ARTICLE IN PRESS
International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Testing causality between two vectors in multivariate GARCH models

Tomasz Woźniak *

Department of Economics, University of Melbourne, 111 Barry Street, Carlton 3053, Australia

ARTICLE INFO

Keywords:
Second-order Granger causality
Bayesian methods
Posterior odds
GARCH models
Volatility spillovers

ABSTRACT

The family of Constant Conditional Correlation GARCH models is used to model the risk associated with financial time series and to make inferences about Granger-causal relationships between second conditional moments. The restrictions for second-order Granger noncausality between two vectors of variables are derived and assessed using posterior odds ratios. This Bayesian method constitutes an alternative to classical tests and can be employed regardless of the form of the restrictions on the parameters of the model. This approach enables the assumptions about the existence of higher-order moments of the processes that are required in classical tests to be relaxed. In the empirical example, a bidirectional second-order causality between the pound-to-Euro and US dollar-to-Euro exchange rates is found.

1. Introduction

The concept of *Granger causality* was introduced to econometrics by Granger (1969) and Sims (1972). One vector of variables does not Granger-cause another vector of variables if past information about the former cannot improve the forecast of the latter. Thus, Granger causality or noncausality refers to the forecast of the conditional mean process. The basic definition is set for a forecast of the one-period-ahead value. The conditions imposed on the parameters of the linear vector autoregressive moving average model for Granger noncausality were derived by Boudjellaba, Dufour, and Roy (1992, 1994). However, the forecast horizon in the definition may be generalized to h or up to h periods ahead, and h may have its limit at infinity (see Dufour & Renault, 1998; Lütkepohl, 1993). Irrespective of the forecast horizon, the restrictions imposed on the parameters assuring noncausality may be nonlinear. This fact motivated the development of nonstandard testing procedures that allow for the empirical verification of hypotheses (see Boudjellaba et al., 1992; Dufour, Pelletier, & Renault, 2006; Lütkepohl & Burda, 1997).

Information about Granger-causal relationships between the conditional variances of time series is important, because it provides an understanding of the structure of the underlying time series. When Granger-causal relationships are introduced into a multivariate volatility model, they allow for the modeling of dynamic interdependencies between conditional variances. This feature changes the way in which persistence in the volatility is modeled, making it more flexible and more accurate, as has been shown by multiple studies (see e.g. Conrad & Karanasos, 2010; Nakatani & Teräsvirta, 2009). Therefore, Granger causality testing allows an appropriate forecasting model to be formed. Such modeling is potentially important in all applications that are based on volatility forecasting, such as risk management, asset allocation and macroeconomic fluctuations analysis.

The present paper examines Granger causality for conditional variances. Consequently, we refer to the concept of *second-order Granger causality*, which was introduced by Robins, Granger, and Engle (1986) and formally distinguished from the *Granger causality in variance* by Comte

* Tel.: +61 3 8344 5310.
E-mail address: tomasz.wozniak@unimelb.edu.au.
URL: http://bit.ly/tomaszw.

http://dx.doi.org/10.1016/j.ijforecast.2015.01.005
0169-2070/

and Lieberman (2000). The intuition is as follows: one vector of variables does not second-order Granger-cause another vector of variables if past information about the variability of the former cannot improve the forecast of the conditional variances of the latter. The formal definition of *second-order noncausality* (introduced in Section 2) assumes that Granger causal relationships can exist in the conditional mean process. However, they should be modeled and filtered out. Otherwise, such relationships may impact on the parameters responsible for causal relationships in conditional variances (see the empirical illustration of the problem by Karolyi, 1995).

Comte and Lieberman (2000) derived restrictions for one-period-ahead second-order noncausality for the family of BEKK-GARCH models. These conditions take the form of zero restrictions that are imposed on several nonlinear functions of the original parameters of the model. However, no good test for such restrictions has been established. The problem is that the matrix of the first partial derivatives of the restrictions, with respect to the parameters of the model, may not be of full rank. This fact translates to the unknown asymptotic properties of classical tests, even if the asymptotic distribution of the estimator is normal. As a consequence, the testing strategy proposed by Comte and Lieberman (2000) and Hafner and Herwartz (2008b) is to derive linear (zero) restrictions on the parameters that are a sufficient condition for the original restrictions. A Wald test is then applied to such conditions.

The conditions for the one-period-ahead second-order noncausality of the family of extended constant conditional correlation GARCH models of Jeantheau (1998) are derived in this paper. In this setting, all of the variables considered are split into two vectors, between which we investigate causal relationships in conditional variances. Then, the conditions for one-period-ahead second-order noncausality appear to be the same as those for second-order noncausality in all future periods. When compared with the work of Comte and Lieberman (2000), these conditions result in a smaller number of restrictions. This improvement has a practical meaning when computing the restricted models, and may also potentially have a significant impact on the properties of tests applied to the problem.

In order to assess the credibility of the noncausality hypotheses, posterior odds ratios are employed. In the context of Granger noncausality hypothesis testing, Bayes factors and posterior odds ratios were used by Droumaguet and Woźniak (2012) for Markov-switching VAR models. In the same context, Woźniak (2012) used a Lindley-type test for VARMA-GARCH models. Moreover, in order to assess the hypothesis of exogeneity, a concept that is related to Granger noncausality, Pajor (2011) used Bayes factors for models with latent variables, and in particular for multivariate stochastic volatility models. Jarociński and Maćkowiak (2013) used Savage–Dickey ratios for the VAR model.

The inference is performed using posterior odds ratios. Therefore, the inference is based on the exact finite sample results. In effect, referring to the asymptotic results becomes pointless. This finding enables a relaxation of the assumptions that are required in the classical inference about the existence of higher-order moments. For instance, in order to test the second-order noncausality hypothesis, the existence of fourth order unconditional moments is required in Bayesian inference. The solutions that currently exist in classical testing require the existence of sixth-order moments (see Ling & McAleer, 2003). Note that this assumption for testing such a hypothesis cannot be relaxed further in the context of the causal inference on second-conditional moments modeled with GARCH models. This finding is justified by the fact that this assumption comes, not from the properties of the test, but from the derivation of the restrictions for second-order noncausality. However, one should be aware of the tradeoffs that arise from the use of Bayesian inference instead of the frequentist tests based on asymptotic derivations. Bayesian inference is valid given certain assumptions about the model, whereas the quasi maximum likelihood inference is valid for more general assumptions regarding the distribution of the error term (see Comte & Lieberman, 2003; Hafner & Preminger, 2009).

The structure of the paper is as follows. Section 2 introduces the model considered and this work's main theoretical finding, namely the restrictions for second-order Granger noncausality. The assumptions behind the causal analysis are discussed. In Section 3, we present and discuss the existing classical approaches to testing for Granger noncausality. Since they are of limited use in the context considered, we further present the posterior odds ratios as the solution. Section 4 presents an empirical illustration of the methodology for two main exchange rates of the Eurozone, and Section 5 concludes.

2. Second-order noncausality for multivariate GARCH models

2.1. Model

First, the notation is set following Boudjellaba et al. (1994). Let $\{y_t : t \in \mathbb{Z}\}$ be a $N \times 1$ multivariate square integrable stochastic process on the integers $\mathbb{Z}$. Let $\mathbf{y} = (y_1, \ldots, y_T)'$ denote a time series of T observations. Write

$$y_t = (y'_{1t}, y'_{2t})', \tag{1}$$

for all $t = 1, \ldots, T$, where y_{it} is a $N_i \times 1$ vector such that $y_{1t} = (y_{1t}, \ldots, y_{N_1.t})'$ and $y_{2t} = (y_{N_1+1.t}, \ldots, y_{N_1+N_2.t})'$ ($N_1, N_2 \geq 1$ and $N_1 + N_2 = N$). y_1 and y_2 contain the variables of interest between which we want to study causal relationships. Further, let $I(t)$ be the Hilbert space generated by the components of y_τ, for $\tau \leq t$, i.e., an information set generated by the past realizations of y_t. Then, $\epsilon_{t+h} = y_{t+h} - P(y_{t+h}|I(t))$ is an error component. For any subspace I_t of $I(t)$ and for $N_1 + 1 \leq i \leq N_1 + N_2$, we denote by $P(y_{it+h}|I_t)$ the affine projection of y_{it+h} on I_t, i.e. the best linear prediction of y_{it+h} based on the variables in I_t and a constant term. Let $I^2(t)$ be the Hilbert space generated by the product of variables $\epsilon_{i\tau}\epsilon_{j\tau}$, $1 \leq i, j \leq N$ for $\tau \leq t$. $I_{-1}(t)$ is the closed subspace of $I(t)$ generated by the components of $y'_{2\tau}$ and $I^2_{-1}(t)$ is the closed subspace of $I^2(t)$ generated by the variables $\epsilon_{i\tau}\epsilon_{j\tau}$, $N_1 + 1 \leq i, j \leq N$ for $\tau \leq t$.

The model under consideration is the vector autoregressive process of Sims (1980) for the conditional mean, and the extended constant conditional correlation generalized autoregressive conditional heteroskedasticity process of Jeantheau (1998) for conditional variances. The conditional mean part models linear relationships between current and lagged observations of the variables considered:

$$y_t = \alpha_0 + \alpha(L)y_t + \epsilon_t \tag{2a}$$

$$\epsilon_t = D_t r_t \tag{2b}$$

$$r_t \sim i.i.St^N(\mathbf{0}, \mathbf{C}, \nu), \tag{2c}$$

for all $t = 1, \dots, T$, where y_t is a $N \times 1$ vector of data at time t, $\alpha(L) = \sum_{i=1}^{p} \alpha_i L^i$ is a lag polynomial of order p, ϵ_t and r_t are $N \times 1$ vectors of residuals and standardized residuals respectively, $D_t = \operatorname{diag}(\sqrt{h_{1t}}, \dots, \sqrt{h_{Nt}})$ is a $N \times N$ diagonal matrix with conditional standard deviations on the diagonal. The standardized residuals follow a N-variate standardized Student t distribution with a vector of zeros as a location parameter, a matrix $\mathbf{C}$ as a scale matrix and $\nu > 2$ as a degrees of freedom parameter. The choice of the distribution is motivated, on the one hand, by its ability to model potential outlying observations in the sample (for $\nu < 30$). On the other hand, it is a good approximation of the normal distribution when the value of the degrees of freedom parameter exceeds 30.

The conditional covariance matrix of the residual term ϵ_t is decomposed into:

$$H_t = D_t \mathbf{C} D_t \quad \forall t = 1, \dots, T. \tag{3}$$

For the matrix H_t to be a positive definite covariance matrix, h_t must be positive for all t and $\mathbf{C}$ must be positive definite (see Bollerslev, 1990). A $N \times 1$ vector of current conditional variances is modeled with lagged squared residuals, $\epsilon_t^{(2)} = (\epsilon_{1t}^2, \dots, \epsilon_{Nt}^2)'$, and lagged conditional variances:

$$h_t = \omega + A(L)\epsilon_t^{(2)} + B(L)h_t, \tag{4}$$

for all $t = 1, \dots, T$, where ω is a $N \times 1$ vector of constants, and $A(L) = \sum_{i=1}^{q} A_i L^i$ and $B(L) = \sum_{i=1}^{r} B_i L^i$ are lag polynomials of orders q and r of ARCH and GARCH effects respectively. The vector of conditional variances is given by $E[\epsilon_{t+1}^{(2)} | I^2(t)] = \frac{\nu}{\nu-2} h_{t+1}$, and the best linear predictor of $\epsilon_{t+1}^{(2)}$ in terms of a constant and $\epsilon_{t+1-i}^{(2)}$ for $i = 1, 2, \dots$ is $P(\epsilon_{t+1}^{(2)} | I^2(t)) = h_{t+1}$. Eq. (4) has a form respecting the partitioning of the vector of the data in Eq. (1):

$$\begin{bmatrix} h_{1t} \\ h_{2t} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t}^{(2)} \\ \epsilon_{2t}^{(2)} \end{bmatrix} + \begin{bmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{bmatrix} \begin{bmatrix} h_{1t} \\ h_{2t} \end{bmatrix}. \tag{5}$$

2.2. Assumptions and properties

Let $\theta \in \Theta \subset \mathbb{R}^k$ be a vector of size k, collecting all of the parameters of the model that are described with Eqs. (2)–(4). Then, the likelihood function has the following form:

$$p(\mathbf{y}|\theta) = \prod_{t=1}^{T} \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left((\nu-2)\pi\right)^{-\frac{N}{2}} |H_t|^{-\frac{1}{2}} \times \left(1 + \frac{1}{\nu-2}\epsilon_t' H_t^{-1} \epsilon_t\right)^{-\frac{\nu+N}{2}}. \tag{6}$$

This model has its origins in the constant conditional correlation GARCH (CCC-GARCH) model proposed by Bollerslev (1990), which consisted of N univariate GARCH equations describing the vector of conditional variances, h_t. The CCC-GARCH model is equivalent to Eq. (4) with diagonal matrices $A(L)$ and $B(L)$. Its extended version, with non-diagonal matrices $A(L)$ and $B(L)$, was analyzed by Jeantheau (1998). He and Teräsvirta (2004) call this model the extended CCC-GARCH (ECCC-GARCH). Such a formulation of the GARCH process allows for the modeling of volatility spillovers, as matrices of the lag polynomials $A(L)$ and $B(L)$ are not diagonal. Therefore, causality between variables in second-conditional moments may be analyzed.

For the purpose of deriving the restrictions for second-order Granger noncausality, four assumptions are imposed on the parameters of the conditional variance process.

Assumption 1. The parameters ω, $A = \left(\operatorname{vec}(A_1)', \dots, \operatorname{vec}(A_q)'\right)'$ and $B = \left(\operatorname{vec}(B_1)', \dots, \operatorname{vec}(B_r)'\right)'$ are such that the conditional variances, h_t, are positive for all t.

Assumption 2. All of the roots of $|I_N - A(z) - B(z)| = 0$ are outside the complex unit circle.

Assumption 3. All of the roots of $|I_N - B(z)| = 0$ are outside the complex unit circle.

Assumption 4. The multivariate GARCH(q, r) model is minimal, in the sense of Jeantheau (1998).

