European
University
Institute

DEPARTMENT
OF
ECONOMICS

Granger-Causal Analysis of Conditional Mean and Volatility Models

Tomasz Woźniak

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, December 2012

 DOI: 10.2870/63858

European University Institute
Department of Economics

Granger-Causal Analysis of Conditional Mean and Volatility Models

Tomasz Woźniak

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Examining Board

Prof. Helmut Lütkepohl, DIW Berlin and Freie Universität (External Supervisor)
Prof. Massimiliano Marcellino, European University Institute
Prof. Jacek Osiewalski, Cracow University of Economics
Prof. Giampiero Gallo, University of Florence

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

 DOI: 10.2870/63858

Contents

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

 DOI: 10.2870/63858

Abstract

Recent economic developments have shown the importance of spillover and contagion effects in financial markets as well as in macroeconomic reality. Such effects are not limited to relations between the levels of variables but also impact on the volatility and the distributions. Granger causality in conditional means and conditional variances of time series is investigated in the framework of several popular multivariate econometric models. Bayesian inference is proposed as a method of assessment of the hypotheses of Granger noncausality.

First, the family of ECCC-GARCH models is used in order to perform inference about Granger-causal relations in second conditional moments. The restrictions for second-order Granger noncausality between two vectors of variables are derived.

Further, in order to investigate Granger causality in conditional mean and conditional variances of time series VARMA-GARCH models are employed. Parametric restrictions for the hypothesis of noncausality in conditional variances between two groups of variables, when there are other variables in the system as well are derived. These novel conditions are convenient for the analysis of potentially large systems of economic variables.

Bayesian testing procedures applied to these two problems, Bayes factors and a Lindley-type test, make the testing possible regardless of the form of the restrictions on the parameters of the model. This approach also enables the assumptions about the existence of higher-order moments of the processes required by classical tests to be relaxed.

Finally, a method of testing restrictions for Granger noncausality in mean, variance and distribution in the framework of Markov-switching VAR models is pro-

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

posed. Due to the nonlinearity of the restrictions derived by Warne (2000), classical tests have limited use. Bayesian inference consists of a novel Block Metropolis-Hastings sampling algorithm for the estimation of the restricted models, and of standard methods of computing posterior odds ratios. The analysis may be applied to financial and macroeconomic time series with changes of parameter values over time and heteroskedasticity.

Keywords: Granger Causality, Second-Order Causality, Volatility Spillovers, Hypothesis Testing, Bayesian Testing, Bayes Factors, Posterior Odds Ratio, Block Metropolis-Hastings Sampling, GARCH Models, VARMA-GARCH Models, Markov-switching Models

JEL classification: C11, C12, C32, C53, E32

Acknowledgements

Agnieszka Majer

Professor Helmut Lütkepohl

Matthieu Droumaguet

Professor Massimiliano Marcellino
Professor Giampiero Gallo
Professor Jacek Osiewalski

Dr Michał Markun

Thank you

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute
DOI: 10.2870/63858

Chapter 1

Testing Causality Between Two Vectors in Multivariate GARCH Models

Abstract. Spillover and contagion effects have gained significant interest in the recent years of financial crisis. Attention has not only been directed to relations between returns of financial variables, but to spillovers in risk as well. 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 relations between second conditional moments. The restrictions for *second-order Granger noncausality* between two vectors of variables are derived, and are 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 required in classical tests to be relaxed. In the empirical example, the pound-to-Euro exchange rate is found to second-order cause the US dollar-to-Euro exchange rate, whereas the causal relation in the other direction is not supported by the data, which confirms the *meteor shower* hypothesis of Engle, Ito and Lin (1990).

This paper was presented at BMRC-QASS Conference in London in May 2010, FindEcon Conference in Łódź in May 2010 under the title *Bayesian testing of second-order causality*. The author

1.1 Introduction

The concept of *Granger causality* was introduced in econometrics by Granger (1969) and Sims (1972). One vector of variables does not Granger-cause the other 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 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 et al. (1992) and Boudjellaba et al. (1994). The forecast horizon in the definition may, however, be generalized to h or up to h periods ahead, and h may have its limit in infinity (see Lütkepohl, 1993; Dufour and Renault, 1998). Irrespective of the forecast horizon, the restrictions imposed on 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; Lütkepohl and Burda, 1997; Dufour et al., 2006).

The present paper examines the Granger causality for conditional variances. Consequently, we refer to the concept of the *second-order Granger causality*, introduced by Robins et al. (1986), and formally distinguished from the *Granger causality in variance* by Comte and Lieberman (2000). One vector of variables does not second-order Granger-cause the other vector of variables if past information about the variability of the former cannot improve the forecast of conditional variances of the latter. The definition of the *second-order noncausality* assumes that Granger causal relations might exist in the conditional mean process, however, they should be modeled and filtered out. Otherwise, such relations may impact on the parameters responsible for causal relations in conditional variances (see the empirical illustration of the problem in Karolyi, 1995).

thanks Helmut Lütkepohl, Jacek Osiewalski, Helmut Herwartz, Massimiliano Marcellino, Vance Martin and Timo Teräsvirta for useful comments and remarks on the study. Also, the author thanks Michał Markun and colleagues from the EUI Time Series Econometrics Working Group for multiple discussions and suggestions.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Restrictions already exist for the one-period-ahead second-order noncausality for the family of BEKK-GARCH models, delivered by Comte and Lieberman (2000); which take the form of several nonlinear functions of original parameters of the model. However, no good test has been established for such restrictions. 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 developed by Comte and Lieberman (2000) and Hafner and Herwartz (2008b) is to derive linear (zero) restrictions on the parameters, which would be a sufficient condition for the original restrictions, and then to apply to them a Wald test.

The conditions for the one-period-ahead second-order noncausality for the family of Extended Constant Correlation GARCH models of Jeantheau (1998) are derived in this paper. In this setting, all the considered variables are split in two vectors, between which we investigate causal relations in conditional variances. Then, the conditions for the 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 has a practical meaning in 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, a standard Bayesian procedure. 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 hypotheses of exogeneity, a concept related to Granger noncausality, Pajor (2011) used Bayes factors for models with latent variables and in particular to multivariate Stochastic Volatility models, whereas Jarociński and Maćkowak (2011) used a Savage-Dickey Ratios for the VAR model.

Since the inference is performed using the posterior odds ratios, it is based on

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

the exact finite sample results. Therefore, referring to the asymptotic results becomes pointless. This finding enables a relaxing of the assumptions required in the classical inference about the existence of the higher-order moments. For instance, in order to test the second-order noncausality hypothesis, the existence of the fourth order unconditional moments is required in Bayesian inference, whereas in classical testing the currently existing solutions require the existence of sixth-order moments (see Ling and McAleer, 2003). Notice that this assumption for testing such a hypothesis cannot be further relaxed 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 does not come from the properties of the test, but from the derivation of the restrictions for the second-order noncausality.

