T+s}^{T+1,z} | Y_T^1, M_0)} = \\
 \int_{\Theta \times H} & \left[p(W_{T+s}^{T+1} | Z_{T+s}^{T+1}, H_{T+s}^{T+1,w}, H_T^{1,w}, \theta_1, Y_T^1, M_0) \times \right. \\
 & \times p(H_{T+s}^{T+1,w} | H_T^{1,w}, \theta_1, Y_T^1, M_0) \times p(H_T^{1,w} | \theta_1, Y_T^1, M_0) \left. \right] d\theta_1 dH_T^{1,w}.
 \end{aligned}$$

Finally, we obtain

$$\begin{aligned}
 p(W_{T+s}^{T+1}, H_{T+s}^{T+1,w} | Z_{T+s}^{T+1}, Y_T^1, M_0) = & p(W_{T+s}^{T+1}, H_{T+s}^{T+1,w} | Z_{T+s}^{T+1}, H_{T+s}^{T+1,z}, Y_T^1, M_0) = \\
 \int_{\Theta \times H} & \left[p(W_{T+s}^{T+1} | Z_{T+s}^{T+1}, H_{T+s}^{T+1,w}, H_T^{1,w}, \theta_1, Y_T^1, M_0) \times \right. \\
 & \times p(H_{T+s}^{T+1,w} | H_T^{1,w}, \theta_1, Y_T^1, M_0) \times p(H_T^{1,w} | \theta_1, Y_T^1, M_0) \left. \right] d\theta_1 dH_T^{1,w}.
 \end{aligned}$$

That is, the full predictive density $p(Y_{T+s}^{T+1}, H_{T+s}^{T+1} | Y_T^1, M_0)$ is not necessary for forecasting Z_{T+s}^{T+1} . Forecasts of Z_{T+s}^{T+1} and $H_{T+s}^{T+1,z}$ can be constructed from the marginal model for Z_{T+s}^{T+1} and $H_{T+s}^{T+1,z}$ and then forecasts of W_{T+s}^{T+1} and $H_{T+s}^{T+1,w}$ can be obtained from the conditional model (without loss of relevant sample information).

Definition 11 (*strong exogeneity*) We say that z_t is strongly exogenous over the sample period $\{1, \dots, T\}$ for $f(H_T^1, \theta)$ and for prediction over $\{T+1, \dots, T+s\}$ if and only if

- (i) z_t is weakly exogenous over the sample period $\{1, \dots, T\}$ for $f(H_T^1, \theta)$,
- (ii) z_t is predictively exogenous over the forecasting period $\{T+1, \dots, T+s\}$.

It is well known that in the classical approach strong exogeneity requires both weak exogeneity and Granger non-causality (see Engle . 1983). Similarly, but not identically, is in the Bayesian approach. We can prove the following theorem:

Theorem 12 *If*

- (i) $\forall_{t \in \{1, \dots, T+s\}}$ the pair $[(\theta_1; h_t^w; w_t), (\theta_2; h_t^z; z_t)]$ operates a Bayesian one-shot cut,
 (ii) $\forall_{t \in \{1, \dots, T+s\}}$ w_t does not cause z_t given $H_t^{1,z}$ and θ_2 ,

then z_t is strongly exogenous over the period $\{1, \dots, T\}$ for $f_0(H_T^{1,w}, \theta_1)$ and for prediction over $\{T+1, \dots, T+s\}$.

If there exists a Bayesian sequential cut and non-causality given $H_t^{1,z}$ and θ_2 for $t \in \{1, \dots, T+s\}$, we can write $p(Z_{T+s}^{T+1}, H_{T+s}^{T+1,z} | H_T^{1,z}, \theta_2, Y_T^1, M_0) = p(Z_{T+s}^{T+1}, H_{T+s}^{T+1,z} | Z_T^1, H_T^{1,z}, \theta_2, M_0)$.

Furthermore, we have:

$$p(H_T^{1,z}, \theta_2 | Y_T^1, M_0) \propto p(\theta_2 | M_0) \prod_{t=1}^T p(z_t | H_t^{1,z}, Z_{t-1}^1, \theta_2, M_0) p(h_t^z | H_t^{1,z}, \theta_2, M_0),$$

thus

$$p(H_T^{1,z}, \theta_2 | Y_T^1, M_0) = p(H_T^{1,z}, \theta_2 | Z_T^1, M_0),$$

and, consequently,

$$p(Z_{T+s}^{T+1}, H_{T+s}^{T+1,z} | Y_T^1, M_0) = \int_{\Theta \times H^z} \left[p(Z_{T+s}^{T+1}, H_{T+s}^{T+1,z} | Z_T^1, H_T^{1,z}, \theta_2, M_0) p(H_T^{1,z}, \theta_2 | Z_T^1, M_0) \right] d\theta_2 dH_T^{1,z}.$$

It follows that the marginal model suffices (without a loss of information) for predictive inference on z_t and h_t^z . On the other hand, strong exogeneity of z_t permits to forecast w_t and $H_{T+s}^{T+1,w}$ from the conditional model, given the forecast of z_t from the marginal model.

5 Exogeneity in the VECM-SV model

Let us consider a simple vector autoregressive (VAR) model with multivariate stochastic volatility, which can be written in the vector error correction mechanism (VECM) form. The VECM form for the 3×1 vector time series process u_t is:

$$y_t - \delta = R(y_{t-1} - \delta) + \alpha\beta' u_{t-1} + \xi_t, \quad t = 1, 2, \dots, T \quad (15)$$

where $y_t = \Delta u_t = u_t - u_{t-1}$, R is a 3×3 matrix of the autoregressive coefficients. It is assumed that u_t is integrated of order one (thus $y_t \sim I(0)$) and there exists

Anna Pajor

one cointegrating relation ($r = 1$) with known cointegrating vector β , so that $\beta'u_{t-1} = ECM_{t-1}$ is the error correction mechanism. More specifically:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} - \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \left(\begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} - \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} \right) + \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} ECM_{t-1} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}, \quad t = 1, \dots, T.$$