Define a process $v_t = \epsilon_t^{(2)} - h_t$. Then, $\epsilon_t^{(2)}$ follows a VARMA process given by:

$$\phi(L)\epsilon_t^{(2)} = \omega + \psi(L)v_t, \tag{7}$$

where $\phi(L) = I_N - A(L) - B(L)$ and $\psi(L) = I_N - B(L)$ are matrix polynomials of the VARMA representation of the GARCH(q, r) process. Suppose that $\epsilon_t^{(2)}$ and v_t are partitioned as y_t in Eq. (1). Then Eq. (7) can be written in the following form:

$$\begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) \\ \phi_{21}(L) & \phi_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t}^{(2)} \\ \epsilon_{2t}^{(2)} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}. \tag{8}$$

Given Assumption 3, the VARMA process in Eq. (7) is invertible and can be written in a VAR form:

$$\Pi(L)\epsilon_t^{(2)} - \omega^* = v_t, \tag{9}$$

where $\Pi(L) = \psi(L)^{-1}\phi(L) = [I_N - B(L)]^{-1}[I_N - A(L) - B(L)]$ is a matrix polynomial of the VAR representation of the GARCH(q, r) process and $\omega^* = \psi(L)^{-1}\omega$ is a constant term. Again, partitioning the vectors, we can rewrite Eq. (9) in the form:

$$\begin{bmatrix} \Pi_{11}(L) & \Pi_{12}(L) \\ \Pi_{21}(L) & \Pi_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t}^{(2)} \\ \epsilon_{2t}^{(2)} \end{bmatrix} - \begin{bmatrix} \omega_1^* \\ \omega_2^* \end{bmatrix} = \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}. \tag{10}$$

Under Assumptions 2 and 3, the processes in Eqs. (7) and (9) are both stationary. The GARCH model has well-established properties under Assumptions 1–4. Under Assumption 1, conditional variances are positive. The least restrictive conditions required for this assumption to hold are derived by Conrad and Karanasos (2010), and do not require all of the parameters of the GARCH model to be positive. However, for the sake of feasible Bayesian inference, the non-negativity conditions suggested by Bollerslev (1990) are used in this study. Further, Jeantheau (1998) proves that the GARCH(r, s) model, as in Eq. (4), has a unique, ergodic, weakly and strictly stationary solution when Assumption 2 holds. Under Assumptions 2–4, the GARCH(r, s) model, and thus v_t as well, are covariance stationary (Silvennoinen & Teräsvirta, 2009) and identifiable (Jeantheau, 1998). Jeantheau showed that the minimum contrast estimator for the multivariate GARCH model is strongly consistent under conditions of stationarity and identifiability, among others. Ling and McAleer (2003) proved strong consistency of the quasi maximum likelihood estimator (QMLE) for the VARMA-GARCH model under Assumptions 2–4, and when all of the parameters of the GARCH process are positive. Moreover, they have set asymptotic normality of the QMLE, provided that $E\|y_t\|^6 < \infty$. The extension of the asymptotic results under the conditions of Conrad and Karanasos (2010) has not yet been established. Finally, He and Teräsvirta (2004) give sufficient conditions for the existence of the fourth moment and derive a complete fourth-moment structure.

2.3. Estimation

Classical estimation consists of maximizing the likelihood function in Eq. (6). This is possible using one of the available numerical optimization algorithms. Due to the complexity of the problem, the algorithms require derivatives of the likelihood function. Hafner and Herwartz (2008a) give analytical solutions for the first and second partial derivatives of a normal likelihood function, whereas Fiorentini, Sentana, and Calzolari (2003) derive numerically reliable analytical expressions for the score, Hessian and information matrix for the models with conditional multivariate Student t distributions. Bayesian estimation requires numerical methods in order to simulate the posterior density of the parameters, given data. Unfortunately, neither the posterior distribution of the parameters nor the full conditional distribution has the form of any known distribution. The Metropolis–Hastings algorithm (see Chib & Greenberg, 1995, and references therein) was applied to multivariate GARCH models by Vrontos, Dellaportas, and Politis (2003).

The posterior distribution of the parameters of the model is proportional to the product of the likelihood function in Eq. (6) and the prior distribution of the parameters:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta). \tag{11}$$

For the unrestricted VAR-GARCH model, the following prior specification is assumed. All of the parameters of the VAR process are *a priori* normally distributed with a vector of zeros as a mean and a diagonal covariance matrix with 100s on the diagonal. A similar prior distribution is assumed for the constant terms of the GARCH process, with the difference that the distribution is truncated to the constrained parameter space for ω. The parameters modeling the dynamic part of the GARCH process, collected in matrices A and B, follow a truncated normally-distributed prior with a zero mean and diagonal covariance matrix with the hyper-parameter $\bar{s}$ on the diagonal. Each of the models in this study is estimated twice with two different values of the hyper-parameter, $\bar{s} \in \{0.1, 100\}$. This allows the sensitivity of the hypothesis assessment to the prior assumptions to be investigated. The truncation of the distribution to the parameter space imposes Assumptions 1–3. Each of the correlation parameters of the correlation matrix $\mathbf{C}$ follows a uniform distribution on the interval $[a_i, b_i]$, where a_i and b_i are determined for each correlation coefficient ρ_i separately, such that the resulting correlation matrix is positive definite (see Barnard, McCulloch, & Meng, 2000, for the algorithm determining a_i and b_i and for the consequences of these marginal prior distribution assumptions for the joint distribution of the correlation matrix). Finally, for the degrees of freedom parameter, the prior distribution proposed by Deschamps (2006) is assumed. To summarize, the prior specification for the model considered has a detailed form of:

$$p(\theta) = p(\alpha)p(\omega, A, B)p(\nu) \prod_{i=1}^{N(N-1)/2} p(\rho_i), \tag{12}$$

where each of the prior distributions is assumed:

$$\begin{aligned}
&\alpha \sim \mathcal{N}^{N+pN^2}\left(\mathbf{0}, 100 \cdot I_{N+pN^2}\right)\\
&\omega \sim \mathcal{N}^{N}\left(\mathbf{0}, 100 \cdot I_{N}\right)\mathcal{I}(\theta \in \Theta)\\
&\left(\mathrm{vec}(A)', \mathrm{vec}(B)'\right)'\\
&\quad \sim \mathcal{N}^{N^2(q+r)}\left(\mathbf{0}, \bar{s} \cdot I_{N^2(q+r)}\right)\mathcal{I}(\theta \in \Theta)\\
&\nu \sim 0.04\exp\left[-0.04(\nu-2)\right]\mathcal{I}(\nu \geq 2)\\
&\rho_i \sim \mathcal{U}(a_i, b_i) \quad \text{for } i = 1, \ldots, N(N-1)/2,
\end{aligned}$$

where $\alpha = (\alpha_0', \mathrm{vec}(\alpha_1)', \ldots, \mathrm{vec}(\alpha_p)')'$ stacks all of the parameters of the VAR process in a vector of size $N+pN^2$. I_n is an identity matrix of order n. $\mathcal{I}(.)$ is an indicator function that takes a value equal to 1 if the condition in the brackets holds and 0 otherwise. Finally, ρ_i is the ith element of a vector stacking all of the elements below the diagonal of the correlation matrix, $\rho = (\mathrm{vecl}(\mathbf{C}))$.

Such prior assumptions, with only proper distributions, have several consequences. First, together with the bounded likelihood function, the proposed prior distribution guarantees the existence of the posterior distribution (see Geweke, 1997). Second, the proper prior distribution for the degrees of freedom parameter of the Student t-distributed error term is required in order for the posterior distribution to be integrable, as was proven by Bauwens and Lubrano (1998). The prior distribution of the degrees of freedom parameter gives more than a 32% chance that its value will be higher than 30. Finally, it gives rise to a subjective interpretation of the probability, which is a feature of the Bayesian inference.

2.4. Second-order noncausality conditions

We focus on the question of the causal relationships between variables in conditional variances. Therefore, the proper concept to refer to is *second-order Granger noncausality*:

Definition 1. y_1 does not second-order Granger-cause y_2 h periods ahead if:

$$P\left[\epsilon^{(2)}(y_{2t+h}|I(t))|I^2(t)\right] = P\left[\epsilon^{(2)}(y_{2t+h}|I(t))|I^2_{-1}(t)\right], \quad (13)$$

for all $t \in \mathbb{Z}$, where $\epsilon_{2t+h} = \epsilon(y_{2t+h}|I(t)) = y_{2t+h} - P(y_{2t+h}|I(t))$ is an error component, $[.]^{(2)}$ means that we square each element of a vector, and $h \in \mathbb{Z}$.

A common part of both sides of Eq. (13) is that the potential Granger causal relationships in the conditional mean process are filtered out in the first step. This is represented by a projection of the forecasted value, y_{2t+h}, on the Hilbert space generated by the full set of variables, $P(y_{2t+h}|I(t))$. In the second stage, the square of the error component, $\epsilon^{(2)}(y_{2t+h}|I(t))$, is projected on the Hilbert space generated by cross-products of the full vector of the error component, $I^2(t)$ (on the LHS), and on the Hilbert space generated by the cross-products of a sub-vector of the error component, $I^2_{-1}(t)$ (on the RHS). If the two projections are equivalent, it means that $\epsilon^{(2)}(y_{2t+h}|I(t)) - P\left[\epsilon^{(2)}(y_{2t+h}|I(t))|I^2_{-1}(t)\right]$ is orthogonal to $I^2(t)$ for all t (see Comte & Lieberman, 2000; Florens & Mouchart, 1985a). Note also the difference between this definition of second-order noncausality and the definition of Comte and Lieberman (2000). In Definition 1, the Hilbert space $I^2(t)$ is generated by square integrable cross-products of the error components ϵ_τ, whereas in the definition of Comte and Lieberman it is generated by the cross-products of the variables $y_{i\tau}$ for $\tau \leq T$.

The definition was proposed in its original form, for one-period-ahead noncausality ($h = 1$), by Robins et al. (1986), and distinguished from *Granger noncausality in variance* by Comte and Lieberman (2000). The difference is that the definition of *Granger noncausality in variance* also assumes *Granger noncausality in mean*, while there is no such assumption in the definition of *second-order noncausality*. However, any existing causal relationship in conditional means needs to be modeled and filtered out before causality for the conditional variances process is analyzed.

The main theoretical contribution of this study is the theorem stating the restrictions for second-order Granger noncausality for the ECCC-GARCH model.

Theorem 1. *Let $\epsilon_t^{(2)}$ follow a stationary vector autoregressive moving average process as in Eq. (7), partitioned as in Eq. (8), for which Assumptions 1–4 hold. Then, y_1 does not second-order Granger-cause y_2 one period ahead if and only if:*

$$\Gamma^{so}_{ij}(z) = \det\begin{bmatrix} \phi^{.j}_{11}(z) & \psi_{11}(z) \\ \varphi_{n_1+i,j}(z) & \psi^{i.}_{21}(z)\end{bmatrix} = 0 \quad \forall z \in \mathbb{C}, \quad (14)$$

for $i = 1, \dots, N_2$ and $j = 1, \dots, N_1$, where $\phi^{.j}_{lk}(z)$ is the jth column of $\phi_{lk}(z)$, $\psi^{i.}_{lk}(z)$ is the ith row of $\psi_{lk}(z)$, and $\varphi_{n_1+1,j}(z)$ is the (i,j)-element of $\phi_{21}(z)$.

Theorem 1 establishes the restrictions on the parameters of the ECCC-GARCH model for second-order noncausality one period ahead between two vectors of variables. Its proof, presented in Appendix A, is based on the theory introduced by Florens and Mouchart (1985a) and applied by Boudjellaba et al. (1992) to VARMA models for the conditional mean. It is applicable to any specification of the GARCH(q, r) process, irrespective of the order of the model, (q, r), and the size of the time series, N.

Due to the setting proposed in this study, in which the vector of variables is split into two parts, the establishment of one-period-ahead second-order Granger noncausality is equivalent to establishing the noncausality relationship at all horizons up to infinity. This result is formalised in a corollary.

Corollary 1. *Suppose that the vector of observations is partitioned as in Eq. (1), and that y_1 does not second-order Granger-cause y_2 one period ahead, such that the condition in Eq. (14) holds. Then, y_1 does not second-order Granger-cause y_2 h periods ahead for all $h = 1, 2, \dots$.*

Corollary 1 is a direct application of Corollary 2.2.1 of Lütkepohl (2005, p. 45) to the GARCH process in the VAR form given in Eq. (10). For the proof of the restrictions for second-order Granger noncausality for the GARCH process in the VAR form, the reader is referred to Appendix A.

Corollary 1 shows the features of the particular setting considered in this work, i.e., the setting in which all of the variables are split between two vectors. If one is interested in the second-order causality relationships at all horizons at once, then one may use just one set of restrictions. However, the restrictions imply a very strong result. If a more detailed analysis with respect to the forecast horizon is required, then one must consider deriving other solutions.

In order to interpret the parameter restrictions derived, a bivariate model is analyzed.

Example 1. Suppose that y_t follows the bivariate ECCC-GARCH(1, 1) process of Jeantheau (1998) ($N = 2$, and $q = r = 1$). The VARMA process for $\epsilon_t^{(2)}$ is as follows:

$$\begin{bmatrix} 1-(A_{11}+B_{11})L & -(A_{12}+B_{12})L \\ -(A_{21}+B_{21})L & 1-(A_{22}+B_{22})L \end{bmatrix}\begin{bmatrix}\epsilon^2_{1t}\\ \epsilon^2_{2t}\end{bmatrix}$$
$$= \begin{bmatrix}\omega_1\\ \omega_2\end{bmatrix} + \begin{bmatrix} 1-B_{11}L & -B_{12}L \\ -B_{21}L & 1-B_{22}L\end{bmatrix}\begin{bmatrix}\nu_{1t}\\ \nu_{2t}\end{bmatrix}. \quad (15)$$

From Theorem 1, it follows that y_1 does not second-order Granger-cause y_2 if and only if:

$$\det\begin{bmatrix} 1-(A_{11}+B_{11})z & 1-B_{11}z \\ -(A_{21}+B_{21})z & -B_{21}z\end{bmatrix} \equiv \mathbf{0}, \quad (16)$$

which leads to the following set of restrictions:

$$\mathbf{R}^I_1(\theta) = A_{21} = 0, \quad \text{and} \quad \mathbf{R}^I_2(\theta) = B_{21}A_{11} = 0. \quad (17)$$

The final noncausality condition for the model presented in Eq. (17) consists of two restrictions. The first is a linear zero restriction, and the second is a zero condition for a product of two parameters. A very intuitive interpretation of the restrictions suggests itself for the case when the restrictions take the form: $A_{21} = B_{21} = 0$. These two parameters are the marginal effects of lagged squared residuals and the lagged conditional variance of y_1, respectively, on the current conditional variance of y_2. However, there is also another case in which the marginal effect of $h_{1.t-1}$ on $h_{2.t}$ is not zero, $B_{21} \neq 0$, and the condition for noncausality holds. This claim is correct when $h_{1.t}$ is not influenced by the lagged squared residual of y_1, $A_{11} = 0$, but only $\epsilon^2_{2.t-1}$ and the lagged conditional variances h_{t-1} affect it.