The structure of the paper is as follows. Section 1.2 introduces the considered model and the main theoretical finding of this work: namely, the restrictions for the second-order Granger noncausality. The assumptions behind the causal analysis are discussed. In Section 1.3, we present and discuss the existing classical approaches to testing for Granger noncausality. Since they have limited use in the considered context, we further present the posterior odds ratios as the solution. In Section 1.4, the empirical illustration of the methodology for two main exchange rates of the Eurozone is presented, and Section 1.5 concludes. The proofs are presented in A.1, while A.2 and B.2 report figures and tables respectively.

1.2 Second-order noncausality for multivariate GARCH models

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

$$y_t = (y_{1t}', y_{2t}')', \tag{1.1}$$

for all $t = 1, \dots, T$, where y_{it} is a $N_i \times 1$ vector such that $y_{1t} = (\mathrm{y}_{1t}, \dots, \mathrm{y}_{N_1.t})'$ and $y_{2t} = (\mathrm{y}_{N_1+1.t}, \dots, \mathrm{y}_{N_1+N_2.t})'$ ($N_1, N_2 \geq 1$ and $N_1 + N_2 = N$). y_1 and y_2 contain the variables of interest between which we want to study causal relations. 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. 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 $\tau \leq t$. 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.

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 relations between current and lagged observations of the considered variables:

$$y_t = \alpha_0 + \alpha(L)y_t + \epsilon_t \tag{1.2a}$$

$$\epsilon_t = D_t r_t \tag{1.2b}$$

$$r_t \sim i.i.St^N(\mathbf{0}, \mathbf{C}, \nu), \tag{1.2c}$$

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, $D_t = diag(\sqrt{h_{1t}}, \dots, \sqrt{h_{Nt}})$ 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 $\nu > 2$ 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 degrees of freedom parameter exceeds 30.

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

$$H_t = D_t \mathbf{C} D_t \quad \forall t = 1, \dots, T. \tag{1.3}$$

For the matrix H_t to be a positive definite covariance matrix, h_t must be positive for all t and $\mathbf{C}$ 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:

$$h_t = \omega + A(L)\epsilon_t^{(2)} + B(L)h_t, \tag{1.4}$$

for all $t = 1, \dots, T$, where ω is a $N \times 1$ vector of constants, $A(L) = \sum_{i=1}^{q} A_i L^i$ 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 $P(\epsilon_{t+1}^{(2)} | I^2(t)) = h_{t+1}$. Equation (1.4) has a form respecting the partitioning of the vector of data (1.1):

$$\begin{bmatrix} h_{1t} \\ h_{2t} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t}^{(2)} \\ \epsilon_{2t}^{(2)} \end{bmatrix} + \begin{bmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{bmatrix} \begin{bmatrix} h_{1t} \\ h_{2t} \end{bmatrix}. \tag{1.5}$$

Assumptions and properties Let $\theta \in \Theta \subset \mathbb{R}^k$ be a vector of size k, collecting all the parameters of the model described with equations (1.2)–(1.4). Then the likelihood function has the following form:

$$p(\mathbf{y}|\theta) = \prod_{t=1}^{T} \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} ((\nu-2)\pi)^{-\frac{N}{2}} |H_t|^{-\frac{1}{2}} \left(1 + \frac{1}{\nu-2} \epsilon_t' H_t^{-1} \epsilon_t\right)^{-\frac{\nu+N}{2}}. \tag{1.6}$$

This model has its origins in the Constant Conditional Correlation GARCH (CCC-GARCH) model proposed by Bollerslev (1990). That model consisted of N univariate GARCH equations describing the vector of conditional variances, h_t. The

CCC-GARCH model is equivalent to equation (1.4) 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. Parameters $\omega, A = \left(\text{vec}(A_1)', \ldots, \text{vec}(A_q)'\right)'$ and $B = (\text{vec}(B_1)', \ldots, \text{vec}(B_r)')'$ are such that the conditional variances, h_t, are positive for all t (see Conrad and Karanasos, 2010, for the detailed restrictions).

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

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

Assumption 4. The multivariate GARCH(r,s) 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:

$$\phi(L)\epsilon_t^{(2)} = \omega + \psi(L)v_t, \tag{1.7}$$

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 (1.1). Then (1.7) can be written in the following form:

$$\begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) \\ \phi_{21}(L) & \phi_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t}^{(2)} \\ \epsilon_{2t}^{(2)} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}. \tag{1.8}$$

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Given Assumption 3, the VARMA process (1.7) is invertible and can be written in a VAR form:

$$\Pi(L)\epsilon_t^{(2)} - \omega^* = v_t, \tag{1.9}$$

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 (1.9) in the form:

$$\begin{bmatrix} \Pi_{11}(L) & \Pi_{12}(L) \\ \Pi_{21}(L) & \Pi_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t}^{(2)} \\ \epsilon_{2t}^{(2)} \end{bmatrix} - \begin{bmatrix} \omega_1^* \\ \omega_2^* \end{bmatrix} = \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}. \tag{1.10}$$

Under Assumptions 2 and 3, processes (1.7) and (1.9) are both stationary. One more assumption is required for the inference about second-order noncausality in the GARCH model:

Assumption 5. The process v_t is covariance stationary.

The GARCH model has well-established properties under Assumptions 1–5. Under Assumption 1, conditional variances are positive. This result does not require that all the parameters of the model are positive (see Conrad and Karanasos, 2010). Further, Jeantheau (1998) proves that the GARCH(r,s) model, as in (1.4), has a unique, ergodic, weakly and strictly stationary solution when Assumption 2 holds. Under Assumptions 2–4 the GARCH(r,s) model is stationary and identifiable. Jeantheau (1998) showed that the minimum contrast estimator for the multivariate GARCH model is strongly consistent under conditions of, among others, stationarity and identifiability. 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 the parameters of the GARCH process are positive. Moreover, they have set asymptotic normality of 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 moments and derive complete fourth-moment structure.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Estimation Classical estimation consists of maximizing the likelihood function (1.6). 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 first and second partial derivatives of normal likelihood function, whereas Fiorentini et al. (2003) derive numerically reliable analytical expressions for the score, Hessian and information matrix for the models with conditional multivariate Student t distribution. Bayesian estimation requires numerical methods in order to simulate the posterior density of the parameters. Unfortunately, neither the posterior distribution of the parameters nor full conditional distributions have the form of some known distribution. The Metropolis-Hastings algorithm (see Chib and Greenberg, 1995, and references therein) was proposed by Vrontos et al. (2003) and used in Osiewalski and Pipień (2002, 2004).