The 3×1 vector of errors, ξ_t , follows a trivariate SV process. We assume that conditionally on the latent variable vector, h_t , and the parameter vector, θ , ξ_t follows a trivariate Gaussian distribution with mean vector $0_{[3 \times 1]}$ and covariance matrix Σ_t , i.e. $\xi_t | h_t, \theta \sim N(0_{[3 \times 1]}, \Sigma_t)$, $t = 1, \dots, T$. We employ the specification of the conditional covariance matrix as in Tsay (2002). Thus, the Cholesky decomposition is used:

$$\Sigma_t = L_t G_t L_t', \quad (16)$$

where L_t is a lower triangular matrix with unitary diagonal elements, and G_t is a diagonal matrix with positive diagonal elements:

$$L_t = \begin{bmatrix} 1 & 0 & 0 \\ q_{21,t} & 1 & 0 \\ q_{31,t} & q_{32,t} & 1 \end{bmatrix}, \quad G_t = \begin{bmatrix} q_{11,t} & 0 & 0 \\ 0 & q_{22,t} & 0 \\ 0 & 0 & q_{33,t} \end{bmatrix}.$$

Hence, we have

$$\Sigma_t = \begin{bmatrix} q_{11,t} & q_{11,t}q_{21,t} & q_{11,t}q_{31,t} \\ q_{11,t}q_{21,t} & q_{11,t}q_{21,t}^2 + q_{22,t} & q_{11,t}q_{21,t}q_{31,t} + q_{22,t}q_{32,t} \\ q_{11,t}q_{31,t} & q_{11,t}q_{21,t}q_{31,t} + q_{22,t}q_{32,t} & q_{11,t}q_{31,t}^2 + q_{22,t}q_{32,t}^2 + q_{33,t} \end{bmatrix}, \quad (17)$$

where $\{q_{ij,t}\}$, and $\{\ln q_{jj,t}\}$ ($i, j = 1, 2, 3, i > j$), as in the univariate SV, are standard univariate autoregressive processes of order one, namely:

$$\ln q_{jj,t} - \gamma_{jj} = \phi_{jj} (\ln q_{jj,t-1} - \gamma_{jj}) + \sigma_{jj} \eta_{jj,t}, \quad j = 1, 2, 3,$$

$$q_{ij,t} - \gamma_{ij} = \phi_{ij} (q_{ij,t-1} - \gamma_{ij}) + \sigma_{ij} \eta_{ij,t}, \quad i > j,$$

$$\eta_t = (\eta_{11,t}, \eta_{22,t}, \eta_{33,t}, \eta_{21,t}, \eta_{31,t}, \eta_{32,t})', \quad \{\eta_t\} \sim iid N_6(0_{[6 \times 1]}, I_6), \quad t \in \{1, \dots, T\},$$

$$h_t = (q_{11,t}, q_{22,t}, q_{33,t}, q_{21,t}, q_{31,t}, q_{32,t})'.$$

Note that positive definiteness of Σ_t is achieved by modelling $\ln q_{jj,t}$ instead of $q_{jj,t}$ (Σ_t is positive defined if $q_{jj,t} > 0$ for $j = 1, 2, 3$). If $|\phi_{ij}| < 1$ ($i, j = 1, 2, 3; i \geq j$)

then $\{\ln q_{jj,t}\}$ and $\{q_{ij,t}\}$ are covariance stationary and their marginal distributions are Normal with mean γ_{ij} and variance $\sigma_{ij}^2 / (1 - \phi_{ij}^2)$. It can easily be shown that if the absolute values of ϕ_{ij} are less than one, the SV process is a white noise (see Pajor 2005a).

Thus, the conditional distribution of y_t (given the past of the process Y_{t-1} , the parameters and the latent variable vector, h_t) is a trivariate Normal with mean vector $\mu_t = \delta + R(y_{t-1} + \delta) + \alpha ECM_{t-1}$ and covariance matrix Σ_t :

$$(y_{1,t}, y_{2,t}, y_{3,t})' | \delta, \alpha, R, h_t, Y_{t-1} \sim N_3(\mu_t, \Sigma_t).$$

We can partition y_t and Σ_t , conformably, into:

$$y_t = \begin{bmatrix} z_t \\ w_t \end{bmatrix} \text{ and } \Sigma_t = \begin{bmatrix} \Sigma_{11,t} & \Sigma_{12,t} \\ \Sigma_{21,t} & \Sigma_{22,t} \end{bmatrix}, \quad (18)$$

where z_t is a scalar: $z_t = y_{1,t}$, and w_t has two elements: $w_t = (y_{2,t}, y_{3,t})'$. After partitioning, Equation (15) becomes:

$$\begin{bmatrix} z_t \\ w_t \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_{-1} \end{bmatrix} + \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \left(\begin{bmatrix} z_{t-1} \\ w_{t-1} \end{bmatrix} - \begin{bmatrix} \delta_1 \\ \delta_{-1} \end{bmatrix} \right) + \begin{bmatrix} \alpha_1 \\ \alpha_{-1} \end{bmatrix} ECM_{t-1} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \quad (19)$$

where $\delta_{-1} = (\delta_2, \delta_3)$, $\alpha_{-1} = (\alpha_2, \alpha_3)$, $\varepsilon_{2,t} = (\xi_{2,t}, \xi_{3,t})$. The VECM form in (19) can be reparameterized as the conditional and marginal distributions in (12) and (13):

$$z_t - \delta_1 = R_{11}(z_{t-1} - \delta_1) + R_{12}(w_{t-1} - \delta_{-1}) + \alpha_1 ECM_{t-1} + \varepsilon_{1,t} \quad (20)$$

$$w_t - \delta_{-1} = (R_{22} - D_t R_{12})(w_{t-1} - \delta_{-1}) + (R_{21} - D_t R_{11})(z_{t-1} - \delta_1) + (\alpha_{-1} - D_t \alpha_1) ECM_{t-1} + D_t(z_t - \delta_1) + (\varepsilon_{2,t} - D_t \varepsilon_{1,t}), \quad (21)$$

where

$$\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} - D_t \varepsilon_{1,t} \end{bmatrix} | h_t \sim N_3 \left(0, \begin{bmatrix} \Sigma_{11,t} & 0 \\ 0 & \Sigma_{22.1,t} \end{bmatrix} \right),$$

$$\Sigma_{22.1,t} = \Sigma_{22,t} - \Sigma_{21,t} \Sigma_{11,t}^{-1} \Sigma_{12,t},$$

$$D_t = \Sigma_{21,t} \Sigma_{11,t}^{-1} = \begin{bmatrix} q_{21,t} \\ q_{31,t} \end{bmatrix}, \quad \Sigma_{11,t} = [q_{11,t}], \quad \Sigma_{22,t} = \begin{bmatrix} q_{22,t} & q_{22,t} q_{32,t} \\ q_{22,t} q_{32,t} & q_{33,t} + q_{22,t} q_{32,t}^2 \end{bmatrix}.$$