Nakatani and Teräsvirta (2009) propose the Lagrange Multiplier test for the hypothesis of no volatility spillovers in a bivariate ECCC-GARCH model. The restrictions that they test are zero restrictions on the off-diagonal elements of the matrix polynomials $A(L)$ and $B(L)$ from the GARCH in Eq. (5). Consequently, the null hypothesis is represented by the CCC-GARCH model of Bollerslev (1990) and the alternative hypothesis by the ECCC-GARCH of Jeantheau (1998). Note that if all of the parameters on the diagonal of the matrices of the lag polynomial $A_{11}(L)$ are assumed to be strictly greater than zero (which can be tested and which in fact is the case for numerous time series that have been considered in applied studies), then the null hypothesis of Nakatani and Teräsvirta (2009) is equivalent to the second-order Granger noncausality condition, as in Example 1. In a general case, for any dimension of the time series, the zero restrictions on the off-diagonal elements of the matrix polynomials $A(L)$ and $B(L)$ represent only a sufficient condition for second-order noncausality.

Theorem 1 has an equivalent for other models from the GARCH family, namely the BEKK-GARCH models. The restrictions were introduced by Comte and Lieberman (2000). However, there are serious differences between the approach presented by Comte and Lieberman and that considered in this study. In order to understand the differences, it is worth studying the noncausality conditions for a BEKK-GARCH model.

Example 2. Suppose that y_t follows a bivariate BEKK-GARCH(1, 1) process ($N = 2$ and $q = r = 1$) that has a vech representation:

$$\begin{bmatrix} h_{1t} \\ h_{12t} \\ h_{2t} \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \epsilon^2_{1t-1} \\ \epsilon_{1t-1}\epsilon_{2t-1} \\ \epsilon^2_{2t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} h_{1t-1} \\ h_{12t-1} \\ h_{2t-1} \end{bmatrix}, \tag{18}$$

where h_{12t} is a conditional covariance between y_1 and y_2. From Theorem 1 of Comte and Lieberman (2000) and the determinant conditions given on page 545 of Comte and Lieberman (2000), it follows that y_2 does not second-order Granger-cause y_1 if and only if:

$\mathbf{R}_1 : a_{12} = 0$

$\mathbf{R}_2 : a_{13} = 0$

$\mathbf{R}_3 : b_{13}(b_{23} - a_{23} - b_{32}) - a_{22}b_{12} = 0$

$\mathbf{R}_4 : b_{13}(b_{22} + b_{23} - a_{33}) - b_{12}(b_{33} + b_{23} + a_{23}) = 0$

$\mathbf{R}_5 : a_{22}(b_{12}b_{33} - b_{13}b_{23}) + a_{32}(b_{13}b_{22} - b_{12}b_{23}) + b_{32}(b_{13}b_{22} - b_{12}b_{23}) + b_{23}(b_{12}b_{23} - b_{13}b_{22}) = 0$

$\mathbf{R}_6 : a_{23}(b_{12}b_{33} - b_{13}b_{23}) + a_{33}(b_{13}b_{22} - b_{12}b_{23}) = 0.$

First, in a bivariate model for the hypothesis that one variable does not second-order cause the other, the restrictions of Comte and Lieberman lead to six restrictions, whereas Example 1 shows that only two restrictions are required in order to test such a hypothesis. The difference in the number of restrictions increases with the dimension of the time series. Second, due to the formulation of the BEKK-GARCH model, the noncausality conditions are much more complicated than those for the ECCC-GARCH model considered here. They are simply much more complex functions of the original parameters of the model. Both of these arguments have consequences for testing that require the estimation of the restricted model or the employment of the delta method. A large number of restrictions may have a strongly unfavorable impact on the size and power properties of the tests. However, the ECCC-GARCH model assumes that the correlations are time invariant, which is not the case for the BEKK-GARCH model.

To illustrate how the level of complexity of the restrictions from Theorem 1 increases with the number of variables, a trivariate example for the ECCC-GARCH process is analyzed.

Example 3. Let y_t follow a trivariate ECCC-GARCH(1, 1) process ($N = 3$ and $q = r = 1$). The VARMA process for $\epsilon^{(2)}_t$ is as follows:

$$\begin{bmatrix} 1-(A_{11}+B_{11})L & -(A_{12}+B_{12})L & -(A_{13}+B_{13})L \\ -(A_{21}+B_{21})L & 1-(A_{22}+B_{22})L & -(A_{23}+B_{23})L \\ -(A_{31}+B_{31})L & -(A_{32}+B_{32})L & 1-(A_{33}+B_{33})L \end{bmatrix} \times \begin{bmatrix} \epsilon^2_{1t} \\ \epsilon^2_{2t} \\ \epsilon^2_{3t} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} + \begin{bmatrix} 1-B_{11}L & -B_{12}L & -B_{13}L \\ -B_{21}L & 1-B_{22}L & -B_{23}L \\ -B_{31}L & -B_{32}L & 1-B_{33}L \end{bmatrix} \begin{bmatrix} \nu_{1t} \\ \nu_{2t} \\ \nu_{3t} \end{bmatrix}. \tag{19}$$

From Theorem 1, it follows that $y_1 = y_1$ does not second-order Granger-cause $y_2 = (y_2, y_3)$ if and only if:

$$\det \begin{bmatrix} 1-(A_{11}+B_{11})z & 1-B_{11}z \\ -(A_{i1}+B_{i1})z & -B_{i1}z \end{bmatrix} = 0 \quad \text{for } i = 2, 3, \tag{20}$$

which results in the following restrictions:

$$\mathbf{R}^{\mathrm{II}}_1(\psi) = A_{11}B_{21} = 0 \quad \text{and} \quad \mathbf{R}^{\mathrm{II}}_2(\psi) = A_{21} = 0 \tag{21a}$$

$$\mathbf{R}^{\mathrm{II}}_3(\psi) = A_{11}B_{31} = 0 \quad \text{and} \quad \mathbf{R}^{\mathrm{II}}_4(\psi) = A_{31} = 0. \tag{21b}$$

However, $y_1 = (y_1, y_2)$ does not second-order Granger-cause $y_2 = y_3$ if and only if:

$$\det \begin{bmatrix} 1-(A_{11}+B_{11})z & 1-B_{11}z & -B_{13}z \\ -(A_{21}+B_{21})z & -B_{21}z & -B_{23}z \\ -(A_{31}+B_{31})z & -B_{31}z & 1-B_{33}z \end{bmatrix} \equiv \mathbf{0}, \tag{22}$$

which leads to the following set of restrictions:

$$\mathbf{R}^{\mathrm{III}}_1(\psi) = A_{11}(B_{23}B_{31} - B_{21}B_{33}) + A_{31}(B_{13}B_{21} - B_{11}B_{23}) = 0 \tag{23a}$$

$$\mathbf{R}^{\mathrm{III}}_2(\psi) = A_{11}B_{21} + A_{31}B_{23} = 0 \tag{23b}$$

$$\mathbf{R}^{\mathrm{III}}_3(\psi) = A_{21} = 0. \tag{23c}$$

2.5. Discussion

The proposed analysis of Granger causality in conditional variances is based on the linear theory introduced by Florens and Mouchart (1985a) (see also Florens & Mouchart, 1982, 1985b). The analysis is fully consistent with this theory, and therefore the properties developed by Florens and Mouchart (1982) apply. However, this comes at the cost of the linear assumptions and has practical implications for the possibilities of extending the model. Definition 1 uses the notion of the best linear predictor. Its application excludes the possibility of extending the model by including nonlinear relationships in the conditional mean process, the conditional variances process, and the process for the conditional correlations, among others. Therefore, Granger causality analysis for a GARCH model with leverage effects or dynamic processes for conditional correlation, such as the DCC process of Engle (2002) or the time-varying correlations process of Tse and Tsui (2002), is excluded. Such a causality analysis would need to refer to a different definition of Granger causality that would be based on a conditional expectation operator.

3. Bayesian hypotheses assessment

The restrictions derived in Section 2 can be tested. We propose to use the Bayesian approach to assess the hypotheses of second-order noncausality represented by the restrictions. However, before presenting the approach, we discuss the classical tests proposed so far and their limitations.

3.1. Classical testing

The testing of second-order noncausality has been considered only for the family of BEKK-GARCH and vec-GARCH models. Comte and Lieberman (2000) did not propose any test because asymptotic normality of the maximum likelihood estimator had not yet been established at that time. The asymptotic result was presented by Comte and Lieberman (2003) and Hafner and Preminger (2009). However, this finding does not solve the problem of testing the nonlinear restrictions imposed on the parameters of the model (see Example 2). In the easy case, when the restrictions are linear, the asymptotic normality of the estimator implies that the Wald, Lagrange Multiplier and Likelihood Ratio test statistics have asymptotic χ^2 distributions. Therefore, the Wald test statistic for the linear restrictions (which are the only sufficient condition for the original restrictions) proposed by Comte and Lieberman is χ^2-distributed. Similar procedures were presented by Hafner and Herwartz (2008b) for the Wald test and Hafner and Herwartz (2006) for the LM test. For the ECCC-GARCH model, Nakatani and Teräsvirta (2009) proposed the Lagrange Multiplier test for the hypothesis of no volatility spillovers. The test statistic is shown to be asymptotically normally distributed. Again, Nakatani and Teräsvirta (2009) tested only the linear zero restrictions.

In this study, the necessary and sufficient conditions for second-order noncausality between variables are tested. The restrictions, unlike the conditions of Comte and Lieberman, Hafner and Herwartz and Nakatani and Teräsvirta, may be nonlinear (see Examples 1 and 3). In such a case, a matrix of the first partial derivatives of the restrictions with respect to the parameters may not be of full rank. To see this, consider the bivariate ECCC-GARCH(1, 1) model from Example 1. In this case, the Jacobian matrix of parameter conditions from Eq. (17) with respect to the parameters occurring in the restrictions is given by:

$$\begin{bmatrix} 0 & 1 & 0 \\ B_{21} & 0 & A_{11} \end{bmatrix},$$

and under the null hypothesis it may be, not of full rank, but of rank equal to 1. Thus, the asymptotic covariance matrix of the restrictions constructed with the delta method is singular. In effect, the asymptotic distribution of the Wald test statistic is no longer normal. The asymptotic normality is distorted when the estimates of parameters A_{11} and B_{21} are close to zero as well. Consequently, the critical values of the usual Wald test cannot be used to test the necessary and sufficient conditions for second-order noncausality in multivariate GARCH models.

This problem is well known in the studies on the testing of parameter conditions for Granger noncausality in multivariate models. Boudjellaba et al. (1992) derive conditions for Granger noncausality for VARMA models that result in multiple nonlinear restrictions on the original parameters of the model. As a solution to the problem of testing the restrictions, they propose a sequential testing procedure. This method has two main drawbacks. First, the test may still appear inconclusive in spite of being performed properly, and second, the confidence level is given in the form of inequalities. Dufour et al. (2006) propose solutions based on the linear regression techniques that are applied for h-step-ahead Granger noncausality for VAR models. Unfortunately, though, the solutions proposed apply only to linear models for first conditional moments. Lütkepohl and Burda (1997) proposed a modified Wald test statistic as a solution to the problem of testing the nonlinear restrictions for h-step-ahead Granger noncausality for VAR models. This method could be applied to the problem of testing the nonlinear restrictions for second-order noncausality in GARCH models. However, more studies on the applicability and properties of this test are required.

Asymptotic results for the models and tests discussed here are established under the following moment conditions. For the BEKK-GARCH models, the Wald tests proposed by Comte and Lieberman (2000) and Hafner and Herwartz (2008b) require asymptotic normality of the quasi maximum likelihood estimator. This result is derived under the existence of bounded moments of order four by Hafner and Preminger (2009). For the ECCC-GARCH model considered in this study, the asymptotic normality of the quasi maximum likelihood estimator is derived by Ling and McAleer (2003) under the existence of moments of order six. However, this assumption is relaxed by Nakatani and Teräsvirta (2009) for the purpose of testing for the existence of volatility spillovers. Their Lagrange Multiplier test statistic requires the existence of fourth-order moments. The Bayesian test presented below relaxes this assumption further.

3.2. Bayesian hypotheses assessment

In order to compare the models restricted according to the noncausality restrictions derived in Theorem 1, Bayes factors are used, while posterior odds ratios of the hypotheses are employed in order to assess hypotheses of noncausality.

Bayes factors are a well-known method for comparing econometric models (see Geweke, 1995; Kass & Raftery, 1995). Denote the m models representing competing hypotheses by $\mathcal{M}_i$, for $i = 1, \ldots, m$. Let

$$p(\mathbf{y}|\mathcal{M}_i) = \int_{\theta\in\Theta} p(\mathbf{y}|\theta, \mathcal{M}_i)p(\theta|\mathcal{M}_i)d\theta \tag{24}$$

be marginal distributions of data corresponding to the various models, for $i = 1, \ldots, m$. $p(\mathbf{y}|\theta, \mathcal{M}_i)$ and $p(\theta|\mathcal{M}_i)$ are the likelihood function in Eq. (6) and the prior distribution in Eq. (12) respectively. The extended notation respecting conditioning on one of the models is used here. The marginal density of the data is a constant normalizing kernel of the posterior distribution in Eq. (11).

A Bayes factor is a ratio of the marginal densities of data for the two models selected:

$$\mathcal{B}_{ij} = \frac{p(\mathbf{y}|\mathcal{M}_i)}{p(\mathbf{y}|\mathcal{M}_j)}, \tag{25}$$

where $i, j = 1, \ldots, m$ and $i \neq j$. The Bayes factor takes positive values, and its value above one is interpreted as evidence for model $\mathcal{M}_i$, whereas its value below one is evidence for model $\mathcal{M}_j$. For the further interpretation of the value of the Bayes factor, the reader is referred to the paper by Kass and Raftery (1995).