The posterior distribution of the parameters of the model is proportional to the product of the likelihood function (1.6) and the prior distribution of the parameters:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta). \tag{1.11}$$

For the unrestricted VAR-GARCH model, the following prior specification is assumed. All 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 for ω the distribution is truncated to the constrained parameter space. The parameters modeling the dynamic part of the GARCH process, collected in matrices A and B follow a truncated normally-distributed prior with zero mean and diagonal covariance matrix with 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 way the sensitivity of the hypotheses assessment to Bartlett's paradox is investigated; see Section 1.3 for more details. The truncation of the distribution to the parameter space imposes Assumptions 1–5. All of the correlation parameters of the corre-

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

lation matrix **C** follow a uniform distribution on the interval $[-1,1]$. Finally, for the degrees of freedom parameter, the prior distribution proposed by Deschamps (2006) is assumed. To summarize, the prior specification for the considered model has a detailed form of:

$$p(\theta) = p(\alpha)p(\omega, A, B)p(\nu) \prod_{i=1}^{N(N-1)/2} p(\rho_i), \tag{1.12}$$

where each of the prior distributions is assumed:

$$\begin{aligned}
\alpha &\sim \mathcal{N}^{N+pN^2}\left(\mathbf{0}, 100 \cdot I_{N+pN^2}\right) \\
\omega &\sim \mathcal{N}^{N+pN^2}\left(\mathbf{0}, 100 \cdot I_{N+pN^2}\right) \mathcal{I}(\theta \in \Theta) \\
(A', B')' &\sim \mathcal{N}^{N+N^2(q+r)}\left(\mathbf{0}, \bar{s} \cdot I_{N+N^2(q+r)}\right) \mathcal{I}(\theta \in \Theta) \\
\nu &\sim .04 \exp\left[-.04(\nu - 2)\right] \mathcal{I}(\nu \geq 2) \\
\rho_i &\sim \mathcal{U}(-1, 1) \quad \text{for } i = 1, \ldots, N(N-1)/2,
\end{aligned}$$

where $\alpha = (\alpha_0', \text{vec}(\alpha_1)', \ldots, \text{vec}(\alpha_p)')'$ stacks all 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 taking 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 the elements below the diagonal of the correlation matrix, $\rho = (\text{vecl}(\mathbf{C}))$.

Such prior assumptions, with only proper distributions, have serious consequences. First, together with the bounded likelihood function, the proposed prior distribution guarantees the existence of the posterior distribution (see Geweke, 1997). Also, the proper prior distribution for the degrees of freedom parameter of the Student t sampling density of the observed series, the likelihood function, is required for the posterior distribution to be integrable, as proven by Bauwens and Lubrano (1998). However, note that prior distributions for all of the parameters, except ν, do not in fact discriminate any of the values that these parameters may take. The prior distribution of the degrees of freedom parameter gives more than a 32 percent chance that its value will be higher than 30, values close to conditional

 DOI: 10.2870/63858

normality (see e.g. Osiewalski and Pipień, 2002).

Second-Order Noncausality Conditions We focus on the question of the causal relations 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 h periods ahead y_2 if:

$$P\Big[\epsilon^{(2)}(y_{2t+h}|I(t))\Big|I^2(t)\Big] = P\Big[\epsilon^{(2)}(y_{2t+h}|I(t))\Big|I^2_{-1}(t)\Big], \tag{1.13}$$

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 and $[.]^{(2)}$ means that we square each element of a vector and $h \in \mathbb{Z}$.

A common part of both sides of (1.13) is that, in the first step, the potential Granger causal relations in the conditional mean process are filtered out. 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 $\epsilon^{(2)}(y_{2t+h}|I(t)) - P\left[\epsilon^{(2)}(y_{2t+h}|I(t))\Big|I^2_{-1}(t)\right]$ is orthogonal to $I^2(t)$ for all t (see Florens and Mouchart, 1985; Comte and Lieberman, 2000). 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 intergrable 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 $\mathsf{y}_{i\tau}$ and $\tau \leq T$.

The definition, in its original form, for one-period-ahead noncausality ($h = 1$), was proposed by Robins et al. (1986) and distinguished from *Granger noncausality in variance* by Comte and Lieberman (2000). The difference is that in the definition of *Granger noncausality in variance* there is another assumption of *Granger noncausality* *in mean*. On the contrary, in the definition of *second-order noncausality* there is no such assumption. However, any existing causal relation 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 (1.7) partitioned as in (1.8) that is identifiable and invertible (assumptions 1–5). Then y_1 does not second-order Granger-cause one period ahead y_2 if and only if:*

$$\Gamma_{ij}^{so}(z) = \det\begin{bmatrix} \phi_{11}^{\cdot j}(z) & \psi_{11}(z) \\ \varphi_{n_1+i,j}(z) & \psi_{21}^{i\cdot}(z) \end{bmatrix} = 0 \quad \forall z \in \mathbb{C} \tag{1.14}$$

for $i = 1, ..., N_2$ and $j = 1, ..., N_1$; where $\phi_{lk}^{\cdot j}(z)$ is the jth column of $\phi_{lk}(z)$, $\psi_{lk}^{i\cdot}(z)$ 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 parameters of the ECCC-GARCH model for the second-order noncausality one period ahead between two vectors of variables. Its proof, presented in A.1, is based of the theory introduced by Florens and Mouchart (1985) 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 for 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 relation 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 (1.1), and that y_1 does not second-order Granger-cause one period ahead y_2, such that the condition (1.14) holds. Then y_1 does not second-order Granger-cause h periods ahead y_2 for all $h = 1, 2, \ldots$.*

Corollary 1 is a direct application of Corollary 2.2.1 of (Lütkepohl, 2005, p. 45) to the GARCH process in the VAR form (1.10). For the proof of the restrictions for second-order Granger noncausality for the GARCH process in the VAR form, the reader is referred to A.1.

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

The theorem has equivalent for other models from the GARCH family, namely the BEKK-GARCH models. The restrictions were introduced by Comte and Lieberman (2000). There are, however, serious differences between the approaches presented by Comte and Lieberman and by this study. 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, in Example 1, we show that in order to test such a hypothesis, only two restrictions are required. The difference in the number of restrictions increases with the dimension of the time series. Secondly, due to the formulation of the BEKK-GARCH model, the noncausality conditions are much more complicated than the conditions for the ECCC-GARCH model considered here. They are simply much more complex functions of the original parameters of the model. Both these arguments have consequences in testing that require estimation of the restricted model or employment of the delta method. A high number of restrictions may have a strongly negative impact on the size and power properties of tests. However, the ECCC-GARCH model assumes that the correlations are time invariant, which is not the case for the BEKK-GARCH model.

Nakatani and Teräsvirta (2009) propose the Lagrange Multiplier test for the hypothesis of no volatility spillovers in a bivariate ECCC-GARCH model. The restrictions they test are zero restrictions on the off-diagonal elements of matrix polynomials $A(L)$ and $B(L)$ from the GARCH equation (1.5). Consequently, the

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

null hypothesis is represented by the CCC-GARCH model of Bollerslev (1990) and the alternative hypothesis by ECCC-GARCH of Jeantheau (1998). Note, that if all 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 considered in applied studies), then the null hypothesis of Nakatani and Teräsvirta (2009) is equivalent to the second-order Granger noncausality condition, as in the Example 1. In a general case, for any dimension of the time series, the zero restrictions on the off-diagonal elements of matrix polynomials $A(L)$ and $B(L)$ represent a sufficient condition for the second-order noncausality.

To conclude, the condition (1.14) leads to the finite number of nonlinear restrictions on the original parameters of the model. Several examples will clarify how they are set.