The parameters and latent variables of the conditional model (21) are

$$\begin{aligned} \left(\text{vec} \left(H_T^{1,w} \right)', \theta_1' \right)' &= \left((R_{21} - D_1 R_{11})', \dots, (R_{21} - D_T R_{11})', \right. \\ &\quad \text{vec} (R_{22} - D_1 R_{12})', \dots, \text{vec} (R_{22} - D_T R_{12})', \\ &\quad (\alpha_{-1} - D_1 \alpha_1)', \dots, (\alpha_{-1} - D_T \alpha_1)', D_1', \dots, D_T', \\ &\quad \left. \text{vec} (\Sigma_{22.1,1})', \dots, \text{vec} (\Sigma_{22.1,T})', \delta_1, \delta_{-1}', \theta_1^{h,w'} \right)', \end{aligned}$$

Anna Pajor

and those of the marginal model (20) are

$$\left(\text{vec} \left(H_T^{1,z} \right)', \theta_2' \right)' = \left(\Sigma_{11,1}, \dots, \Sigma_{11,T}, \delta_1, \delta'_{-1}, R_{11}, R_{12}, \alpha_1, \theta_2^{h,z'} \right)',$$

where $\theta_t^{h,w}$ and $\theta_t^{h,z}$ are vectors of parameters in the equations for $h_t^{1,w}$ and $h_t^{1,z}$, respectively.

It must be noticed that the conditional and marginal models contain two completely separate sets of latent variables. Thus, under some additional assumptions, we can make efficient inference about the parameters and latent variables in the conditional model, and the marginal model can be neglected due to no information loss.

Lemma 13 *If*

$$(i) \quad R_{11} = 0, R_{12} = 0, \alpha_1 = 0, \delta_1 = 0,$$

$$(ii) \quad \left(\delta'_{-1}, R'_{21}, \text{vec} (R_{22})', \alpha'_{-1}, \theta_1^{h,w'} \right)' \text{ and } \theta_1^{h,z} \text{ are a priori independent,}$$

or

$$(i) \quad \delta_1 = 0, \forall t \in \{1, \dots, T\} \quad D_t = 0,$$

$$(ii) \quad \left(\delta'_{-1}, R'_{21}, \text{vec} (R_{22})', \alpha'_{-1}, \theta_1^{h,w'} \right)' \text{ and } \left(R_{11}, R_{12}, \alpha_1, \theta_2^{h,z'} \right)' \text{ are a priori independent,}$$

then z_t is weakly exogenous for $f_0 \left(H_T^{1,w}, \theta_1 \right)$.

The latent variables in the conditional process do not appear in the marginal process for z_t . Furthermore, the latent variables characterizing the marginal $(H_T^{1,z})$ and the conditional $(H_T^{1,w})$ processes are variation free. The prior independence of the parameters θ_1 and θ_2 ensures that $\left(\text{vec} (H_T^{1,w})', \theta_1' \right)'$ and $\left(\text{vec} (H_T^{1,z})', \theta_2' \right)'$ are a priori independent. Thus, assumptions (i) and (ii) in lemma 13 are sufficient for the Bayesian sequential cut and, consequently, for weak exogeneity of z_t for $\left(\text{vec} (H_T^{1,w})', \theta_1' \right)'$.

Conditions (i) and (ii) in lemma 13 are sufficient but not necessary for weak exogeneity. In lemma 13 we present an alternative set of sufficient conditions that guarantee weak exogeneity of z_t for $\left(\text{vec} (H_T^{1,w})', \theta_1' \right)' \left(\theta_1, H_T^{1,w} \right)$: $D_t = 0$ ($D_t \neq 0$ implying contemporaneous conditional correlation between w_t and z_t), $t = 1, \dots, T$, $\delta_1 = 0$, and the prior independence of $\left(\delta'_{-1}, R'_{21}, \text{vec} (R_{22})', \alpha'_{-1}, \theta_1^{h,w'} \right)'$ and $\left(R_{11}, R_{12}, \alpha_1, \theta_1^{h,z'} \right)'$. It is worth stressing that the Bayesian concept of weak exogeneity requires the prior independence between θ_1 and θ_2 , which in the VECM-SV

model implies the prior independence of $(\text{vec}(H_T^{1,w})', \theta_1')'$ and $(\text{vec}(H_T^{1,z})', \theta_2')'$, and naturally leads to the posterior independence of $(\text{vec}(H_T^{1,w})', \theta_1')'$ and $(\text{vec}(H_T^{1,z})', \theta_2')'$.

Once conditions (i) and (ii) in lemma 13 are satisfied, the conditional model of w_t given z_t assumes the form:

$$\begin{aligned} \begin{bmatrix} y_{2,t} \\ y_{3,t} \end{bmatrix} - \begin{bmatrix} \delta_2 \\ \delta_3 \end{bmatrix} &= \begin{bmatrix} r_{22} & r_{23} \\ r_{32} & r_{33} \end{bmatrix} \left(\begin{bmatrix} y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} - \begin{bmatrix} \delta_2 \\ \delta_3 \end{bmatrix} \right) + \begin{bmatrix} q_{21,t} \\ q_{31,t} \end{bmatrix} y_{1,t} + \\ &+ \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix} y_{1,t-1} + \begin{bmatrix} \alpha_2 \\ \alpha_3 \end{bmatrix} ECM_{t-1} + \begin{bmatrix} \xi_{2,t} \\ \xi_{3,t} \end{bmatrix}, \end{aligned} \quad (22)$$

where $(\xi_{2,t}, \xi_{3,t})'$ is the bivariate SV process, $t = 1, 2, \dots, T$.