The posterior probability of the ith model is computed using the Bayes formula:

$$\Pr(\mathcal{M}_i|\mathbf{y}) = \frac{p(\mathbf{y}|\mathcal{M}_i)\Pr(\mathcal{M}_i)}{\sum_{j=1}^{m} p(\mathbf{y}|\mathcal{M}_j)\Pr(\mathcal{M}_j)}, \tag{26}$$

where $\Pr(\mathcal{M}_j)$ is *a priori* a probability of model j.

The hypotheses of interest, $\mathcal{H}_i$, where i denotes the particular hypothesis, may be assessed using posterior probabilities of hypotheses, $\Pr(\mathcal{H}_i|\mathbf{y})$. These are computed by summing the posterior probabilities of the non-nested models representing hypothesis i:

$$\Pr(\mathcal{H}_i|\mathbf{y}) = \sum_{j\in\mathcal{H}_i} \Pr(\mathcal{M}_j|\mathbf{y}). \tag{27}$$

This approach to the assessment of hypotheses requires the estimation of all of the models representing the hypotheses considered, as well as the corresponding marginal densities of data (Eq. (24)). Moreover, in order to use the Bayes formula from Eq. (26) and to compute the posterior probabilities of hypotheses as in Eq. (27), it is required that the models considered in the assessment be mutually exclusive. This assumption may be problematic, as the hypothesis of noncausality is represented by models that are nested in the unrestricted model representing the hypothesis of causality by imposing zero restrictions on the parameters. In this case, the assumption is satisfied formally by excluding the value zero, $\Theta\backslash\{0\}$, from the parameter space of more general models. However, this constraint is not very restrictive, bearing in mind that a single point in a continuous parameter space has a zero probability mass. Thus, Bayesian estimation algorithms need not be modified seriously.

The sensitivity of the model assessment with respect to the specification of the prior distribution is checked by assuming two different prior distributions for the matrices of parameters A and B for each of the models estimated. The distributions, defined in Section 2, differ in the variances of the distributions. One of the variances is equal to 0.1, representing a shrinkage prior distribution, and the other is equal to 100, representing a diffuse prior distribution.

3.3. Estimation of models

The form of the posterior distribution in Eq. (11) for all of the parameters, θ, for the GARCH models is not in the form of any known distribution function, even with the prior distribution set to a proper distribution function, as in Eq. (12). Moreover, none of the full conditional densities for any sub-group of the parameter vector has a form corresponding to a standard distribution. Still, the posterior distribution, although it is known only up to a normalizing constant, exists; this is ensured by the bounded likelihood function and the proper prior distribution. Therefore, the posterior distribution may be simulated using a Monte Carlo Markov Chain (MCMC) algorithm. Due to the above-mentioned problems with the form of the posterior and the full conditional densities, an example of a proper algorithm for sampling the posterior distribution in Eq. (11) would be the Metropolis–Hastings algorithm (see Chib & Greenberg, 1995, and references therein). This algorithm was adapted for multivariate GARCH models by Vrontos et al. (2003).

Suppose that the starting point of the Markov Chain is some value $\theta_0 \in \Theta$. Let $q(\theta^{(s)}, \theta'|\mathbf{y}, \mathcal{M}_i)$ denote the proposal density (candidate-generating density) for the transition from the current state of the Markov chain $\theta^{(s)}$ to a candidate draw θ'. The candidate density for model $\mathcal{M}_i$ depends on the data $\mathbf{y}$. In this study, a multivariate Student t distribution is used, with the location vector set to the current state of the Markov chain, $\theta^{(s)}$, the scale matrix Ω_q, and a degrees of freedom parameter of five. The scale matrix, Ω_q, should be determined by preliminary runs of the MCMC algorithm, such that it is close to the covariance matrix of the posterior distribution. Such a candidate-generating density should enable the algorithm to draw from the posterior density relatively efficiently. A new candidate θ' is accepted with the probability:

$$\alpha(\theta^{(s)}, \theta'|\mathbf{y}, \mathcal{M}_i) = \min\left[1, \frac{p(\mathbf{y}|\theta', \mathcal{M}_i)p(\theta'|\mathcal{M}_i)}{p(\mathbf{y}|\theta^{(s)}, \mathcal{M}_i)p(\theta^{(s)}|\mathcal{M}_i)}\right], \tag{28}$$

and if it is rejected, then $\theta^{(s+1)} = \theta^{(s)}$. The sample drawn from the posterior distribution with the Metropolis–Hastings algorithm, $\{\theta^{(s)}\}_{s=1}^{S}$, should be diagnosed so as to ensure that it is a good sample from the stationary posterior distribution (see e.g. Geweke, 1999; Plummer et al., 2006).

3.4. Estimation of marginal densities of data

Having estimated the models, the marginal densities of the data (MDD) may be computed using one of the available methods. Since the estimation of the models is performed using the Metropolis–Hastings algorithm, suitable estimators of the MDD are presented by Chib and Jeliazkov (2001) and Geweke (1999, 2005). However, any estimator of the marginal density of data that is applicable to the problem might be used (see Miazhynskaia & Dorffner, 2006, who review the estimators of MDD for univariate GARCH models). A Bayesian comparison of bivariate GARCH models using Bayes factors was presented by Osiewalski and Pipień (2004).

In the empirical part of this paper, we report the results using the estimator of Chib and Jeliazkov (2001). This estimator of the MDD is computed using the so-called marginal likelihood identity:

$$\log p\left(\mathbf{y}|\mathcal{M}_i\right) = \log p\left(\mathbf{y}|\theta^*, \mathcal{M}_i\right) + \log p\left(\theta^*|\mathcal{M}_i\right) + \log p\left(\theta^*|\mathbf{y}, \mathcal{M}_i\right), \tag{29}$$

where θ^* is a point in the parameter space for which the posterior density has a high value (posterior means are used in the empirical example of Section 4), and the right hand side of the formula above includes the ordinates of the likelihood function, the prior density function and the posterior density function respectively, evaluated at θ^*. Given our prior assumptions specified in Eq. (12), where the parameter space of ω, A and B is constrained, the computations of $\log p\left(\mathbf{y}|\mathcal{M}_i\right)$ require adjustments for a normalising constant that takes into account the constraints imposed on Θ. For the parameters ω, A and B, the condition $\theta \in \Theta$ denotes the following constraints imposed on the parameter space: $\omega > 0$, $A, B \geq 0$, and that the absolute value of the largest eigenvalue of the matrix $A + B$ is less than one.

Given the truncation of the parameter space, the ordinate of the posterior distribution evaluated at θ^* is estimated by:

$$\hat{p}\left(\theta^*|\mathbf{y}, \mathcal{M}_i\right) = \frac{S^{-1}\sum_{s=1}^{S}\alpha(\theta^{(s)}, \theta^*|\mathbf{y}, \mathcal{M}_i)q(\theta^{(s)}, \theta^*|\mathbf{y}, \mathcal{M}_i)}{J^{-1}\sum_{j=1}^{J}\alpha(\theta^*, \theta^{(j)}|\mathbf{y}, \mathcal{M}_i)}, \tag{30}$$

where $\{\theta^{(j)}\}_{j=1}^{J}$ is a sample drawn from $q(\theta^*, \theta'|\mathbf{y}, \mathcal{M}_i)$. The adjustment of the estimator of the ordinate of the posterior distribution for the truncation consists of setting the probabilities of acceptance, α, in the numerator and the denominator of Eq. (30) to zeros for the draws for which the condition $\theta \in \Theta$ does not hold.

The ordinate of the constrained prior density function is computed by:

$$p\left(\theta^*|\mathcal{M}_i\right) = \frac{\phi\left(\theta^*\right)\mathcal{I}\left(\theta \in \Theta\right)}{p_\Theta},$$

where ϕ is a probability density function of a normal distribution with appropriate moments, and p_Θ is a normalizing constant that is the probability of condition $\theta \in \Theta$ given measure ϕ. Due to the fact that the conditions for the stationarity of the GARCH model are specified in terms of conditions on the eigenvalues of matrix $A + B$, p_Θ is approximated for each model by simulation techniques. In order to estimate the normalising constant p_Θ efficiently, firstly, we compute analytically a probability that $\omega_i > 0$, $0 \leq A_{ij} \leq 1.8$ and $0 \leq B_{ij} \leq 1.8$, for $i, j = 1, \dots, N$. Secondly, 10 million independent draws from the normal prior distribution for parameters ω, A and B, truncated to the interval (0; 1.8), are used to compute the probability that Assumption 2 holds. This probability is denoted by $\Pr\left[|\text{eigenvalue}(A+B)| < 1|0 \leq A, B \leq 1.8\right]$. We report the values of these simulated probabilities for each model in Table 12. Then, p_Θ is computed by:

$$\begin{aligned}\Pr\left[\omega > 0 \wedge A, B \geq 0 \wedge |\text{eigenvalue}(A+B)| < 1\right] &= \Pr\left[\omega > 0\right] \cdot \Pr\left[|\text{eigenvalue}(A+B)| < 1|0 \leq A,\right.\\ &\left.B \leq 1.8\right] \cdot \Pr\left[0 \leq A, B \leq 1.8\right].\end{aligned}$$

The conditioning on other elements is suspended in this formula. The truncation of the normal distribution to the interval (0; 1.8) used in the simulations is determined such that the probability that the stationarity conditions hold outside the interval is numerically zero, and this was checked in auxiliary simulations.

3.5. Discussion

The proposed approach to testing the second-order noncausality hypothesis for GARCH models has several appealing features. First of all, the proposed Bayesian testing procedure makes it possible to test the parameter conditions. It avoids the singularities that may appear in classical tests, in which the restrictions imposed on the parameters are nonlinear.

Secondly, since the competing hypotheses are compared with Bayes factors, they are treated symmetrically. Thanks to the interpretation of the Bayes factors that comes from the posterior odds ratio, the outcome of the test is a positive argument *in favour of* the most likely *a posteriori* hypothesis. Moreover, unlike with classical testing, a choice is being made between all of the competing hypotheses at once, rather than only between the unrestricted and one of the restricted models (for a discussion of the argument, see Hoogerheide, van Dijk, & van Oest, 2009).

Furthermore, as the testing outcome is based on the posterior analysis, the inference has an exact finite sample justification, making it unnecessary to refer to asymptotic theory. In consequence, the assumptions required for testing the restrictions may now be relaxed. In order to test the second-order noncausality hypothesis, Assumptions 1–4 must hold. This requires the existence of fourth-order unconditional moments that can be ensured by the restrictions derived by He and Teräsvirta (2004). No classical test of the restrictions has been proposed for the ECCC-GARCH model thus far. However, such a potential test would depend on the moment conditions required for the asymptotic normality of the estimator, and thus, the existence of unconditional moments of an order of at least six would be required (see Ling & McAleer, 2003).

For the testing of volatility spillovers in the ECCC-GARCH model, the assumption may be relaxed further. Here, the strict assumption for Florens and Mouchart's (1985a) linear theory of noncausality need not hold. In fact, when testing the zero restrictions for the no-volatility-spillovers hypothesis, the only assumption required about the moments of the process is that the conditional variances must exist and be bounded. Not even the existence of the second unconditional moments of the process is required. Again, this result is an improvement, in comparison with the test of Nakatani and Teräsvirta (2009), which requires the existence of fourth-order unconditional moments in order for the Lagrange Multiplier test statistic to be asymptotically χ^2-distributed.

Thus, the improvements in moment conditions are established for both kinds of hypotheses. This fact may be crucial for the testing of the hypotheses on the financial time series. In multiple applied studies, such data have been shown to have the distribution of the residual term, with thicker tails than those of the normal distribution. Then, distributions modeling this property, such as the Student t distribution function, are employed. We follow this methodological finding, assuming exactly this distribution function.

The main cost of the proposed approach is the necessity for all of the unrestricted and restricted models to be estimated. This simply requires some time-consuming computations. While bivariate GARCH models may be able to be estimated reasonably quickly (depending on the order of the process, and thus on the number of the parameters), trivariate models require significant amounts of time and computational power. Also Bayesian inference is conditioned on the assumptions of the model. In particular, the assumptions regarding the distribution of the error term require careful specification, and the inference may be vulnerable to its potential misspecification. In the frequentist approach, the quasi maximum likelihood inference is valid under a more general true distribution of the error term, even if the assumed distribution is normal. Moreover, second-order causality testing should be performed by comparing the restricted and unrestricted models belonging to the same class of models, e.g. ECCC-GARCH or BEKK-GARCH models. The restricted and unrestricted models should have the same characteristics except for the noncausality restrictions. Otherwise, the Bayes factors would be informative about the in-sample fit of the model rather than about the hypotheses themselves.

4. Granger causal analysis of exchange rates

The restrictions derived in Section 2 for second-order noncausality for GARCH models, along with the Bayesian testing procedure described in Section 3, are now used in the analysis of a bivariate system of two exchange rates.

4.1. Data

The system under consideration consists of daily exchange rates of the British pound (GBP/EUR) and the US dollar (USD/EUR) to the Euro. Logarithmic rates of returns, expressed in percentage points, are analyzed, $y_{it} = 100(\ln x_{it} - \ln x_{it-1})$ for $i = 1, 2$, where x_{it} are the nominal values of the exchange rates. The data available span the period from 4 January 1999 to 31 December 2013, which gives 3841 observations, and were downloaded from the European Central Bank website (http://sdw.ecb.int/browse.do?node=2018794). However, the period of primary interest starts the day after Lehman Brothers filed for Chapter 11 bankruptcy protection, i.e., on 16 September 2008, which gives $T = 1357$ observations, counting through to the last day of year 2013.

The data set contains the two most liquid exchange rates in the Eurozone. The period chosen for analysis starts just after an event that had a very strong impact on the turmoil in financial markets: the bankruptcy of Lehman Brothers Holding Inc. The proposed analysis of the second-order causality between the series may therefore be useful for both financial and public institutions located in the Eurozone, the performances of which depend on forecasts of exchange rates. Such institutions include the governments of the countries belonging to the Eurozone, which keep their debts in currencies, mutual funds and banks, and all of the participants in the exchange rates market.