Example 1. Suppose that y_t follows a bivariate GARCH(1,1) process, ($N = 2$, and $p = q = 1$). The VARMA process for $\epsilon_t^{(2)}$ is as follows:

$$\begin{bmatrix} 1-(A_{11}+B_{11})L & -(A_{12}+B_{12})L \\ -(A_{21}+B_{21})L & 1-(A_{22}+B_{22})L \end{bmatrix} \begin{bmatrix} \epsilon_{1t}^2 \\ \epsilon_{2t}^2 \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} 1-B_{11}L & -B_{12}L \\ -B_{21}L & 1-B_{22}L \end{bmatrix} \begin{bmatrix} \nu_{1t} \\ \nu_{2t} \end{bmatrix}. \quad (1.15)$$

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

$$\det \begin{bmatrix} 1-(A_{11}+B_{11})z & 1-B_{11}z \\ -(A_{21}+B_{21})z & -B_{21}z \end{bmatrix} \equiv \mathbf{0}, \quad (1.16)$$

which leads to the following set of restrictions:

$$\mathbf{R}_1^I(\theta) = A_{21} = 0, \quad \text{and} \quad \mathbf{R}_2^I(\theta) = B_{21}A_{11} = 0. \quad (1.17)$$

Example 2. Let y_t follow a trivariate GARCH(1,1) process ($N = 3$ and $r = s = 1$).

The VARMA process for $\epsilon_t^{(2)}$ is as follows:

$$\begin{bmatrix} 1-(A_{11}+B_{11})L & -(A_{12}+B_{12})L & -(A_{13}+B_{13})L \\ -(A_{21}+B_{21})L & 1-(A_{22}+B_{22})L & -(A_{23}+B_{23})L \\ -(A_{31}+B_{31})L & -(A_{32}+B_{32})L & 1-(A_{33}+B_{33})L \end{bmatrix} \begin{bmatrix} \epsilon_{1t}^2 \\ \epsilon_{2t}^2 \\ \epsilon_{3t}^2 \end{bmatrix} =$$
$$= \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} + \begin{bmatrix} 1-B_{11}L & -B_{12}L & -B_{13}L \\ -B_{21}L & 1-B_{22}L & -B_{23}L \\ -B_{31}L & -B_{32}L & 1-B_{33}L \end{bmatrix} \begin{bmatrix} \nu_{1t} \\ \nu_{2t} \\ \nu_{3t} \end{bmatrix}. \quad (1.18)$$

From Theorem 1, we see, that $y_1 = \mathsf{y}_1$ does not second-order Granger-cause $y_2 = (\mathsf{y}_2, \mathsf{y}_3)$ if and only if:

$$\det \begin{bmatrix} 1-(A_{11}+B_{11})z & 1-B_{11}z \\ -(A_{i1}+B_{i1})z & -B_{i1}z \end{bmatrix} = 0 \quad \text{for } i = 2,3, \quad (1.19)$$

which results in the following restrictions:

$$\mathbf{R}_1^{II}(\psi) = A_{11}B_{21} = 0 \quad \text{and} \quad \mathbf{R}_2^{II}(\psi) = A_{21} = 0 \quad (1.20a)$$

$$\mathbf{R}_3^{II}(\psi) = A_{11}B_{31} = 0 \quad \text{and} \quad \mathbf{R}_4^{II}(\psi) = A_{31} = 0. \quad (1.20b)$$

However, $y_1 = (\mathsf{y}_1, \mathsf{y}_2)$ does not second-order Granger-cause $y_2 = \mathsf{y}_3$ if and only if:

$$\det \begin{bmatrix} 1-(A_{11}+B_{11})z & 1-B_{11}z & -B_{13}z \\ -(A_{21}+B_{21})z & -B_{21}z & -B_{23}z \\ -(A_{31}+B_{31})z & -B_{31}z & 1-B_{33}z \end{bmatrix} \equiv \mathbf{0}, \quad (1.21)$$

which leads to the following set of restrictions:

$$\mathbf{R}_1^{III}(\psi) = A_{11}(B_{23}B_{31} - B_{21}B_{33}) + A_{31}(B_{13}B_{21} - B_{11}B_{23}) = 0 \quad (1.22a)$$

$$\mathbf{R}_2^{III}(\psi) = A_{11}B_{21} + A_{31}B_{23} = 0 \quad (1.22b)$$

$$\mathbf{R}_3^{III}(\psi) = A_{21} = 0. \quad (1.22c)$$

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

1.3 Bayesian hypotheses assessment

The restrictions derived in Section 1.2 can be tested. We propose to use the Bayesian approach to assess the hypotheses of second-order noncausality represented by the restrictions. Before the approach is presented, however, the classical tests proposed so far and their limitations are discussed.

Classical testing 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 been established at that time. The asymptotic result was presented in Comte and Lieberman (2003). This finding, however, does not solve the problem of testing the nonlinear restrictions imposed on the parameters of the model. 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. A similar procedure was presented in Hafner and Herwartz (2008b) for the Wald test, and in 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, contrary to the conditions of Comte and Lieberman, Hafner and Herwartz and Nakatani and Teräsvirta, may be nonlinear (see Example 2). 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. Thus, the asymptotic distribution of the Wald test statistic is no longer normal. In fact, for the time being it is unknown. Consequently, the Wald test statistic cannot be

 DOI: 10.2870/63858

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 original parameters of the model. As a solution to the problem of testing the restrictions, they propose a sequential testing procedure. There are two main drawbacks in this method. First, despite being properly performed, the test may still appear inconclusive, 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. The proposed solutions, unfortunately, are only applicable 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 the h-step ahead Granger noncausality for VAR models. This method could be applied to the problem of testing the nonlinear restrictions for the second-order noncausality in GARCH models. More studies are required, however, on the applicability and properties of this test.

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 Hafner and Herwartz (2008b) and Comte and Lieberman (2000) require asymptotic normality of the Quasi Maximum Likelihood Estimator. This result is derived under the existence of bounded moments of order 8 by Comte and Lieberman (2003). For the ECCC-GARCH model considered in this study, the asymptotic normality of the Quasi Maximum Likelihood Estimator is derived in Ling and McAleer (2003) under the existence of moments of order 6. This assumption is, however, relaxed for the purpose of testing the existence of volatility spillovers by Nakatani and Teräsvirta (2009). Their Lagrange Multiplier test statistic requires the existence of fourth-order moments. The Bayesian test presented below further relaxes this assumptions.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

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

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

$$p(\mathbf{y}|\mathcal{M}_i) = \int_{\theta\in\Theta} p(\mathbf{y}|\theta, \mathcal{M}_i)p(\theta|\mathcal{M}_i)d\theta \tag{1.23}$$

be marginal distributions of data corresponding to each of the model, for $i = 1, \dots, m$. $p(\mathbf{y}|\theta, \mathcal{M}_i)$ and $p(\theta|\mathcal{M}_i)$ are the likelihood function (1.6) and the prior distribution (1.12) respectively. The extended notation respecting conditioning on one of the models is used here. The marginal density of data is a constant normalizing kernel of the posterior distribution (1.11).