When conditions (iii) and (iv) in lemma 13 hold, Equations (20) and (21) can be written as:

$$z_t = R_{11}z_{t-1} + R_{12}(w_{t-1} - \delta_{-1}) + \alpha_1 ECM_{t-1} + \varepsilon_{1,t}, \quad (23)$$

$$w_t - \delta_{-1} = R_{22}(w_{t-1} - \delta_{-1}) + R_{21}z_{t-1} + \alpha_{-1} ECM_{t-1} + \varepsilon_{2,t}, \quad (24)$$

Note that conditions (i) and (ii) in lemma 13 remain the same for strong exogeneity of z_t for $(\text{vec}(H_T^{1,w})', \theta_1')'$. However, conditions (iii) and (iv) are not sufficient for strong exogeneity of z_t for $(\text{vec}(H_T^{1,w})', \theta_1')'$. To ensure it, we must assume, in addition, that $\alpha_1 = 0$ and $R_{12} = 0$.

Lemma 14 *If*

$$(i) \delta_1 = 0, \forall_{t \in \{1, \dots, T+s\}} D_t = 0,$$

$$(ii) \alpha_1 = 0, R_{12} = 0,$$

$$(iii) \left(\delta'_{-1}, R'_{21}, \text{vec}(R_{22})', \alpha'_{-1}, \theta_1^{h,w} \right)' \text{ and } \left(R_{11}, \theta_2^{h,z} \right)' \text{ are a priori independent,}$$

then z_t is strongly exogenous over the period $\{1, \dots, T\}$ for $f_0(H_T^{1,w}, \theta_1)$ and for prediction over $\{T+1, \dots, T+s\}$.

It must be emphasized that these strong exogeneity conditions are only sufficient, and not necessary. It also seems that they are rather restrictive. In practice, the conditions for exogeneity should be tested. Unfortunately, tests of weak or strong exogeneity hypothesis require the joint model to be specified.

Anna Pajor

6 An empirical illustration

For an empirical illustration, let us consider two average daily main Polish official exchange rates: the zloty (PLN) values of the US dollar, and the zloty (PLN) values of the euro, over the period from January 3, 2005 to September 30, 2011 (downloaded from the website of the National Bank of Poland). The dataset of the daily logarithmic growth rates (expressed in percentage points), consists of 1662 observations (for each series). The first observation is used to construct initial conditions, thus $T = 1661$. It is well known that the Polish official exchange rates are linked to the exchange rates of the international Forex market. Thus, while building time-series models for the two Polish exchange rates (EUR/PLN ($x_{2,t}$) and USD/PLN ($x_{3,t}$)), we introduce some extra variable from the international Forex market – the euro value of the US dollar, EUR/USD ($x_{1,t}$, downloaded from <http://stooq.com>). The relationship: (EUR/PLN)/(USD/PLN) \approx EUR/USD is introduced by assuming that this relation (in log terms) is a cointegration equation in the sense of Engle and Granger (1987). This yields the cointegrating vector $(1, -1, -1)$ and the long-run (equilibrium) relationship $\ln x_{2,t} - \ln x_{3,t} = \ln x_{1,t}$. The assumption that $\ln x_{1,t}$, $\ln x_{2,t}$ and $\ln x_{3,t}$ are cointegrated has been checked informally using the Phillips and Perron test applied to the series $ECM_t = 100 (\ln x_{2,t} - \ln x_{3,t} - \ln x_{1,t})$. The Phillips and Perron statistic (-39.96) supports covariance stationarity.

The vector of growth rates, $y_t = (y_{1,t}, y_{2,t}, y_{3,t})$, where $y_{j,t} = 100 (\ln x_{j,t} - \ln x_{j,t-1})$ ($j = 1, 2, 3$), is modelled within the VECM framework defined in (15):

$$y_t - \delta = R(y_{t-1} - \delta) + \alpha ECM_{t-1} + \xi_t, \quad t = 1, \dots, T,$$

where $\{\xi_t\}$ is the trivariate SV process, and T denotes the number of the observations used in the estimation. Note that in this model, the EUR/USD exchange rate is treated as an endogenous variable and is modelled jointly with the remaining variables. An alternative method of introducing any variables that would be related to the international Forex market is proposed by Osiewalski and Pipień (2004). However, they consider MGARCH structures with no latent processes, and also they assume exogeneity of the EUR/USD exchange rate without testing it.

To obtain the Bayesian model we need to specify a prior distribution of the parameters. For all elements of δ , α and R we assume the multivariate standard Normal prior $N_{15}(0, I)$, truncated by the restriction that all eigenvalues of R lie inside the unit circle. For the remaining parameters we assume the following prior distributions (see Pajor 2006): $(\gamma_{ij}, \phi_{ij})' \sim N_2(0, 10^2 \cdot I) \cdot I_{(-1,1)}(\phi_{ij})$, $\sigma_{ij}^{-2} \sim G(1, 0.005)$, $\ln q_{ii,0} \sim N_1(0, 10^2)$ for $i, j \in \{1, 2, 3\}$ and $i \geq j$; $q_{ij,0} \sim N_1(0, 10^2)$ for $i, j \in \{1, 2, 3\}$, $i > j$, where $N_n(a, B)$ denotes the n -variate Normal distribution with mean vector a and covariance matrix B , $G(a, b)$ denotes the Gamma distribution with shape parameter a and precision parameter b , the mean being $\frac{a}{b}$, and $I_{(-1,1)}(\cdot)$ is the indicator function of the interval $(-1, 1)$. The prior distributions used are relatively non-informative.

The data are plotted in Figure 1, 2 and 3. It can be seen from the graphs that the

growth rates seem to be centered around zero, with time-varying volatility and the presence of outliers. The daily exchange rates are more volatile in the period of the financial crisis 2008-2009. However, the volatility of the official exchange rates of the National Bank of Poland is higher than that of the EUR/USD exchange rates. Summary statistics for the time series are shown in Table 1. The arithmetic means of the growth rates of the official Polish exchange rates are positive, but the standard deviations are relatively high. We see that the kurtosis (much larger than 3) suggests that distributions of the growth series are leptokurtic. As expected, the EUR/PLN and USD/PLN exchange rates are positively correlated.