Table 1
Summary statistics of the exchange rate: logarithmic rates of return expressed in percentage points.

	GBP/EUR	USD/EUR
Mean	0.0041	0.0041
Median	−0.007	0.008
Standard deviation	0.502	0.653
Correlation	–	0.507
Minimum	−2.657	−4.735
Maximum	3.461	4.204
Excess kurtosis	3.726	2.499
Excess kurtosis (robust)	0.136	0.180
Skewness	0.35	0.038
Skewness (robust)	0.031	−0.007
LJB test stat.	2300.19	1000.858
LJB p-value	0.000	0.000
T	3841	3841

Note: The excess kurtosis (robust) and skewness (robust) coefficients are outlier-robust versions of the excess kurtosis and skewness coefficients, as described by Kim and White (2004). The LJB test and LJB p-values describe the tests of normality from Lomnicki (1961) and Jarque and Bera (1980).

Fig. 1 plots the time series, with the vertical black line marking 15 September 2008. The time series clearly follow the pattern of persistent periods of low and high volatility. An analysis of the plot to the right of the black line shows the first period of length – of nearly a year – which may be characterized by a high level of volatility of the exchange rates. The subsequent period is characterized by a slightly lower volatility for both series. The evident heteroskedasticity, together with the volatility clustering, seem to provide a strong argument in favor of specifying the GARCH models that are capable of modeling such features in the data.

Table 1 reports summary statistics of the two series considered. Both of the returns series have sample means and medians that are close to zero. The US dollar has a slightly larger sample standard deviation than the British pound. Both series are leptokurtic, as is evidenced by the fact that the excess kurtosis coefficient exceeds the level of 2. Both of the exchange rates are slightly positively skewed; however, the robust skewness coefficient is slightly negative for the dollar. Neither series can be described well using a normal distribution. These features of the returns seem to confirm the choice of a Student t-distributed likelihood function. We also consider a skewed multivariate t distribution in order to check the robustness of the results due to the potential skewness of the empirical distribution.

4.2. Testing strategy and estimation results

For the bivariate time series of exchange rates, the vector autoregression of order one with the extended CCC-GARCH(1, 1) conditional variance process is fitted with two

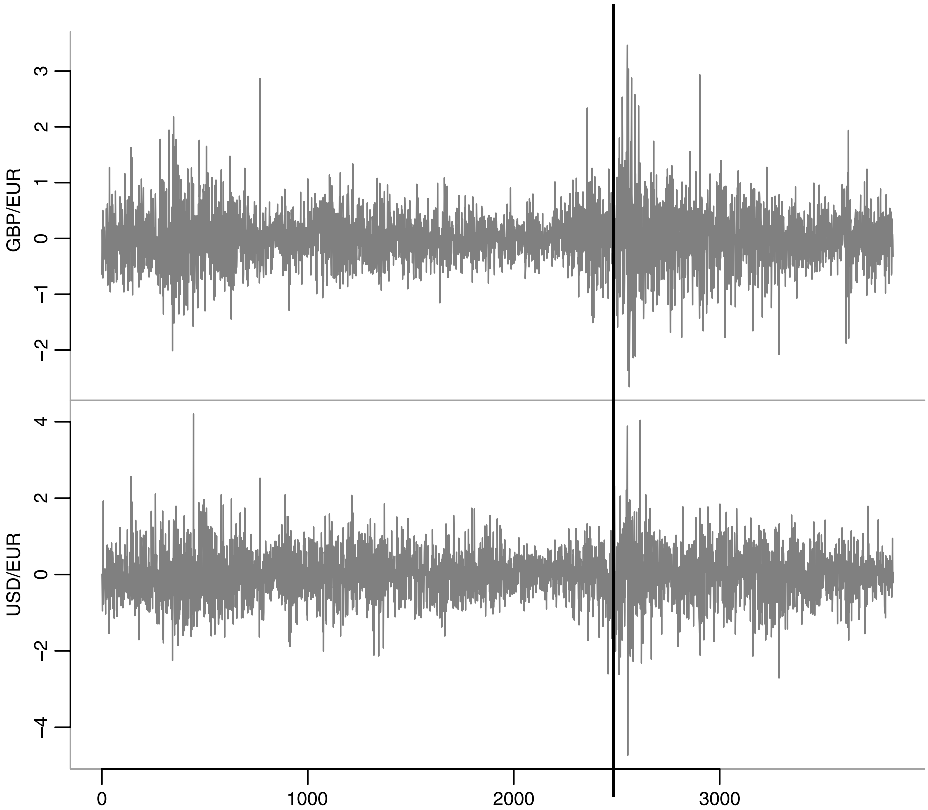

Fig. 1. Plot of data. Note: the graph presents the daily logarithmic rates of returns of two exchange rates, expressed in percentage points, namely the British pound and the US dollar to the Euro. The data span the period from 4 January 1999 to 31 December 2013. The vertical black line marks the day Lehman Brothers filed for Chapter 11 bankruptcy protection, i.e. 15 September 2008.

different assumptions regarding the prior distribution. The first prior distribution, referred to as the diffuse prior, is as specified in Section 2 and Eq. (12) with the value of the hyper-parameter $\bar{s}$ set to 100. The second assumed prior distribution, referred to as the shrinkage distribution, has the value of this hyper-parameter set to 0.1. The estimated models are as follows. The unrestricted model defined by Eqs. (2)–(4) allows for second-order causality in both directions, from GBP/EUR to USD/EUR and the reverse. Restricted models represent different hypotheses of noncausality, and are restricted according to the conditions stated in Theorem 1. All of the models, both the unrestricted and the restricted, are estimated twice with the two different prior distributions.

Three hypotheses of second-order noncausality are investigated. The restrictions resulting from Theorem 1 for the hypothesis of second-order noncausality from the British pound to the US dollar, denoted by GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR, are presented in Example 1, and are given by:

$$A_{21} = 0 \quad \text{and} \quad B_{21}A_{11} = 0,$$

while the restrictions for the hypothesis of second-order noncausality from the dollar to the pound, denoted by USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR, are:

$$A_{12} = 0 \quad \text{and} \quad B_{12}A_{22} = 0.$$

The third hypothesis of second-order noncausality in both directions results in the restrictions being a logical conjunction of the two restrictions presented above.

The strategy for the assessment of the hypotheses is as follows. For each of the hypotheses, a full set of sufficient conditions for the restrictions representing the hypothesis is derived. The sufficient conditions are in the form of zero restrictions imposed on individual parameters. All of the restricted models are estimated and the respective marginal distributions of data are computed. Posterior probabilities of both the models and the hypotheses are computed. The hypotheses are compared using posterior odds ratios. Table 3 presents all of the hypotheses and the models that represent them, with restrictions being necessary and sufficient conditions resulting from Theorem 1 for each of the hypotheses.

We begin our analysis of the results with several comments on the parameters of the unrestricted models. Table 2 reports the posterior means and the posterior standard deviations of the parameters of these models, estimated with the two different prior assumptions. Note that there are no significant differences between the values of the parameters in these two models, and thus, all of the following comments concern both of the specifications. The unrestricted model with a skew t distribution was also estimated for both of the prior specifications. While the extension improved the in-sample fit of the models (the logarithms of base 10 of the Bayes factors of the extended model to the symmetric one were 3.34 and 2.12 for the diffuse and shrinkage priors respectively), it did not change

Please cite this article in press as: Woźniak, T., Testing causality between two vectors in multivariate GARCH models. International Journal of Forecasting (2015), http://dx.doi.org/10.1016/j.ijforecast.2015.01.005

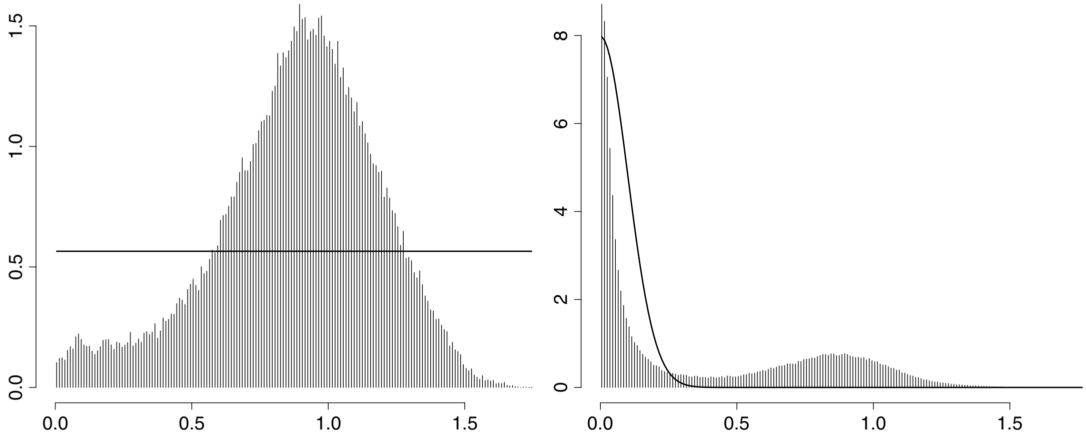

Fig. 2. Marginal posterior densities of the parameter B_{21} for diffuse and shrinkage prior distributions. Note: the marginal posterior densities of B_{21} for the two prior specifications (diffuse on the LHS and shrinkage on the RHS) represent the densities for the models that do not restrict B_{21} to zero, weighted by the posterior probabilities of the models: $p(B_{12}|\mathbf{y}) = \sum_i \Pr(\mathcal{M}_i|\mathbf{y})p(B_{21}|\mathbf{y}, \mathcal{M}_i)$, for i such that parameter B_{21} is not restricted to zero in model $\mathcal{M}_i$. The black lines are the marginal prior distributions of these parameters.

Table 2
Summary of the estimation of the unrestricted VAR(1)-ECCC-GARCH(1, 1) models for the sample from September 16, 2008, to December 31, 2013.

	VAR(1)			*GARCH*(1, 1)					ν
	α_0	α_1		ω	A		B		ρ_{12}
Panel A: estimation results for model $\mathcal{M}_0$ with a diffuse prior distribution									
GBP/EUR	0.002	0.044	−0.031	0.003	0.055	0.009	0.557	0.230	9.072
	(0.012)	(0.032)	(0.024)	(0.003)	(0.023)	(0.007)	(0.285)	(0.174)	(1.443)
USD/EUR	0.013	0.032	−0.044	0.013	0.053	0.040	0.698	0.460	0.490
	(0.016)	(0.040)	(0.031)	(0.020)	(0.034)	(0.020)	(0.441)	(0.305)	(0.022)
Panel B: estimation results for model $\mathcal{M}_0$ with a shrinkage prior distribution									
GBP/EUR	0.002	0.043	−0.031	0.002	0.053	0.009	0.599	0.206	9.043
	(0.012)	(0.031)	(0.024)	(0.003)	(0.021)	(0.006)	(0.282)	(0.173)	(1.404)
USD/EUR	0.014	0.031	−0.043	0.014	0.054	0.040	0.724	0.443	0.491
	(0.016)	(0.039)	(0.031)	(0.020)	(0.036)	(0.020)	(0.442)	(0.302)	(0.023)

Note: The table summarizes the estimation of the VAR(1)-ECCC-GARCH(1, 1) model described by Eqs. (2)–(4) and the likelihood function in Eq. (6). The prior distributions are specified as in Eq. (12). The posterior means and the posterior standard deviations (in brackets) of the parameters are reported. Summaries of the characteristics of the simulations of the posterior densities of the parameters for all of the models are reported in Tables 8 and 9.

the parameter estimates. Moreover, the asymmetry parameter has a high posterior density for the value zero, corresponding to the symmetric case. We therefore proceed with the analysis using symmetrically t-distributed error terms.

None of the parameters of the vector autoregressive part can be considered to be statistically different from zero. This finding proves that the two exchange rates do not Granger cause each other in conditional means. All of the parameters of the GARCH(1, 1) part are assumed to be nonnegative. However, a significant part of the posterior probability mass is concentrated around zero for most of parameters of the matrices ω, A and B. The parameter B_{21} is of particular interest and has a very high posterior mean, around 0.7, but has a large posterior standard deviation and a non-negligible posterior probability mass concentrated around zero for both the diffuse and shrinkage prior distributions. This parameter is responsible for the volatility transmission from the lagged value of the conditional variance of the GBP/EUR exchange rate to the current conditional variance of the variable USD/EUR.

Fig. 2 presents its marginal posterior densities for both assumed prior densities. These posterior densities are integrated over all of the models for which the parameter B_{21} is not set to zero, and the model-specific densities are weighted with the corresponding posterior probabilities of the models (and thus, this analysis does not correspond fully to the analysis of the posterior means for the unrestricted model). The posterior densities differ for both of the prior specifications. The introduction of a shrinkage prior distribution does indeed result in the posterior probability mass being concentrated more around zero than for the diffuse specification. The distribution for the shrinkage prior specification is bimodal, with a part of the posterior probability mass near zero and a part concentrated around 0.8. The distribution for the diffuse prior specification is unimodal, with the mode around 0.9, but is also very wide-spread, with a non-negligible part near zero. These findings are reflected in the results of the assessment of the hypotheses of second-order noncausality.

4.3. Assessment of hypotheses

The credibility of the hypotheses of second-order noncausality between the exchange rates of the British pound

Table 3
Marginal densities of data for the models for the sample from September 16, 2008, to December 31, 2013.