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

$$\mathcal{B}_{ij} = \frac{p(\mathbf{y}|\mathcal{M}_i)}{p(\mathbf{y}|\mathcal{M}_j)}, \tag{1.24}$$

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

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

$$\Pr(\mathcal{M}_i|\mathbf{y}) = \frac{p(\mathbf{y}|\mathcal{M}_i)\Pr(\mathcal{M}_i)}{\sum_{j=1}^{m} p\left(\mathbf{y}|\mathcal{M}_j\right)\Pr\left(\mathcal{M}_j\right)}, \tag{1.25}$$

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

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

$$\Pr(\mathcal{H}_i|\mathbf{y}) = \sum_{j\in\mathcal{H}_i} \Pr\left(\mathcal{M}_j|\mathbf{y}\right). \tag{1.26}$$

This approach to assessment of hypotheses requires the estimation of all the models representing considered hypotheses, as well as the estimation of the corresponding to the models marginal densities of data (1.23).

Bartlett's paradox Using Bayes factors for the comparison of the models is not uncontroversial. It appears that Bayes factors are sensitive to the specification of the prior distributions for the parameters being tested. The more diffuse a prior distribution the more informative it is about the the parameter tested with a Bayes factor. This phenomenon is called Bartlett's paradox (see Bartlett, 1957) and is a version of the Lindley paradox. Moreover, Strachan and van Dijk (2011) show that assuming a diffuse prior distribution for the parameters of the model, results in wrongly defined Bayes factors. As a solution to this problem Strachan and van Dijk recommend using a prior distribution belonging to a class of shrinkage distributions.

The sensitivity of the model assessment with respect to the specification of the prior distribution is checked by assuming for each of the estimated model two different prior distributions for the matrices of parameters A and B. The distributions, defined in Section 1.2, 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.

Estimation of models The form of the posterior distribution (1.11) for all of the parameters, θ, for the GARCH models, even with the prior distribution set to a proper distribution function, as in (1.12), is not in a form of any known distribution function. 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

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

distribution. Therefore, the posterior distribution may be simulated with a Monte Carlo Markov Chain (MCMC) algorithm. Due to the above mentioned problems with the form of the posterior and full conditional densities, a proper algorithm to sample the posterior distribution (1.11) is, e.g. the Metropolis-Hastings algorithm (see Chib and Greenberg, 1995, and references therein). The algorithm was adapted for multivariate GARCH models by Vrontos et al. (2003).

Suppose 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 the degrees of freedom parameter set to 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 relatively efficiently from the posterior density. A new candidate θ' is accepted with the probability:

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

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 to ensure that it is a good sample from the stationary posterior distribution (see e.g. Geweke, 1999; Plummer et al., 2006).

Estimation of the marginal distribution of data Having estimated the models, the marginal densities of the data may be computed using one of the available methods. Since the estimation of the models is performed using the Metropolis-Hastings algorithm, a suitable estimator of the marginal density of data is presented by Geweke (1997). However, any estimator of the marginal density of data applicable to the problem might be used (see Miazhynskaia and Dorffner, 2006,

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

who review the estimators of marginal density of data for univariate GARCH models). The Bayesian comparison of bivariate GARCH models using Bayes factors was presented by Osiewalski and Pipień (2004).

The Modified Harmonic Mean estimator of the marginal density of data proposed by Geweke (1997) is:

$$\hat{p}(\mathbf{y}|\mathcal{M}_i) = \left[S^{-1} \sum_{s=1}^{S} \frac{f(\theta^{(s)})}{p(\mathbf{y}|\theta^{(s)}, \mathcal{M}_i)p(\theta^{(s)}|\mathcal{M}_i)} \right]^{-1}, \tag{1.27}$$

where $f(\theta^{(s)})$ is a multivariate truncated normal distribution, with the mean vector equal to the posterior mean and covariance matrix set to the posterior covariance matrix. The truncation is chosen such that $f(\theta^{(s)})$ have thinner tails than the posterior distribution.

In comparison to the Harmonic Mean estimator of Newton and Raftery (1994), given the assumptions regarding the prior distribution, the Modified Harmonic Mean Estimator has that advantage that it is bounded as explained by Frühwirth-Schnatter (2004).

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 testing of the parameter conditions possible. 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 coming from the Posterior Odds Ratio, the outcome of the test is a positive argument *in favour of* the most likely *a posteriori* hypothesis. Moreover, contrary to classical testing, a choice is being made between all the competing hypotheses at once, not only between the unrestricted and one of the restricted models (see Hoogerheide et al., 2009, for a discussion of the argument).

Further, 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, Assumptions required to test the restrictions may now be relaxed. In order to test the second-order noncausality hypothesis, Assumptions 1–5 must hold. This requires the existence of the fourth-order unconditional moment that is ensured by the restrictions derived by He and Teräsvirta (2004). No classical test of the restrictions has been proposed so far for the ECCC-GARCH model. Note that the Bayesian testing in the proposed form may be applied to BEKK-GARCH and vec-GARCH models without any complications, and while preserving all the advantages. Then, the assumption of existence of moments of order four is a significant improvement, in comparison with the result of Comte and Lieberman (2003). There, the asymptotic distribution of the QMLE is established under the existence of the eighth-order moments.

For the testing of volatility spillovers in the ECCC-GARCH model, the assumption may be further relaxed. Here, the strict assumption for the linear theory for noncausality of Florens and Mouchart (1985) need not hold. In fact, for testing the zero restrictions for the no volatility spillovers hypothesis, the only required assumption 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 required the existence of fourth-order moments for the Lagrange Multiplier test statistic to be asymptotically χ^2-distributed.

The improvements in moment conditions are, therefore, established for both kinds of hypothesis. This fact may be crucial for the testing of the hypotheses on the financial time series. In multiple applied studies, such data are 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 costs of the proposed approach is the necessity to estimate all the unrestricted and restricted models. This simply requires some time-consuming

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 1.1: Summary statistics of the exchange rate logarithmic rates of return expressed in percentage points

	GBP/EUR	USD/EUR
Mean	0.012	-0.006
Median	0.011	0.016
Standard Deviation	0.707	0.819
Correlation	.	0.368
Minimum	-2.657	-4.735
Maximum	3.461	4.038
Excess kurtosis	2.430	2.683
Excess kurtosis (robust)	0.060	0.085
Skewness	0.344	-0.091
Skewness (robust)	0.010	-0.016
LJB test	206.525	234.063
LJB p-value	0.000	0.000
T	777.0	777.0

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

computations. While bivariate GARCH models may (depending on the order of the process, and thus on the number of the parameters) be estimated reasonably quickly, trivariate models require significant amounts of time and computational power.

1.4 Granger causal analysis of exchange rates

The restrictions derived in Section 1.2 for second-order noncausality for GARCH models, along with the Bayesian testing procedure described in Section 1.3, are now used in an analysis of the bivariate system of two exchange rates.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Data The system under consideration consists of daily exchange rates of the British pound (GBP/EUR) and the US dollar (USD/EUR), both denominated in Euro. Logarithmic rates of return expressed in percentage points are analyzed, $y_{it} = 100(\ln x_{it} - \ln x_{it-1})$ for $i = 1, 2$, where x_{it} are levels of the assets. The data spans the period from 16 September 2008 to 22 September 2011, which gives $T = 777$ prices, and was downloaded from the European Central Bank website (http://sdw.ecb.int/browse.do?node=2018794). The analyzed period starts the day after Lehman Brothers filed for Chapter 11 bankruptcy protection.