All presented results were obtained with the use of the Metropolis and Hastings algorithm within the Gibbs sampler using 10^6 iterations after $5 \cdot 10^4$ burn-in Gibbs steps (see Gamerman 1997, Tsay 2002 and Pajor 2005a, 2007 for details). The Bayes factors were calculated using the Newton and Raftery method (see Newton and Raftery, 1994). Although the harmonic mean estimator is sensitive to outliers with small likelihood values, it is stable enough to provide a rough approximation of the Bayes factor of the contrasting hypotheses.

Table 1: Summary statistics of $y_{i,t}$

	USD/PLN	EUR/PLN	EUR/USD
mean	0.0047	0.0050	-0.0001
std dev.	0.0280	0.0182	0.0166
skewness	0.4802	0.2530	0.1858
kurtosis	7.6101	8.8404	6.1284
minimum	-6.7485	-4.5895	-3.0031
maximum	6.2677	4.1467	4.6208
correlations			
USD/PLN	1	0.7502	-0.4431
EUR/PLN		1	-0.1641
EUR/USD			1

Table 2 reports the posterior means and standard deviations (in parentheses) of all parameters of the VECM-SV model. The main posterior characteristics of matrix R indicate that the daily growth rates of EUR/PLN ($y_{2,t}$) and USD/PLN ($y_{3,t}$) significantly depend on the past of the daily growth rates of EUR/USD ($y_{1,t-1}$). In other words, the main Polish exchange rates are affected by the past movements of the Forex market. Obviously, the Forex market is insignificantly affected by the past movement of the Polish market.

The posterior means of α_i ($i = 1, 2, 3$) are equal to -0.090 , -0.355 and 0.611 , respectively. The signs of the posterior means of α_i , for $i = 2, 3$, are correct with interpretation in terms of the error correction mechanism. The posterior standard deviations of α_1 , α_2 and α_3 are equal to 0.054 , 0.044 and 0.057 , respectively. These results indicate that the equilibrium correction error ECM_{t-1} "significantly"

Anna Pajor

influences only on $(y_{2,t}, y_{3,t})'$.

Figure 1: Daily growth rates of USD/PLN (January 3, 2005 – September 30, 2011)

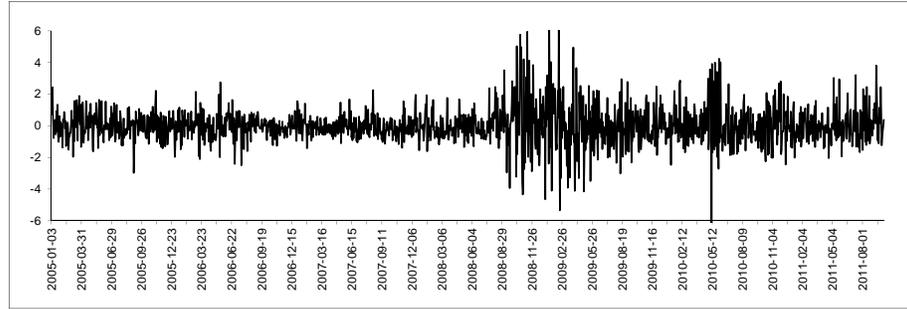


Figure 2: Daily growth rates of EUR/PLN (January 3, 2005 – September 30, 2011)

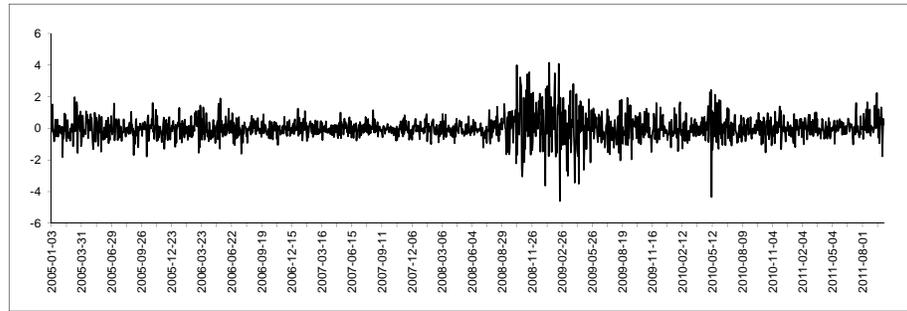


Figure 3: Daily growth rates of EUR/USD (January 3, 2005 – September 30, 2011)

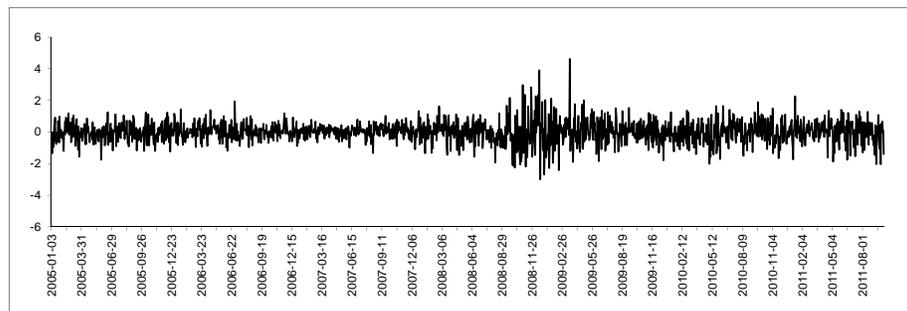
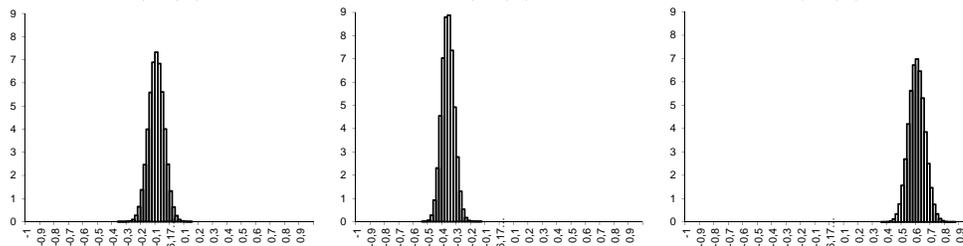


Table 2: The posterior means and standard deviations (in parentheses) of the parameters of the trivariate VECM-SV model