$\mathcal{M}_j$	Restrictions	$\ln p(\mathbf{y}\|\mathcal{M}_j)$	
		Diffuse prior	Shrinkage prior
	$\mathcal{H}_0$: unrestricted model		
$\mathcal{M}_0$	–	−2294.88	−2315.44
	$\mathcal{H}_1$: GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR		
$\mathcal{M}_1$	$A_{21} = B_{21} = 0$	−2305.00	−2320.28
$\mathcal{M}_2$	$A_{11} = A_{21} = 0$	−2328.03	−2332.40
$\mathcal{M}_3$	$A_{11} = A_{21} = B_{21} = 0$	−2325.51	−2338.11
	$\mathcal{H}_2$: USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR		
$\mathcal{M}_4$	$A_{12} = B_{12} = 0$	−2301.21	−2317.19
$\mathcal{M}_5$	$A_{12} = A_{22} = 0$	−2309.19	−2316.59
$\mathcal{M}_6$	$A_{12} = A_{22} = B_{12} = 0$	−2304.21	−2316.72
	$\mathcal{H}_3$: GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR and USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR		
$\mathcal{M}_7$	$A_{12} = A_{21} = B_{12} = B_{21} = 0$	−2302.18	−2313.70
$\mathcal{M}_8$	$A_{12} = A_{21} = A_{22} = B_{21} = 0$	−2365.10	−2374.51
$\mathcal{M}_9$	$A_{12} = A_{21} = A_{22} = B_{12} = B_{21} = 0$	−2357.19	−2361.12
$\mathcal{M}_{10}$	$A_{12} = A_{21} = A_{11} = B_{12} = 0$	−2385.66	−2394.22
$\mathcal{M}_{11}$	$A_{12} = A_{21} = A_{11} = A_{22} = 0$	−2391.30	−2404.153
$\mathcal{M}_{12}$	$A_{12} = A_{21} = A_{11} = A_{22} = B_{12} = 0$	−2390.14	−2398.00
$\mathcal{M}_{13}$	$A_{12} = A_{21} = A_{11} = B_{12} = B_{21} = 0$	−2389.87	−2397.54
$\mathcal{M}_{14}$	$A_{12} = A_{21} = A_{11} = A_{22} = B_{21} = 0$	−2403.37	−2410.06
$\mathcal{M}_{15}$	$A_{12} = A_{21} = A_{11} = A_{22} = B_{12} = B_{21} = 0$	−2421.07	−2426.27

Note: The estimator used for the marginal densities of data is that of Chib and Jeliazkov (2001).

Table 4
Summary of hypothesis testing for the sample from September 16, 2008, to December 31, 2013.

$\mathcal{H}_i$	Hypothesis	Models	$\log_{10}\frac{\Pr(\mathcal{H}_i\|\mathbf{y})}{\Pr(\mathcal{H}_0\|\mathbf{y})}$	
			Diffuse prior	Shrinkage prior
$\mathcal{H}_0$	Unrestricted model	$\mathcal{M}_0$	0.00	0.00
$\mathcal{H}_1$	GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR	$\mathcal{M}_1$–$\mathcal{M}_3$	−4.40	−2.10
$\mathcal{H}_2$	USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR	$\mathcal{M}_4$–$\mathcal{M}_6$	−2.73	−0.11
$\mathcal{H}_3$	GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR and USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR	$\mathcal{M}_7$–$\mathcal{M}_{15}$	−3.17	0.76

Note: The posterior probabilities of the hypotheses were computed using Eq. (27), which uses posterior probabilities of models, as in Eq. (26), and assuming a uniform prior distribution for the hypotheses: $\Pr(\mathcal{H}_i) = 0.25$, which implies a uniform improper prior distribution for the models $p(\mathcal{M}_j) = 0.25/\sum_{j=0}^{15}\mathcal{I}(\mathcal{M}_j \in \mathcal{H}_i)$.

and the US dollar to Euro is evaluated. Altogether, four different hypotheses are formed and assessed within the framework of the VAR-GARCH model. Due to the strategy adopted – testing the full set of sufficient conditions for the original restrictions – some of the hypotheses are represented by several models. Table 3 summarizes the hypotheses, the models representing them, and the restrictions on the parameters according to which the models are restricted. Moreover, the table reports natural logarithms of the marginal densities of the data for each of the models and for both assumed prior distributions.

According to Table 3, the results for model selection are not consistent over different prior specifications. For the diffuse prior specification, the model supported best by the data is model $\mathcal{M}_0$. This model allows for unconstrained volatility transmissions in both directions. Consequently, it represents hypothesis $\mathcal{H}_0$ of Granger causality in either direction. A contradictory result is found when the shrinkage prior specification is considered. In this case, model $\mathcal{M}_7$ is supported best by the data. This model corresponds to the no-volatility-transmissions hypothesis, as it has both matrices, A and B, being diagonal matrices. It implies second-order noncausality from dollar to pound and from pound to dollar, taken jointly, i.e., hypothesis $\mathcal{H}_3$. In effect, these two hypotheses are expected to have the highest posterior probabilities, given the prior distribution class.

Table 4 confirms this claim, by providing posterior odds ratios of the hypotheses. Consider first the hypotheses assessed with the diffuse prior specification. The logarithms of base 10 of the posterior odds ratios for all of the hypotheses compared to the null hypothesis have negative values, which means that hypothesis $\mathcal{H}_0$ has the highest value of the posterior probability. This value is over 500 times larger $\left(10^{2.73}\right)$ than the posterior probability of the second best hypothesis, $\mathcal{H}_2$, and nearly 1500 times larger $\left(10^{3.17}\right)$ than the hypothesis of second-order noncausality in both directions, $\mathcal{H}_3$.

The assessment of the hypotheses for the shrinkage prior distributions is not as sound as in the previous case. Hypothesis $\mathcal{H}_3$ has a posterior probability that is only 5.8 times larger $\left(10^{0.76}\right)$ than hypothesis $\mathcal{H}_0$. Therefore, the hypothesis that is not supported strongly by the prior distribution (which shrinks the distribution of the parameters towards no volatility transmissions) gains a non-negligible posterior probability. Note that, for all of the models, the value of the logarithm of the marginal data density is greater for models with diffuse prior distributions. This is a strong argument in favor of using the results for the diffuse prior specification to form the main findings. This is

Table 5
Summary of the estimation of the unrestricted VAR(1)-ECCC-GARCH(1,1) models for the sample from January 5, 1999, to September 15, 2008.

	VAR(1)			*GARCH*(1, 1)					ν
	α_0	α_1		ω	A		B		ρ_{12}
Panel A: estimation results for model $\mathcal{M}_0$ with a diffuse prior distribution									
GBP/EUR	−0.000	0.068	−0.045	0.002	0.054	0.007	0.888	0.015	10.741
	(0.008)	(0.024)	(0.016)	(0.001)	(0.012)	(0.004)	(0.030)	(0.011)	(1.382)
USD/EUR	0.009	0.050	−0.021	0.007	0.008	0.023	0.019	0.946	0.548
	(0.011)	(0.034)	(0.024)	(0.005)	(0.007)	(0.007)	(0.043)	(0.033)	(0.015)
Panel B: estimation results for model $\mathcal{M}_0$ with a shrinkage prior distribution									
GBP/EUR	−0.000	0.067	−0.044	0.002	0.054	0.007	0.889	0.015	10.805
	(0.007)	(0.024)	(0.016)	(0.001)	(0.012)	(0.004)	(0.032)	(0.012)	(1.416)
USD/EUR	0.009	0.049	−0.021	0.006	0.007	0.023	0.017	0.948	0.547
	(0.011)	(0.034)	(0.024)	(0.004)	(0.007)	(0.006)	(0.030)	(0.027)	(0.015)

Note: The table summarizes the estimation of the VAR(1)-ECCC-GARCH(1, 1) model described by Eqs. (2)–(4) and the likelihood function in Eq. (6). The prior distributions are specified as in Eq. (12). The posterior means and posterior standard deviations (in brackets) of the parameters are reported. Summaries of the characteristics of the simulations of the posterior densities of the parameters for each of the models are reported in Tables 10 and 11.

Table 6
Marginal densities of data for the models for the sample from January 5, 1999, to September 15, 2008.

$\mathcal{M}_j$	Restrictions	$\ln p(\mathbf{y}\mid\mathcal{M}_j)$	
		Diffuse prior	Shrinkage prior
	$\mathcal{H}_0$: unrestricted model		
$\mathcal{M}_0$	–	−3074.00	−3098.16
	$\mathcal{H}_1$: GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR		
$\mathcal{M}_1$	$A_{21} = B_{21} = 0$	−3072.50	−3090.71
$\mathcal{M}_2$	$A_{11} = A_{21} = 0$	−3107.51	−3121.50
$\mathcal{M}_3$	$A_{11} = A_{21} = B_{21} = 0$	−3114.11	−3126.40
	$\mathcal{H}_2$: USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR		
$\mathcal{M}_4$	$A_{12} = B_{12} = 0$	−3079.50	−3095.38
$\mathcal{M}_5$	$A_{12} = A_{22} = 0$	−3083.93	−3100.93
$\mathcal{M}_6$	$A_{12} = A_{22} = B_{12} = 0$	−3085.96	−3097.62
	$\mathcal{H}_3$: GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR and USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR		
$\mathcal{M}_7$	$A_{12} = A_{21} = B_{12} = B_{21} = 0$	−3077.35	−3084.54
$\mathcal{M}_8$	$A_{12} = A_{21} = A_{22} = B_{21} = 0$	−3121.15	−3130.50
$\mathcal{M}_9$	$A_{12} = A_{21} = A_{22} = B_{12} = B_{21} = 0$	−3117.84	−3125.77
$\mathcal{M}_{10}$	$A_{12} = A_{21} = A_{11} = B_{12} = 0$	−3189.61	−3199.45
$\mathcal{M}_{11}$	$A_{12} = A_{21} = A_{11} = A_{22} = 0$	−3176.50	−3207.90
$\mathcal{M}_{12}$	$A_{12} = A_{21} = A_{11} = A_{22} = B_{12} = 0$	−3197.72	−3204.55
$\mathcal{M}_{13}$	$A_{12} = A_{21} = A_{11} = B_{12} = B_{21} = 0$	−3188.19	−3196.60
$\mathcal{M}_{14}$	$A_{12} = A_{21} = A_{11} = A_{22} = B_{21} = 0$	−3203.55	−3211.73
$\mathcal{M}_{15}$	$A_{12} = A_{21} = A_{11} = A_{22} = B_{12} = B_{21} = 0$	−3203.95	−3208.57

Note: The estimator used for the marginal densities of data is that of Chib and Jeliazkov (2001).

important because the overall rank of the credibility of the hypotheses is not robust to the specification of the prior distributions for the parameters of the model. However, most of these findings are confirmed when the estimator of the marginal data density of Geweke (1999, 2005) is used. These results are available on request.

The most important finding of the empirical analysis is that the hypothesis of second-order causality in either direction is supported by the data. That confirms the impression obtained based on the large values of the posterior means of the volatility spillover parameters reported in Table 3.

4.4. Robustness with respect to sample selection

In order to investigate second-order causality between the exchange rates further, an analysis of the sample from before the crisis (from January 4, 1999, to September 15, 2008) was performed. Table 5 documents the fact that there are significant differences between the estimates of the posterior means in these two samples. The main difference is in the estimates of the parameters of the matrix B. For the sample before the crisis, the posterior mass of the off-diagonal parameters is shifted towards zero, and the posterior means are much smaller than their corresponding values for the sample after the breakpoint. At the same time, the posterior means of the diagonal parameters, which are responsible for the persistence of their own conditional variances, are much higher than for the sample analyzed initially.

In this sample, the results for the shrinkage prior specification are the same as in the sample analyzed earlier. According to Table 6, the model supported best by the data is model $\mathcal{M}_7$, representing hypothesis $\mathcal{H}_3$. However, this hypothesis now dominates the other hypotheses strongly, with the value of the logarithm of base 10 of its posterior probability compared to the posterior probability of the null hypothesis being equal to 5.91 (see Table 7). Ap-

Table 7
Summary of hypothesis testing for the sample from January 5, 1999, to September 15, 2008.

$\mathcal{H}_i$	Hypothesis	Models	$\log_{10}\frac{\Pr(\mathcal{H}_i\vert\mathbf{y})}{\Pr(\mathcal{H}_0\vert\mathbf{y})}$	
			Diffuse prior	Shrinkage prior
$\mathcal{H}_0$	Unrestricted model	$\mathcal{M}_0$	0.00	0.00
$\mathcal{H}_1$	GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR	$\mathcal{M}_1$–$\mathcal{M}_3$	0.65	3.24
$\mathcal{H}_2$	USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR	$\mathcal{M}_4$–$\mathcal{M}_6$	−2.38	1.25
$\mathcal{H}_3$	GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR and USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR	$\mathcal{M}_7$–$\mathcal{M}_{15}$	−1.46	5.91

Note: The posterior probabilities of the hypotheses were computed using a formula of Eq. (27) that uses posterior probabilities of models as in Eq. (26), and assuming flat prior distributions for hypotheses: $\Pr(\mathcal{H}_i) = 0.25$, which implies a uniform improper prior distribution for the models $p(\mathcal{M}_j) = 0.25/\sum_{j=0}^{15}\mathcal{I}(\mathcal{M}_j \in \mathcal{H}_i)$.

Table 8
Properties of the simulations of the posterior densities of the models with diffuse prior distributions for the sample from September 16, 2008, to December 31, 2013.