The data set contains the two most liquid exchange rates in the Eurozone. The chosen period of analysis starts just after an event that had a very strong impact on the turmoil in the 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 financial institutions as well as public institutions located in the Eurozone whose performance depends on the forecast of exchange rates. Such institutions include the governments of the countries belonging to the Eurozone that keep their debts in currencies, mutual funds and banks, and all the participants of the exchange rates market.

Figure A.1 from A.2 plots the time series. It clearly shows the first period of length – of nearly a year – which may be characterized by the high level of volatility of the exchange rates. The subsequent period is characterized by a slightly lower volatility for both series. The evident heteroskedasticity, as well as the volatility clustering, seem to provide a strong argument in favor of specifyng the GARCH models that are capable of modeling such features in the data.

Table 2.1 reports summary statistics of the two considered series. Both of the returns series have sample means and medians close to zero. The US dollar has a slightly larger sample standard deviation than the British pound. Both series are leptokurtic, which is evidenced by the excess kurtosis coefficient of around 2.5. The pound is slightly positively and the dollar slightly negatively skewed. Neither series can be well described with a normal distribution.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Testing strategy and estimation results For the bivariate time series of exchange rates the Vector autoregression of order 1 with the Extended CCC GARCH(1,1) conditional variance process is fitted with two different assumptions regarding the prior distribution. The first prior distribution, referred to as a diffuse prior, is as specified in Section 1.2 and equation (1.12) with the value of the hyper-parameter $\bar{s}$ set to 100. The second assumed prior distribution, referred to as shrinkage distribution, has the value oh this hyper-parameter set to 0.1. The models that are estimated are as follows. The unrestricted model defined by equations (1.2), (1.3) and (1.4) allows for second-order causality in both directions: from GBP/EUR to USD/EUR and conversely. Restricted models represent different hypotheses of noncausality, and are restricted according to the conditions stated in Theorem 1. All the models, the unrestricted and 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 two restrictions:

$$A_{21} = 0 \qquad \text{and} \qquad B_{21}A_{11} = 0.$$

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

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

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

The strategy for the assessment of the hypotheses is the following. For each of the hypotheses, a full set of sufficient conditions of the restrictions representing the hypothesis is derived. The sufficient conditions are in the form of zero restrictions

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 1.2: Summary of the estimation of the unrestricted VAR(1)-ECCC-GARCH(1,1) models

Panel A: Estimation results for model $\mathcal{M}_0$ with diffuse prior distribution									
	VAR(1)			*GARCH(1,1)*					ν
	α_0	α_1		ω	A		B		ρ_{12}
GBP/EUR	0.010	0.083	-0.021	0.011	0.049	0.020	0.525	0.263	11.904
	(0.021)	(0.040)	(0.033)	(0.011)	(0.025)	(0.016)	(0.281)	(0.184)	(3.697)
USD/EUR	0.009	0.067	-0.006	0.033	0.078	0.044	0.730	0.342	0.409
	(0.026)	(0.048)	(0.041)	(0.033)	(0.046)	(0.030)	(0.324)	(0.248)	(0.033)
Panel B: Estimation results for model $\mathcal{M}_0$ with shrinkage prior distribution									
	VAR(1)			*GARCH(1,1)*					ν
	α_0	α_1		ω	A		B		ρ_{12}
GBP/EUR	0.011	0.083	-0.022	0.012	0.052	0.020	0.508	0.271	11.535
	(0.021)	(0.041)	(0.033)	(0.013)	(0.028)	(0.014)	(0.282)	(0.186)	(3.326)
USD/EUR	0.009	0.066	-0.007	0.036	0.078	0.045	0.715	0.349	0.410
	(0.026)	(0.049)	(0.041)	(0.038)	(0.046)	(0.030)	(0.330)	(0.252)	(0.033)

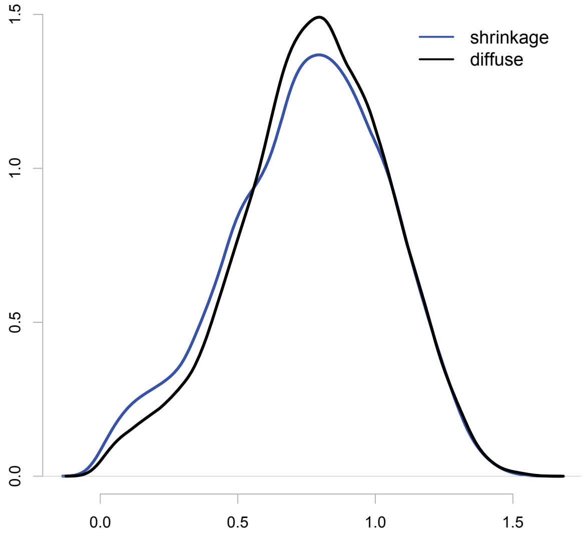
Note: The table summarizes the estimation of the VAR(1)-ECCC-GARCH(1,1) model described by the equations (1.2), (1.3), (1.4) and the likelihood function (1.6). The prior distributions are specified as in equation (1.12). The posterior means and the posterior standard deviations (in brackets) of the parameters are reported. A summary of the characteristics of the simulations of the posterior densities of the parameters for all the models are reported in B.2, in Tables A.1 and A.2.

imposed on single parameters. All of the restricted models are estimated and the respective marginal distributions of data are computed. Posterior probabilities of the models and of the hypotheses are computed. The hypotheses are compared using Posterior Odds Ratios. Table 1.3 presents all the hypotheses and the models representing them with restrictions being a sufficient condition of the conditions resulting from Theorem 1 for each of the hypotheses.

We start the analysis of the results with several comments on the parameters of the unrestricted models. Table 1.2 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 in the

 DOI: 10.2870/63858

Figure 1.1: Marginal posterior densities of parameter B_{21}



Note: Marginal posterior densities for both of the prior specifications represent the marginal posterior densities of B_{21}, 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 in model $\mathcal{M}_i$ parameter B_{21} is not restricted to zero.

parameters' values between these two models. All the following comments, thus, concern both of the specifications. Among the parameters of the vector autoregressive part, only parameter $\alpha_{1.11}$ can be considered 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, for most of parameters of matrices ω, A and B, a significant part of the posterior probability mass is concentrated around zero. Only one parameter, namely B_{21}, has its posterior probability mass not concentrated around zero for both of the specifications, with diffuse and shrinkage prior distribution. Figure 1.1 plots the marginal posterior densities for both assumed prior densities. This parameter is responsible for the volatility transmission from the lagged value of the conditional variance of the GBP/EUR exchange rate on the current conditional variance of variable USD/EUR. Regardless of the prior density specification, this is the only parameter significantly different from zero. These findings, especially regarding the parameters of the matrices A and B, are reflected in the results of the assessment of the hypotheses of second-order noncausality.