δ_1	δ_2	δ_3	r_{11}	r_{12}	r_{13}	r_{21}	r_{22}	r_{23}	r_{31}
0.0172 (0.0127)	-0.0213 (0.012)	-0.0410 (0.016)	-0.0938 (0.049)	-0.0288 (0.0396)	-0.0195 (0.0294)	-0.3031 (0.0375)	-0.0602 (0.0286)	0.0323 (0.0206)	-0.3043 (0.0491)
r_{32}	r_{33}	γ_{11}	ϕ_{11}	σ_{11}^2	γ_{22}	ϕ_{22}	σ_{22}^2	γ_{33}	ϕ_{33}
-0.0427 (0.0381)	0.0299 (0.0277)	-0.5204 (3.1028)	0.9969 (0.0027)	0.0046 (0.0015)	-1.2239 (1.5352)	0.9910 (0.0049)	0.0238 (0.0067)	-1.4775 (3.7415)	0.9966 (0.0032)
σ_{33}^2	γ_{21}	ϕ_{21}	σ_{21}^2	γ_{31}	ϕ_{31}	σ_{31}^2	γ_{32}	ϕ_{32}	σ_{32}^2
0.0127 (0.0048)	-0.0302 (0.0201)	-0.1935 (0.2229)	0.0171 (0.0073)	-0.3525 (0.0244)	-0.0784 (0.2535)	0.0174 (0.0105)	1.0592 (0.0177)	0.2656 (0.4336)	0.0138 (0.009)
$\ln q_{11,0}$	$\ln q_{22,0}$	$\ln q_{33,0}$	$q_{21,0}$	$q_{31,0}$	$q_{32,0}$	α_1	α_2	α_3	
-1.1368 (0.321)	-0.9270 (0.4668)	-2.6033 (0.482)	4.5891 (7.7703)	-3.0127 (8.6023)	0.9122 (4.5254)	-0.0902 (0.0539)	-0.3550 (0.0436)	0.6112 (0.0569)	

Figure 4: Histograms of the posterior distributions for $\alpha_i, i = 1, 2, 3$



The posterior mean and standard deviation of γ_{31} indicate that matrix D_t (more precisely, $q_{31,t}$) is significantly different from 0. The formal Bayesian testing (not presented here) would lead to rejection of the null hypothesis that $D_t = 0$ for $t = 1, 2, \dots, T$.

Table 3 reports the posterior means and standard deviations (in parentheses) of the parameters of the conditional model (22). The main conclusions remain the same.

Suppose that some function of $(\text{vec}(H_T^{1,w}), \theta_1)'$ is of interest. To test the sufficient conditions (presented in lemma 13, i-ii) for the strong exogeneity of the EUR/USD exchange rate we use the Lindley type test (based on the highest posterior density region, see Box and Tiao, 1973, Osiewalski and Steel, 1993) and the Bayes factor (BF, in favor of the model with exogeneity of the EUR/USD exchange rate). We set the null and alternative hypotheses as: $H_0: \theta_0 = 0$ and $H_1: \theta_0 \neq 0$, where θ_0 is the $k \times 1$ vector ($\theta_0 = (\delta_1, \alpha_1, r_{11}, r_{12}, r_{13})'$, $k = 5$). Then, a posterior, the

 Anna Pajor

Table 3: The posterior means and standard deviations (in parentheses) of the parameters of the conditional bivariate VECM-SV model

δ_1	δ_2	r_{11}	r_{12}	r_{21}	r_{22}	γ_{22}	ϕ_{22}	σ_{22}^2
-0.0156 (0.0111)	-0.0296 (0.0132)	-0.0618 (0.0287)	0.0315 (0.0207)	-0.0524 (0.0359)	0.0227 (0.026)	-1.2258 (1.5365)	0.9910 (0.0049)	0.0236 (0.0068)
γ_{33}	ϕ_{33}	σ_{33}^2	γ_{32}	ϕ_{32}	σ_{32}^2	γ_{21}	ϕ_{21}	σ_{21}^2
-1.4818 (3.7611)	0.9966 (0.0032)	0.0125 (0.0048)	1.0591 (0.0319)	0.2581 (0.4296)	0.0145 (0.0092)	-0.0305 (0.0196)	-0.1886 (0.2347)	0.0171 (0.0073)
γ_{31}	ϕ_{31}	σ_{31}^2	α_2	α_3	β_2	β_3	$\ln g_{22,0}$	$\ln g_{33,0}$
-0.3526 (0.0238)	-0.0951 (0.2577)	0.0174 (0.0104)	-0.3581 (0.0438)	0.5802 (0.054)	-0.3071 (0.0377)	-0.3368 (0.0464)	-0.8795 (0.4621)	-2.5421 (0.4893)
$q_{32,0}$	$q_{21,0}$	$q_{31,0}$						
0.8413 (4.5066)	2.8599 (7.4889)	-3.0597 (8.3435)						

quadratic form: $F(\theta_0) = [\theta_0 - E(\theta_0|y)]' V^{-1}(\theta_0|y) [\theta_0 - E(\theta_0|y)]$, where $E(\theta_0|y)$ and $V(\theta_0|y)$ are respectively the vector of posterior means of θ_0 and the posterior covariance matrix (in the notation we omit the initial conditions), is approximately ($T \rightarrow \infty$) χ^2 distributed with k degrees of freedom. We obtain $F(0) = 8.943$, and $p(F(\theta_0) > F(0)|y) \approx 0.11$. Thus, the sufficient conditions of the strong exogeneity are not rejected by the data (see Figure 5). Also, the decimal logarithm of the Bayes factor in favor of the hypothesis that $\theta_0 = 0$ (equal to 4.92) provides the same conclusion.

A different result is obtained when we examine whether $\theta_{01} = 0$, where $\theta_{01} = (\delta_1, r_{11}, r_{12}, r_{13}, \gamma_{21}, \phi_{21}, \sigma_{21}, \gamma_{31}, \phi_{31}, \sigma_{31})'$ (see conditions (iii) and (iv) in lemma 13). For the Lindley type test $F(0) = 34.372$, that is $p(F(\theta_{01}) > F(0)|y) \approx 0.00016$. The decimal logarithm of the Bayes factor in favor of the hypothesis that $\theta_{01} = 0$ is equal to -172.59 . Thus the formal Bayesian testing leads to rejection of the restriction $\theta_{01} = 0$.