Model	RNE			Autocorrelation at lag 1			Autocorrelation at lag 10			Geweke's z			S
	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	
$\mathcal{M}_0$	0.014	0.002	0.045	0.950	0.908	0.987	0.632	0.400	0.906	−0.349	−1.913	2.816	70 000
$\mathcal{M}_1$	0.064	0.002	0.122	0.755	0.590	0.941	0.160	0.070	0.735	−0.446	−1.142	1.248	50 000
$\mathcal{M}_2$	0.006	0.001	0.011	0.976	0.961	0.992	0.844	0.692	0.936	0.116	−0.763	3.617	50 000
$\mathcal{M}_3$	0.082	0.015	0.138	0.813	0.722	0.955	0.191	0.064	0.704	0.078	−1.195	1.135	50 000
$\mathcal{M}_4$	0.047	0.002	0.174	0.767	0.726	0.987	0.169	0.081	0.925	0.219	−1.414	1.765	50 000
$\mathcal{M}_5$	0.008	0.001	0.032	0.962	0.879	0.995	0.706	0.534	0.973	0.242	−4.881	4.134	50 000
$\mathcal{M}_6$	0.077	0.041	0.161	0.843	0.468	0.916	0.223	0.083	0.502	0.398	−1.008	2.398	50 000
$\mathcal{M}_7$	0.010	0.004	0.052	0.973	0.804	0.986	0.786	0.317	0.877	0.162	−1.613	1.770	50 000
$\mathcal{M}_8$	0.149	0.044	0.213	0.717	0.671	0.806	0.052	0.017	0.226	0.012	−1.457	2.071	50 000
$\mathcal{M}_9$	0.005	0.001	0.138	0.914	0.495	0.997	0.645	0.038	0.980	0.008	−1.567	1.895	50 000
$\mathcal{M}_{10}$	0.039	0.001	0.096	0.842	0.665	0.990	0.349	0.124	0.955	0.450	−2.453	2.734	50 000
$\mathcal{M}_{11}$	0.020	0.002	0.181	0.956	0.560	0.979	0.647	0.018	0.863	−0.038	−2.084	1.514	50 000
$\mathcal{M}_{12}$	0.039	0.009	0.326	0.828	0.380	0.954	0.281	0.015	0.725	0.310	−2.029	2.484	45 000
$\mathcal{M}_{13}$	0.073	0.052	0.123	0.849	0.591	0.891	0.235	0.074	0.339	0.721	−2.931	3.693	50 000
$\mathcal{M}_{14}$	0.016	0.005	0.293	0.910	0.297	0.954	0.438	0.017	0.754	0.669	−3.302	3.958	50 000
$\mathcal{M}_{15}$	0.205	0.017	0.347	0.632	0.613	0.934	0.022	0.014	0.658	−0.269	−0.925	0.645	50 000
$\mathcal{M}_a$	0.004	0.001	0.012	0.966	0.856	0.993	0.806	0.486	0.968	−0.074	−4.233	2.452	50 000

Note: The table summarizes the properties of the numerical simulation of the posterior densities of all of the models considered. For each of the statistics, the median of all of the parameters of the model is reported, together with the minimum and maximum. The table reports the relative numerical efficiency coefficient (RNE; Geweke, 1989), autocorrelations of the MCMC draws at lags 1 and 10, and Geweke's z scores for the hypothesis of equal means of the first 10% and the last 50% of draws that follow the standard normal distribution (see Geweke, 1992).

parently, the shrinkage effect dominates and has a deciding effect on the assessment of the hypotheses. When the diffuse prior specification is considered in this sample, the model supported best by the data is model $\mathcal{M}_1$, which does not allow for the volatility transmissions from GBP/EUR to USD/EUR. Consequently, hypothesis $\mathcal{H}_1$ gains most of the posterior probability mass. Note, however, that the posterior probability of hypothesis $\mathcal{H}_0$ is only around 4.5 times smaller $\left(10^{-0.65}\right)$, and gains a non-negligible posterior probability mass. The models with the diffuse prior distribution also have higher values of the marginal data densities in this sample than the models with the shrinkage prior structure.

To summarise, the main finding of the empirical investigation is that the British pound to Euro exchange rate second-order Granger causes the US dollar to Euro exchange rate, and the converse also holds. The evidence is delivered by the analysis of models with diffuse (and proper) prior distributions. A non-negligible posterior probability mass is assigned to this hypothesis for the whole available sample. However, this finding should be qualified. This can be done best by analyzing the marginal posterior distribution of the parameter B_{21} (see e.g. Fig. 2). For the sample during and after the financial crisis, this parameter has a very high value of the posterior mean, which supports the idea of volatility transmissions from GBP/EUR to USD/EUR. This result confirms the analysis of Woźniak (2012), who found second-order causality in the same direction analyzed here in a system of three exchange rates (the two used in this paper and the Swiss frank to Euro exchange rate) for a sample from 16 September 2008 to 22 September 2011. Woźniak (2012) explained this phenomenon based on the *meteor shower* hypothesis of Engle, Ito, and Lin (1990).[1] However, in the bivariate system of

[1] The *meteor shower* hypothesis links the hours of trading activity to the structure of the forecasting model of volatility. Despite the fact that the exchange rate market is open 24 hours a day, traders on different continents are active mainly during their working hours. Over one day, agents in Australia and Asia are active first, then agents in Europe (and Africa) become active; and finally, traders in both Americas start working. Therefore, coming back to our example, on a given day, agents in Europe trade between the Euro zone and the United Kingdom first, and it is only later, when working hours in the United States commence, that agents start trading between the Euro zone and the USA. Such a pattern is captured by the models that represent the hypothesis of second-order noncausality from the pound to the dollar. Note that the triangular GARCH model of Engle et al. (1990), which represents the meteor shower hypothesis, is just one of the models, namely $\mathcal{M}_4$, representing hypothesis $\mathcal{H}_2$.

Table 9
Properties of the simulations of the posterior densities of the models with shrinkage prior distributions for the sample from September 16, 2008, to December 31, 2013.

Model	RNE			Autocorrelation at lag 1			Autocorrelation at lag 10			Geweke's z			S
	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	
$\mathcal{M}_0$	0.018	0.002	0.039	0.949	0.866	0.988	0.624	0.408	0.906	0.293	−1.511	1.385	70 000
$\mathcal{M}_1$	0.073	0.002	0.126	0.763	0.608	0.955	0.184	0.080	0.797	−0.089	−2.086	1.663	50 000
$\mathcal{M}_2$	0.006	0.001	0.014	0.979	0.965	0.993	0.836	0.717	0.955	0.359	−3.281	3.230	50 000
$\mathcal{M}_3$	0.071	0.013	0.152	0.812	0.721	0.959	0.216	0.058	0.734	0.003	−1.103	1.191	50 000
$\mathcal{M}_4$	0.055	0.001	0.220	0.743	0.696	0.994	0.098	0.034	0.959	0.191	−1.978	1.189	50 000
$\mathcal{M}_5$	0.009	0.001	0.027	0.964	0.919	0.995	0.735	0.504	0.969	0.093	−3.280	3.186	50 000
$\mathcal{M}_6$	0.088	0.032	0.160	0.847	0.491	0.910	0.217	0.080	0.485	0.294	−2.694	1.550	50 000
$\mathcal{M}_7$	0.010	0.004	0.026	0.972	0.819	0.986	0.771	0.404	0.873	0.022	−1.168	1.366	50 000
$\mathcal{M}_8$	0.119	0.060	0.253	0.727	0.647	0.819	0.070	0.025	0.264	0.229	−0.720	0.759	50 000
$\mathcal{M}_9$	0.003	0.001	0.186	0.927	0.506	0.999	0.720	0.030	0.994	−2.754	−18.835	11.205	50 000
$\mathcal{M}_{10}$	0.041	0.001	0.106	0.842	0.661	0.994	0.339	0.097	0.972	0.189	−3.875	2.721	50 000
$\mathcal{M}_{11}$	0.017	0.002	0.142	0.959	0.589	0.986	0.672	0.044	0.915	−0.066	−2.836	2.573	50 000
$\mathcal{M}_{12}$	0.049	0.011	0.374	0.823	0.385	0.954	0.274	-0.004	0.713	0.216	−1.487	2.370	50 000
$\mathcal{M}_{13}$	0.079	0.051	0.148	0.852	0.598	0.890	0.234	0.073	0.334	0.390	−0.463	1.652	50 000
$\mathcal{M}_{14}$	0.018	0.003	0.265	0.913	0.312	0.956	0.465	0.010	0.772	0.189	−2.903	2.844	50 000
$\mathcal{M}_{15}$	0.186	0.018	0.266	0.653	0.612	0.838	0.028	0.011	0.393	−0.258	−1.165	1.852	50 000
$\mathcal{M}_a$	0.004	0.002	0.014	0.967	0.869	0.992	0.793	0.509	0.948	0.006	−2.018	2.304	50 000

Note: The table summarizes the properties of the numerical simulation of the posterior densities of all of the models considered. For each of the statistics, the median of all the parameters of the model is reported, together with the minimum and maximum. The table reports the relative numerical efficiency coefficient (RNE; Geweke, 1989), autocorrelations of the MCMC draws at lags 1 and 10, and Geweke's z scores for the hypothesis of equal means of the first 10% and the last 50% of draws that follow the standard normal distribution (see Geweke, 1992).

Table 10
Properties of the simulations of the posterior densities of the models with diffuse prior distributions for the sample from January 5, 1999, to September 15, 2008.

Model	RNE			Autocorrelation at lag 1			Autocorrelation at lag 10			Geweke's z			S
	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	
$\mathcal{M}_0$	0.053	0.003	0.102	0.827	0.747	0.963	0.269	0.128	0.870	−0.243	−2.006	1.359	50 000
$\mathcal{M}_1$	0.045	0.002	0.087	0.827	0.660	0.967	0.320	0.112	0.822	0.124	−1.276	1.774	50 000
$\mathcal{M}_2$	0.019	0.002	0.045	0.927	0.900	0.980	0.499	0.391	0.884	0.830	−1.813	1.653	50 000
$\mathcal{M}_3$	0.047	0.003	0.132	0.867	0.737	0.978	0.324	0.120	0.892	−0.205	−3.748	3.112	50 000
$\mathcal{M}_4$	0.046	0.001	0.125	0.816	0.789	0.999	0.167	0.121	0.995	0.287	−1.954	2.417	50 000
$\mathcal{M}_5$	0.135	0.078	0.284	0.728	0.674	0.808	0.054	0.036	0.200	0.142	−2.209	2.098	50 000
$\mathcal{M}_6$	0.057	0.009	0.115	0.875	0.743	0.969	0.309	0.075	0.775	0.703	−0.958	1.832	50 000
$\mathcal{M}_7$	0.154	0.079	0.292	0.742	0.559	0.791	0.084	0.038	0.117	−0.053	−0.878	2.025	50 000
$\mathcal{M}_8$	0.061	0.008	0.144	0.868	0.709	0.969	0.293	0.052	0.759	0.008	−0.875	1.254	50 000
$\mathcal{M}_9$	0.168	0.001	0.396	0.663	0.647	0.996	0.059	0.007	0.967	0.107	−3.904	4.194	50 000
$\mathcal{M}_{10}$	0.131	0.015	0.218	0.711	0.665	0.884	0.047	0.010	0.496	−0.475	−1.836	1.238	50 000
$\mathcal{M}_{11}$	0.018	0.001	0.072	0.880	0.770	1.000	0.336	0.188	0.999	−0.772	−3.392	3.170	65 600
$\mathcal{M}_{12}$	0.161	0.006	0.273	0.717	0.678	0.833	0.048	0.015	0.564	0.367	−1.326	1.078	50 000
$\mathcal{M}_{13}$	0.163	0.100	0.254	0.709	0.644	0.793	0.039	0.018	0.207	−0.685	−1.532	1.616	50 000
$\mathcal{M}_{14}$	0.139	0.001	0.213	0.672	0.636	0.998	0.076	0.008	0.991	0.201	−0.429	0.863	50 000
$\mathcal{M}_{15}$	0.118	0.002	0.206	0.685	0.630	0.970	0.062	0.020	0.839	0.786	−1.827	2.686	50 000
$\mathcal{M}_a$	0.012	0.002	0.022	0.912	0.764	0.962	0.595	0.325	0.809	0.729	−2.123	3.008	43 500

Note: The table summarizes the properties of the numerical simulation of the posterior densities of all of the models considered. For each of the statistics, the median of all of the parameters of the model is reported, together with the minimum and maximum. The table reports the relative numerical efficiency coefficient (RNE; Geweke, 1989), autocorrelations of the MCMC draws at lags 1 and 10, and Geweke's z scores for the hypothesis of equal means of the first 10% and the last 50% of draws that follow the standard normal distribution (see Geweke, 1992).

the GBP/EUR and USD/EUR considered in the current paper, the meteor shower effect is only temporary. There is strong evidence that the effect was not present before September 2008, as the posterior density of the parameter of the transition from GBP/EUR to USD/EUR was concentrated around zero.

Therefore, it would appear to be correct to conclude that the global financial crisis and its aftermath strengthened the dynamic relationships between the volatilities of exchange rates. This phenomenon could be described more accurately using a model that allowed for endogenous changes in the parameter values, such as a Markov-switching multivariate GARCH model. Structural changes that are determined endogenously within the model are necessary in order to determine the change points correctly. However, the second-order causality analysis based on the linear theory of causality is not suitable for the analysis in such models.

5. Conclusions

In this work, the conditions for analyzing Granger noncausality for the second conditional moments of a GARCH process are derived. The restrictions presented for one-

Please cite this article in press as: Woźniak, T., Testing causality between two vectors in multivariate GARCH models. International Journal of Forecasting (2015), http://dx.doi.org/10.1016/j.ijforecast.2015.01.005

Table 11
Properties of the simulations of the posterior densities of the models with shrinkage prior distributions for the sample from January 5, 1999, to September 15, 2008.

Model	RNE			Autocorrelation at lag 1			Autocorrelation at lag 10			Geweke's z			S
	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	
$\mathcal{M}_0$	0.059	0.004	0.100	0.843	0.774	0.921	0.233	0.155	0.733	0.365	−2.477	2.797	54 200
$\mathcal{M}_1$	0.036	0.002	0.093	0.819	0.651	0.971	0.318	0.075	0.844	−0.624	−2.530	3.076	50 000
$\mathcal{M}_2$	0.031	0.001	0.064	0.907	0.868	0.984	0.413	0.321	0.918	0.242	−2.233	2.275	50 000
$\mathcal{M}_3$	0.043	0.003	0.088	0.866	0.757	0.980	0.321	0.146	0.897	0.209	−0.944	1.366	50 000
$\mathcal{M}_4$	0.049	0.010	0.104	0.873	0.611	0.934	0.291	0.114	0.664	−0.042	−2.450	2.061	50 000
$\mathcal{M}_5$	0.132	0.080	0.230	0.721	0.675	0.771	0.057	0.016	0.128	−0.175	−2.149	1.650	50 000
$\mathcal{M}_6$	0.168	0.121	0.315	0.701	0.672	0.756	0.042	0.025	0.101	−0.078	−0.900	1.241	50 000
$\mathcal{M}_7$	0.009	0.003	0.016	0.974	0.949	0.984	0.798	0.660	0.892	−0.115	−2.062	1.456	50 000
$\mathcal{M}_8$	0.049	0.008	0.162	0.873	0.736	0.973	0.341	0.070	0.780	0.452	−1.860	1.908	50 000
$\mathcal{M}_9$	0.161	0.002	0.264	0.651	0.628	0.996	0.038	0.003	0.962	−0.417	−4.551	4.270	50 000
$\mathcal{M}_{10}$	0.130	0.056	0.212	0.705	0.663	0.844	0.053	0.022	0.298	−0.065	−2.002	2.412	50 000
$\mathcal{M}_{11}$	0.013	0.002	0.019	0.962	0.927	0.982	0.691	0.589	0.893	−0.203	−1.853	1.906	41 000
$\mathcal{M}_{12}$	0.211	0.051	0.323	0.656	0.646	0.797	0.032	0.013	0.344	0.118	−1.462	0.956	41 000
$\mathcal{M}_{13}$	0.158	0.072	0.261	0.707	0.663	0.802	0.049	0.021	0.251	−0.189	−1.896	1.466	41 000
$\mathcal{M}_{14}$	0.110	0.002	0.254	0.661	0.619	0.955	0.048	0.009	0.822	−0.491	−2.062	2.078	50 000
$\mathcal{M}_{15}$	0.114	0.001	0.266	0.653	0.614	0.997	0.056	0.014	0.979	0.027	−2.487	1.375	50 000
$\mathcal{M}_a$	0.011	0.001	0.022	0.904	0.730	0.995	0.562	0.215	0.967	0.029	−1.818	1.900	70 000

Note: The table summarizes the properties of the numerical simulation of the posterior densities of all of the models considered. For each of the statistics, the median of all the parameters of the model is reported, together with the minimum and maximum. The table reports the relative numerical efficiency coefficient (RNE; Geweke, 1989), autocorrelations of the MCMC draws at lags 1 and 10, and Geweke's z scores for the hypothesis of equal means of the first 10% and the last 50% of draws that follow the standard normal distribution (see Geweke, 1992).