Hypotheses assessment The credibility of the hypotheses of second-order noncausality between the exchange rates of the British pound and the US dollar to Euro is evaluated. All together four different hypotheses are formed and assessed within the framework of the VAR-GARCH model. Due to the adopted strategy – testing the full set of sufficient conditions for the original restrictions – some of the hypotheses are represented by several models. Table 1.3 summarizes the hypotheses, the models representing them, as well as the restrictions on the parameters according to which the models are restricted. Moreover, the table reports natural logarithms of the marginal densities of data for each of the models and for both assumed prior distributions.

According to Table 1.3 the model best supported by the data for both of the prior distributions is model $\mathcal{M}_6$. Both of the models: with diffuse and shrinkage prior distributions, have the highest value of the logarithm of the marginal data density. Moreover, five out of six of the models that represent the second hypothesis, of second-order noncausality from dollar to pound, have the values of logarithms of the marginal data densities higher than any other estimated model. In effect, this hypothesis is expected to have the highest posterior probability of all the considered hypotheses. Table 1.4, providing Posterior Odds Ratios of hypotheses one, two and three to the null hypothesis, confirms this claim. Hypothesis $\mathcal{H}_2$ has the highest value of the posterior odds ratio relative to the null hypothesis, and thus, it attracts the biggest part of the posterior probability mass of the hypotheses.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 1.3: Marginal densities of data for models

$\mathcal{M}_j$	Restrictions	$\ln p(\mathbf{y}\vert\mathcal{M}_j)$ diffuse prior	shrinkage prior
	$\mathcal{H}_0$: Unrestricted model		
$\mathcal{M}_0$	-	-1649.034	-1649.100
	$\mathcal{H}_1$: GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR		
$\mathcal{M}_1$	$A_{21} = B_{21} = 0$	-1652.868	-1658.230
$\mathcal{M}_2$	$A_{11} = A_{21} = 0$	-1653.773	-1654.851
$\mathcal{M}_3$	$A_{11} = A_{21} = B_{21} = 0$	-1653.170	-1655.024
	$\mathcal{H}_2$: USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR		
$\mathcal{M}_4$	$A_{12} = B_{12} = 0$	-1646.750	-1652.790
$\mathcal{M}_5$	$A_{12} = A_{22} = 0$	-1646.377	-1646.501
$\mathcal{M}_6$	$A_{12} = A_{22} = B_{12} = 0$	-1645.714	-1645.737
	$\mathcal{H}_3$: GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR and USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR		
$\mathcal{M}_7$	$A_{12} = A_{21} = B_{12} = B_{21} = 0$	-1650.702	-1658.387
$\mathcal{M}_8$	$A_{12} = A_{21} = A_{22} = B_{21} = 0$	-1676.102	-1676.266
$\mathcal{M}_9$	$A_{12} = A_{21} = A_{22} = B_{12} = B_{21} = 0$	-1672.589	-1672.623
$\mathcal{M}_{10}$	$A_{12} = A_{21} = A_{11} = B_{12} = 0$	-1680.278	-1680.443
$\mathcal{M}_{11}$	$A_{12} = A_{21} = A_{11} = A_{22} = 0$	-1685.885	-1685.335
$\mathcal{M}_{12}$	$A_{12} = A_{21} = A_{11} = A_{22} = B_{12} = 0$	-1681.618	-1681.671
$\mathcal{M}_{13}$	$A_{12} = A_{21} = A_{11} = B_{12} = B_{21} = 0$	-1681.372	-1681.238
$\mathcal{M}_{14}$	$A_{12} = A_{21} = A_{11} = A_{22} = B_{21} = 0$	-1687.385	-1687.276
$\mathcal{M}_{15}$	$A_{12} = A_{21} = A_{11} = A_{22} = B_{12} = B_{21} = 0$	-1693.491	-1695.578

Note: An estimator of the marginal densities of data is the Modified Harmonic Mean estimator by Geweke (1997), defined by equation (3.26). A summary of the characteristics of the simulations of the posterior densities of the parameters of the models are reported in B.2, in Tables A.1 and A.2.

Other hypotheses gained a negligible part of the posterior probability mass. Also the rank of the credibility of the hypotheses is not entirely robust to the specification of the prior distributions for the parameters of the model. Note, however, that the unrestricted model is rejected by the data. Also the hypotheses that include the restrictions of second-order noncausality from pound to dollar are rejected. Most of the models representing these two hypotheses restrict the parameter B_{21} to zero. Therefore, if the posterior probability mass for this parameter is far from zero, the restriction cannot hold. In effect, such hypotheses are not supported by the data.

The most important finding of the empirical analysis is that the exchange rate of the USD/EUR does not second-order cause the exchange rate of the GBP/EUR. This means that past information of the variability of dollar's exchange rate is dispensable for the forecast of the conditional variance of pound's exchange rate constructed within the bivariate VAR-GARCH model. This finding is robust to the specification of the prior distribution for the parameters of the model. How to explain this somewhat surprising result?

This phenomenon is in line with the *meteor shower* hypothesis of Engle et al. (1990) that 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, first agents in Australia and Asia are active, then agents in Europe (and Africa) start being active; and finally, traders in both Americas start working. Therefore, coming back to our example, on a particular day, first agents in Europe trade between Euro zone and United Kingdom and only after that, when working hours in the United States commence, agents start trading between Euro zone and the USA. Such a pattern is captured by the models representing the hypothesis of second-order noncausality from pound to dollar. Note that the triangular GARCH model of Engle et al. (1990), representing the *meteor shower* hypothesis, is just one of the models, namely $\mathcal{M}_4$, representing hypothesis $\mathcal{H}_2$.

1.5 Conclusions

In this work the conditions for analyzing Granger noncausality for the second conditional moments of a GARCH process are derived. The presented restrictions for one period ahead second-order noncausality, due to the specific setting of the

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 1.4: Summary of the hypotheses testing

$\mathcal{H}_i$	Hypothesis	Models	$\frac{Pr(\mathcal{H}_i\|\mathbf{y})}{Pr(\mathcal{H}_0\|\mathbf{y})}$ diffuse prior	shrinkage prior
$\mathcal{H}_0$	Unrestricted model	$\mathcal{M}_0$	1	1
$\mathcal{H}_1$	GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR	$\mathcal{M}_1$–$\mathcal{M}_3$	0.0463	0.006
$\mathcal{H}_2$	USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR	$\mathcal{M}_4$–$\mathcal{M}_6$	51.7058	42.338
$\mathcal{H}_3$	GBP/EUR $\overset{so}{\nrightarrow}$ USD/EUR and USD/EUR $\overset{so}{\nrightarrow}$ GBP/EUR	$\mathcal{M}_7$–$\mathcal{M}_{15}$	0.1885	0.0001

Note: Posterior Probabilities of the hypotheses were computed using a formula of equation (1.26) that uses posterior probabilities of models, (1.25), and assuming flat prior distributions of the models: $\Pr(\mathcal{M}_i) = m^{-1}$ for $i = 1, \ldots, m$.

system, in which all the considered variables belong to one of the two vectors, appear to be the restrictions for the second-order noncausality at all future horizons. These conditions may result in several nonlinear restrictions on the parameters of the model, which results in the fact that the available classical tests have limited uses.