It is worth mentioning that the test result of the hypothesis that $\theta_0 = 0$ depends on the source of data (the euro values of the US dollar). For example, when we use the data downloaded from www.federalreserve.gov/releases/, the sufficient conditions of the exogeneity are rejected by the data: $F(0) = 17.913$, and $p(F(\theta_0) > F(0)|y) \approx 0.0031$. It is very important to stress that this test result does not exclude the possibility of exogeneity.

For our data the inference about the conditional correlation coefficients and individual volatility of each time series (measured by the conditional standard deviation) is very similar in the full (complete) model (15) and in the conditional model (22). The time plots of conditional correlation between the growth of rates of USD/PLN and EUR/PLN (for each $t = 1, 2, \dots, T$; $T = 1661$) are presented in Figure 6, where the upper line represents the posterior mean plus the standard deviation, and the lower one - the posterior mean minus the standard deviation. It can be seen from the graph

Figure 5: Histogram of the posterior distribution for $F(\theta_0)$. The black square represents $F(0)$

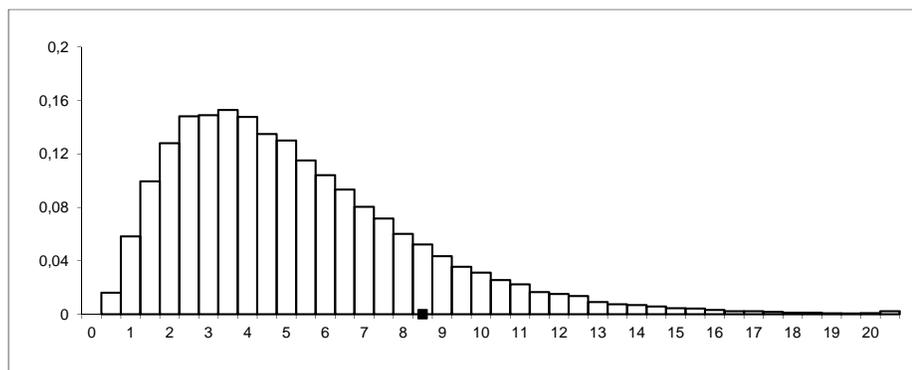
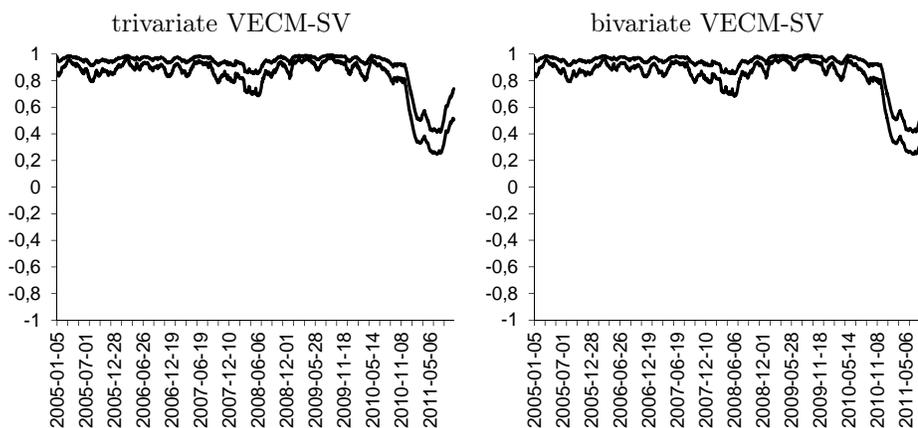


Figure 6: Conditional correlation coefficients between daily growth rates of USD/PLN and EUR/PLN (posterior mean \pm standard deviation)



that these models lead to very similar inferences about the dynamics of the conditional correlations between the growth of rates of the USD/PLN and EUR/PLN. The models clearly exhibit volatility clustering phenomenon and produce volatility peaks at the same time (see Figure 7 and 8). These results are consistent with the test result for exogeneity of the EUR/USD exchange rate.

Anna Pajor

Figure 7: Posterior means of the conditional standard deviations for USD/PLN (posterior mean \pm standard deviation)

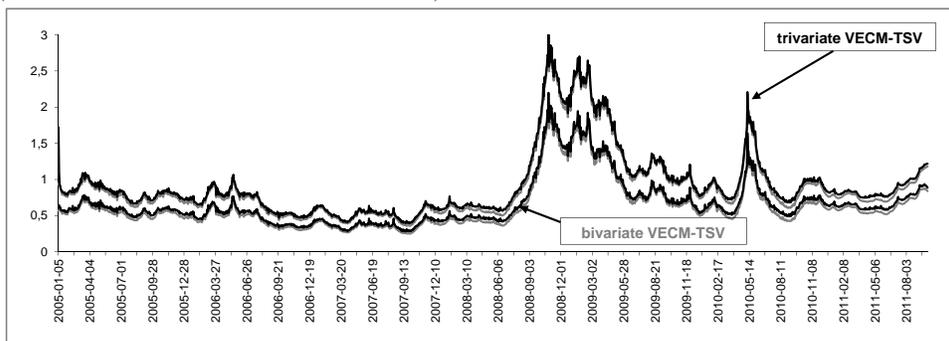
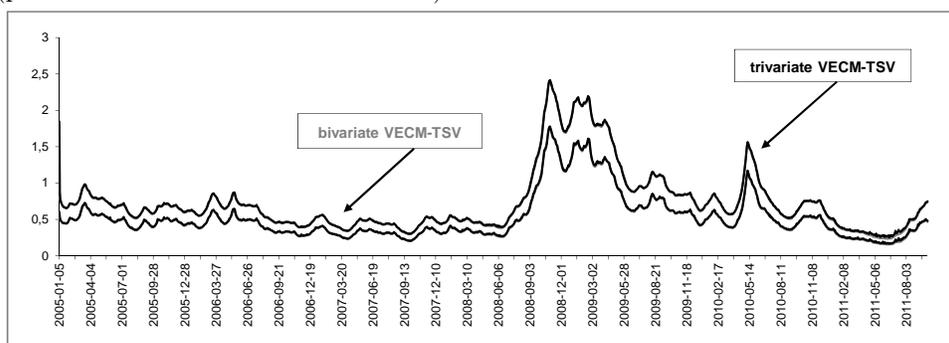


Figure 8: Posterior means of the conditional standard deviations for EUR/PLN (posterior mean \pm standard deviation)



7 Conclusions

In the present paper the concept of exogeneity in models with latent variables is presented. The Bayesian exogeneity concept as defined here validates the use of the conditional model (as a reduction of the complete model) for inference about the parameters of interest and some latent variables. For the purpose of conditional forecasting we define predictive and strong exogeneity. The theory is applied to the VECM-SV model for the main official Polish exchange rates: USD/PLN and EUR/PLN, with the EUR/USD exchange rates from the international Forex market. The sufficient conditions of strong exogeneity of the EUR/USD exchange rates are

not rejected by the data. The same main posterior inferences in the case of the joint (complete) model and in the conditional model alone, confirm the results of formal Bayesian hypothesis testing.