Table 12
Simulated probabilities $\Pr[|\text{eigenvalue}(A + B)| < 1 | 0 \leq A, B \leq 1.8]$.

	Diffuse prior							
Models	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$	$\mathcal{M}_7$
Probabilities	0.00011	0.02431	0.00602	0.08685	0.02431	0.00606	0.08665	0.02426
Models	$\mathcal{M}_8$	$\mathcal{M}_9$	$\mathcal{M}_{10}$	$\mathcal{M}_{11}$	$\mathcal{M}_{12}$	$\mathcal{M}_{13}$	$\mathcal{M}_{14}$	$\mathcal{M}_{15}$
Probabilities	0.08690	0.08695	0.08686	0.07664	0.31080	0.08688	0.31083	0.31075
	Shrinkage prior							
Models	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$	$\mathcal{M}_7$
Probabilities	0.50851	0.90225	0.82998	0.94856	0.90258	0.82983	0.94856	0.90223
Models	$\mathcal{M}_8$	$\mathcal{M}_9$	$\mathcal{M}_{10}$	$\mathcal{M}_{11}$	$\mathcal{M}_{12}$	$\mathcal{M}_{13}$	$\mathcal{M}_{14}$	$\mathcal{M}_{15}$
Probabilities	0.94846	0.94833	0.94847	0.98428	0.99688	0.94836	0.99688	0.99688

Note: Computer code for the MCMC simulations is available from the author's website: http://bit.ly/tomaszw.

period-ahead second-order noncausality appear to be the restrictions for second-order noncausality at all future horizons, due to the specific setting of the system, in which all of the variables considered belong to one of the two vectors. These conditions may result in several nonlinear restrictions on the parameters of the model, meaning that the available classical tests are of limited use.

Therefore, in order to test these restrictions, the basic Bayesian procedure is applied. This involves estimating the models representing the hypotheses of second-order causality and noncausality, then comparing the models and hypotheses based on posterior odds ratios. This well-known procedure overcomes the difficulties that the classical tests have encountered to date when applied to this problem. The Bayesian inference about the second-order causality between variables is based on the finite-sample analysis. Moreover, although the analysis does not refer to the asymptotic results, the strict assumptions about the existence of the higher-order moments of the series that are required in the asymptotic analysis may be relaxed in the Bayesian inference. In effect, the existence of fourth unconditional moments is assumed for the second-order noncausality analysis, and of second conditional moments for the volatility spillovers analysis.

However, several remarks regarding the proposed approach are in order, due to the fact that all of the variables in the system are divided into only two vectors, between which the causality inference is performed. Within such a setting, not all of the hypotheses of interest may be formulated in a system that contains more than two variables (see Example 3). Another limitation is the fact that the restrictions presented serve as restrictions for second-order noncausality at all future horizons at once. This feature is caused by the particular setting considered in this work. However, the most serious limitation of the proposed second-order causality analysis is that it is not suitable for samples in which the parameters of the model change the values over time.

This critique is a motivation for further research on the topic of Granger causality in second conditional moments. First, one might consider the setting in which the causality between two variables is analyzed, when there are also other variables in the system that might be used for modeling and forecasting. This may be necessary in particular for the analysis of the robustness of the causal or noncausal relationships found, as the values of the parameters in the GARCH models are exposed to the omitted variables problem. Second, the second-order noncausality could be analyzed separately at each of the future horizons. Such a decomposition could provide further insights into causal relationships between economic relationships. Finally, in order to analyze financial time series with the most up-to-date observations, models with time-varying parameters are recommended strongly.

Acknowledgments

This paper was presented at the BMRC-QASS Conference in London in May 2010 and at the FindEcon Conference in Łódź in May 2010 under the title *Bayesian testing of second-order causality*, and at the Multivariate Time Series Modelling and Forecasting Workshop in Melbourne in February 2013 and at Freie Universität Berlin in January 2015 under the current title. The author thanks Helmut Lütkepohl, Jacek Osiewalski, Giampiero Gallo, Helmut Herwartz, Massimiliano Marcellino, Vance Martin and Timo Teräsvirta, as well as the participants of the workshops, for their remarks on the study. He especially thanks the editors and the two anonymous referees for their comments that improved the quality of the paper. Also, the author thanks Michał Markun and colleagues from the EUI Time Series Econometrics Working Group for multiple discussions and suggestions.

Appendix A. Proof

A.1. Proof of Theorem 1

The first part of the proof sets the second-order noncausality restrictions for the GARCH process in the VAR form of Eq. (9). Let $\epsilon_t^{(2)}$ follow a stationary VAR process as in Eq. (9), partitioned as in Eq. (10), that is identifiable. Then, y_1 does not second-order Granger-cause y_2 if and only if:

$$\Pi_{21}(z) \equiv 0 \quad \forall z \in \mathbb{C}. \tag{A.1}$$

The condition in Eq. (A.1) may be proven by the application of Proposition 1 of Boudjellaba et al. (1992). However, several changes are required to adjust the proof of that proposition for the vector autoregressive process to the setting considered in Theorem 1 for the GARCH models. Here, one projects the squared elements of the residual term, $\epsilon^{(2)}(y_{2t+1}|I(t)) = [y_{2t+1} - P(y_{2t+1}|I(t))]^{(2)}$, on the Hilbert spaces $I^2(t)$ or $I^2_{-1}(t)$, both defined in Section 2.

The proven condition still leads to an infinite number of restrictions on the parameters. This property excludes the possibility of testing these restrictions. In order to obtain the simplified condition in Eq. (14), apply the matrix transformations, first of Theorem 1 and then of Theorem 2 of Boudjellaba et al. (1994), to Eq. (A.1).

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2015.01.005.

References

Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*, 1281–1311.

Bauwens, L., & Lubrano, M. (1998). Bayesian inference on GARCH models using the Gibbs sampler. *Econometrics Journal*, *1*, C23–C46.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, *72*, 498–505.

Boudjellaba, H., Dufour, J.-M., & Roy, R. (1992). Testing causality between two vectors in multivariate autoregressive moving average models. *Journal of the American Statistical Association*, *87*, 1082–1090.

Boudjellaba, H., Dufour, J.-M., & Roy, R. (1994). Simplified conditions for noncausality between vectors in multivariate ARMA models. *Journal of Econometrics*, *63*, 271–287.

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

Chib, S., & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association*, *96*, 270–281.

Comte, F., & Lieberman, O. (2000). Second-order noncausality in multivariate GARCH processes. *Journal of Time Series Analysis*, *21*, 535–557.

Comte, F., & Lieberman, O. (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis*, *84*, 61–84.

Conrad, C., & Karanasos, M. (2010). Negative volatility spillovers in the unrestricted ECCC-GARCH model. *Econometric Theory*, *26*, 838–862.

Deschamps, P. J. (2006). A flexible prior distribution for Markov switching autoregressions with Student-*t* errors. *Journal of Econometrics*, *133*, 153–190.

Droumaguet, M., & Woźniak, T. (2012). *Bayesian testing of Granger causality in Markov-switching VARs*. Working paper series, European University Institute, Florence, Italy.

Dufour, J.-M., Pelletier, D., & Renault, E. (2006). Short run and long run causality in time series: Inference. *Journal of Econometrics*, *132*, 337–362.

Dufour, J.-M., & Renault, E. (1998). Short run and long run causality in time series: Theory. *Econometrica*, *66*, 1099–1125.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, *20*, 339–350.

Engle, R. F., Ito, T., & Lin, W.-L. (1990). Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. *Econometrica*, *58*, 525–542.

Fiorentini, G., Sentana, E., & Calzolari, G. (2003). Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with Student *t* innovations. *Journal of Business and Economic Statistics*, *21*, 532–546.

Florens, J. P., & Mouchart, M. (1982). A note on noncausality. *Econometrica*, *50*, 583–591.

Florens, J. P., & Mouchart, M. (1985a). A linear theory for noncausality. *Econometrica*, *53*, 157–176.

Florens, J. P., & Mouchart, M. (1985b). Conditioning in dynamic models. *Journal of Time Series Analysis*, *6*, 15–34.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, *57*, 1317–1339.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics 4, Vol. 148*. Oxford: Clarendon Press.

Geweke, J. (1995). *Bayesian comparison of econometric models*. Working paper, Federal Reserve Bank of Minneapolis.

Geweke, J. (1997). Posterior simulators in econometrics. In D. Kreps, & K. F. Wallis (Eds.), *Advances in economics and econometrics: theory and applications*, vol. 3 (pp. 128–165). Cambridge University Press.

Geweke, J. (1999). Using simulation methods for Bayesian econometric models: Inference, development, and communication. *Econometric Reviews*, *18*, 1–73.

Geweke, J. (2005). *Contemporary Bayesian econometrics and statistics*. John Wiley & Sons, Inc.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*, 424–438.

Hafner, C. M., & Herwartz, H. (2006). A Lagrange multiplier test for causality in variance. *Economics Letters*, *93*, 137–141.

Hafner, C. M., & Herwartz, H. (2008a). Analytical quasi maximum likelihood inference in multivariate volatility models. *Metrika*, *67*, 219–239.

Hafner, C. M., & Herwartz, H. (2008b). Testing for causality in variance using multivariate GARCH models. *Annales d'Économie et de Statistique*, *89*, 215–241.

Hafner, C. M., & Preminger, A. (2009). Asymptotic theory for a factor GARCH model. *Econometric Theory*, *25*, 336–363.

He, C., & Teräsvirta, T. (2004). An extended constant conditional correlation GARCH model and its fourth-moment structure. *Econometric Theory*, *20*, 904–926.

Hoogerheide, L. F., van Dijk, H. K., & van Oest, R. (2009). Simulation based Bayesian econometric inference: Principles and some recent computational advances. In *Handbook of computational econometrics* (Chapter 7, pp. 215–280). Wiley.

Jarociński, M., & Maćkowiak, B. (2013). *Granger-causal-priority and choice of variables in vector autoregressions*. Technical report 16. European Central Bank.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*, 255–259.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, *14*, 70–86.

Karolyi, G. A. (1995). A multivariate GARCH model of international transmissions of stock returns and volatility: The case of the United States and Canada. *Journal of Business and Economic Statistics*, *13*, 11–25.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Kim, T., & White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters*, *1*, 56–73.

Ling, S., & McAleer, M. (2003). Asymtotic theory for a vector ARMA-GARCH model. *Econometric Theory*, *19*, 280–310.

Lomnicki, Z. (1961). Test for departure from normality in the case of linear stochastic processes. *Metrika*, *4*, 37–62.

Lütkepohl, H. (1993). *Introduction to multiple time series analysis*. Springer-Verlag.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

Lütkepohl, H., & Burda, M. M. (1997). Modified Wald tests under nonregular conditions. *Journal of Econometrics*, *78*, 315–332.

Miazhynskaia, T., & Dorffner, G. (2006). A comparison of Bayesian model selection based on MCMC with an application to GARCH-type models. *Statistical Papers*, *47*, 525–549.

Nakatani, T., & Teräsvirta, T. (2009). Testing for volatility interactions in the constant conditional correlation GARCH model. *Econometrics Journal*, *12*, 147–163.

Osiewalski, J., & Pipień, M. (2004). Bayesian comparison of bivariate ARCH-type models for the main exchange rates in Poland. *Journal of Econometrics*, *123*, 371–391.

Pajor, A. (2011). A Bayesian analysis of exogeneity in models with latent variables. *Central European Journal of Economic Modelling and Econometrics*, *3*, 49–73.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*, 7–11.

Robins, R. P., Granger, C. W. J., & Engle, R. F. (1986). Wholesale and retail prices: Bivariate time-series modeling with forecastable error variances. In *Model reliability* (pp. 1–17). The MIT Press.

Silvennoinen, A., & Teräsvirta, T. (2009). Multivariate GARCH models. In T. Mikosch, J.-P. Kreiß, R. A. Davis, & T. G. Andersen (Eds.), *Handbook of financial time series*. Berlin, Heidelberg: Springer.

Sims, C. A. (1972). Money, income, and causality. *The American Economic Review*, *62*, 540–552.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*, 1–48.

Tse, Y. K., & Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business and Economic Statistics*, *20*, 351–362.

Vrontos, I. D., Dellaportas, P., & Politis, D. N. (2003). Inference for some multivariate ARCH and GARCH models. *Journal of Forecasting*, *22*, 427–446.

Woźniak, T. (2012). *Granger-causal analysis of VARMA-GARCH models*. Working paper series, European University Institute, Florence, Italy.

Tomasz Woźniak is a lecturer at the Department of Economics of the University of Melbourne. He received his Ph.D. in Economics from the European University Institute in 2012. His research focuses on time series econometrics and Bayesian inference. In particular, he is interested in Granger-causal analysis of conditional mean and volatility models and the Bayesian assessment of hypotheses of noncausality. Recently, he has been working on developing Bayesian methods for empirical macroeconomic research, with a focus on the inference on heteroskedastic structural vector autoregressions.

Please cite this article in press as: Woźniak, T., Testing causality between two vectors in multivariate GARCH models. International Journal of Forecasting (2015), http://dx.doi.org/10.1016/j.ijforecast.2015.01.005