8 Acknowledgements

The author would like to thank Jacek Osiewalski for helpful comments and discussions, and Janusz Jaworski for language verification of the manuscript. Research supported by a grant from Cracow University of Economics.

References

- [1] Box G.E.P., Tiao G.C., (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley Publishing Company, Massachusetts.
- [2] De Luna X., P. Johansson P., (2006), Exogeneity in Structural Equation Models, *Journal of Econometrics* 132, 527-543.
- [3] Engle R.F., Granger C.W.J., (1987), Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica* 55, 251-276.
- [4] Engle R.F., Hendry D.F., Richard J.F., (1983), Exogeneity, *Econometrica* 51, 277-304.
- [5] Ericsson N.R., Hendry D.F., Mizon G.E., (1998), Exogeneity, Cointegration, and Economic Policy Analysis, *Journal of Business and Economic Statistics* 6, 370-387.
- [6] Florens J.P., Mouchart M., (1977), Reduction of Bayesian experiments, *CORE discussion paper* no. 7737, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
- [7] Florens J.P., Mouchart M., (1980), Initial and Sequential Reduction of Bayesian Experiments, *CORE discussion paper* no. 8015, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
- [8] Florens J.P., Mouchart M., (1982), A Note on Noncausality, *Econometrica* 50 (3), 583-591.
- [9] Florens J.P., Mouchart M., (1985), Conditioning in Dynamic Models, *Journal of Time Series Analysis* 53, 15-35.
- [10] Florens J.P., Mouchart M., Rolin J.M., (1990), *Elements of Bayesian Statistics*, Marcel Dekker, Inc, New York and Basel.

Anna Pajor

- [11] Gamerman D., (1997), *Markov Chain Monte Carlo. Statistic Simulation for Bayesian Inference*, Chapman and Hall, London.
- [12] Granger C.W.J., (1969), Investigating Causal Relation by Econometric and Cross-sectional Method, *Econometrica* 37, 424-438.
- [13] Hendry D.F., Richard J.F., (1982), On the Formulation of Empirical Models in Dynamic Econometrics, *Journal of Econometrics* 20, 3-33.
- [14] Hendry D.F., Richard J.F., (1983), The Econometric Analysis of Economic Time Series, *International Statistical Review* 51, 111-163.
- [15] Mouchart M., Russo F., Wunsch G., (2007), Causality, Structural Modelling, and Exogeneity, *IAP Statistics Network*, 1-26.
- [16] Newton M.A., Raftery A.E., (1994), Approximate Bayesian Inference by the Weighted Likelihood Bootstrap [with discussion], *Journal of the Royal Statistical Society B*, 56, 3-48
- [17] O'Hagan A., (1994), *Bayesian Inference*, Halsted Press, New York.
- [18] Osiewalski J., Pipień M., (2004), Bayesian Comparison of Bivariate GARCH Processes. The Role of the Conditional Mean Specification, Chapter 7, [in:] *Contributions to Economic Analysis 269, New Directions in Macromodelling*, [ed.:] A. Welfe, Elsevier, Amsterdam, 173-196.
- [19] Osiewalski J., Steel M.F.J., (1993), Una Perspectiva Bayesiana en Selección de Modelos, *Cuadernos Economicos* 55, 327-351, (typescript: A Bayesian perspective on model selection).
- [20] Osiewalski J., Steel M.F.J., (1996), A Bayesian Analysis of Exogeneity in Models Pooling Time-Series and Cross-section Data, *Journal of Statistical Planning and Inference* 50, 187-206.
- [21] Pajor A., (2005a), Bayesian Analysis of Stochastic Volatility Model and Portfolio Allocation, *Acta Universitatis Lodzensis - Folia Oeconomica* 192, 229-249.
- [22] Pajor A., (2005b), Bayesian comparison of bivariate SV models for two related time series, *Acta Universitatis Lodzensis - Folia Oeconomica* 190, 177-196.
- [23] Pajor A., (2006), Bayesian VECM-SV Models for the Main Polish Exchange Rates, [in:], *32-nd International Conference, MACROMODELS'05*, [ed.:] W. Welfe, A. Welfe, Łódź, 135-154.
- [24] Pajor A., (2007), VECM-TSV Models for Two Polish Official Exchange Rates, [in:] *Financial Markets: Principles of Modeling, Forecasting and Decision-Making, FindEcon Monograph Series: Advance in Financial Market Analysis*, No. 2, [ed.:] W. Milo, P. Wdowiński, Łódź University Press, Łódź, 49-66.

A Bayesian Analysis of Exogeneity . . .

- [25] Pajor A., (2008), A Bayesian Analysis of Weak Exogeneity in VECM-SV Models, [in:] *Metody ilościowe w naukach ekonomicznych (Quantitative Methods in Economic Sciences)*, [ed.:] A. Welfe, Warsaw School of Economics, Warszawa, 235-249.
- [26] Pajor A., (2010), *Wielowymiarowe procesy wariancji stochastycznej w ekonometrii finansowej. Ujęcie bayesowskie (Multivariate Stochastic Variance Processes in Financial Econometrics. Bayesian approach)*, in Polish, Cracow University of Economics, Kraków.
- [27] Rault Ch., (2011), Long-Run Strong-Exogeneity, *Economics Bulletin* 31, 1-8.
- [28] Tsay R.S., (2002), *Analysis of Financial Time Series. Financial Econometrics*, A Wiley-Interscience Publication, John Wiley & Sons, INC, New York.