Therefore, in order to test these restrictions, the basic Bayesian procedure is applied, which consists of the estimation of the models representing the hypotheses of second-order causality and noncausality, and then of the comparison of the models and hypotheses with posterior odds ratios. This well-known procedure overcomes the difficulties that the classical tests applied so far to this problem have met. 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.

Several remarks regarding the proposed approach are, however, in place. These

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

come from 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 the hypotheses of interest may be formulated in the system that contains more than two variables (see Example 2). Another limitation is the fact that the presented restrictions serve as the restrictions for the second-order noncausality at all future horizons at once. This feature is caused by the particular setting considered in this work.

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 the purpose of modeling and forecasting. This may be particularly necessary for the analysis of the robustness of the causal or noncausal relations 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 relations between economic relations. However, the setting considered in this work does not allow for such an analysis.

Appendix A

A.1 Proof

Proof of Theorem 1 The first part of the proof sets the second-order noncausality restrictions for the GARCH process in the VAR form (1.9). Let $\epsilon_t^{(2)}$ follow a stationary VAR process as in (1.9), partitioned as in (1.10), that is identifiable. Then, y_1 does not second-order Granger-cause y_2 if and only if:

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

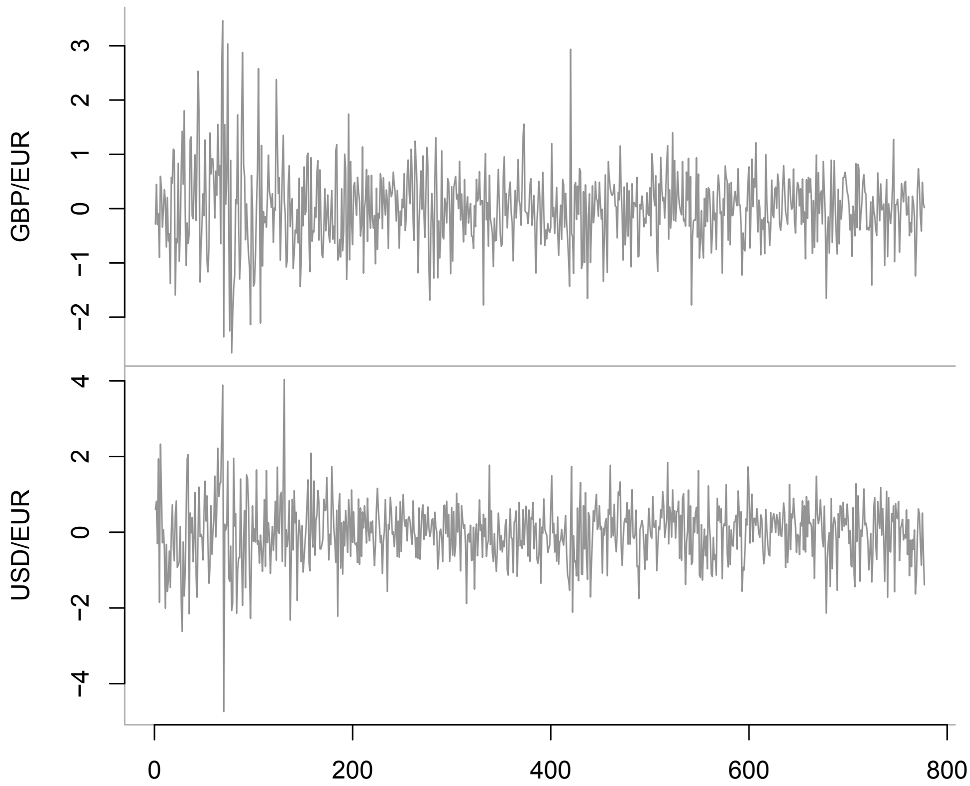
Condition (A.1) may be proven by the application of Proposition 1 of Boudjellaba et al. (1992). Several changes are, however, required to adjust the proof of that Proposition for the vector autoregressive process to the setting considered in Theorem 1 for the GARCH models. Here, one projects the squared 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 1.2.

The proven condition still leads to infinite number of restrictions on parameters. This property excludes the possibility of testing these restrictions. In order to obtain the simplified condition (1.14), apply to (A.1) the matrix transformations, first of Theorem 1 and then of Theorem 2 of Boudjellaba et al. (1994).

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

A.2 Data

Figure A.1: Data plot: (GBP/EUR,USD/EUR)



The graph presents the daily logarithmic rates of return, expressed in percentage points $y_{it} = 100(\ln x_{it} - \ln x_{it-1})$ for $i = 1, 2$, where x_{it} denotes the level of an asset of two exchange rates: the British pound and the US dollar, all denominated in Euro. The data spans the period from 16 September 2008 to 22 September 2011, which gives $T = 777$ observations. It was downloaded from the European Central Bank website (http://sdw.ecb.int/browse.do?node=2018794).

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute
DOI: 10.2870/63858

A.3 Summary of the simulation

 DOI: 10.2870/63858

Table A.1: Properties of the simulations of the posterior densities of the models with diffuse prior distributions

Model	RNE median	RNE min	RNE max	Autocorrelation at lag 1 median	Autocorrelation at lag 1 min	Autocorrelation at lag 1 max	Autocorrelation at lag 10 median	Autocorrelation at lag 10 min	Autocorrelation at lag 10 max	Geweke's z median	Geweke's z min	Geweke's z max	S
$\mathcal{M}_0$	0.203	0.034	0.745	0.531	0.361	0.830	0.047	-0.032	0.386	0.074	-2.530	2.716	4500
$\mathcal{M}_1$	0.683	0.225	1.008	0.208	0.089	0.571	0.004	-0.017	0.026	0.454	-1.966	2.167	3000
$\mathcal{M}_2$	0.035	0.017	0.147	0.834	0.717	0.937	0.370	0.045	0.722	-0.096	-2.542	2.439	4000
$\mathcal{M}_3$	0.204	0.027	0.339	0.623	0.541	0.841	0.024	-0.011	0.450	0.140	-2.307	2.164	5000
$\mathcal{M}_4$	0.639	0.228	0.922	0.188	0.124	0.472	-0.004	-0.050	0.045	-0.175	-1.248	1.790	2800
$\mathcal{M}_5$	0.138	0.025	0.322	0.703	0.551	0.879	0.112	0.017	0.470	0.171	-1.302	1.249	4500
$\mathcal{M}_6$	0.638	0.414	0.834	0.195	0.131	0.335	0.005	-0.012	0.025	-0.001	-1.609	2.187	4500
$\mathcal{M}_7$	0.081	0.055	0.107	0.767	0.731	0.859	0.152	0.078	0.265	-0.307	-1.995	0.974	3000
$\mathcal{M}_8$	0.501	0.267	0.992	0.113	-0.007	0.546	-0.001	-0.020	0.036	0.105	-2.362	2.080	4500
$\mathcal{M}_9$	0.222	0.052	0.457	0.449	0.249	0.779	0.031	0.008	0.262	-0.086	-0.934	1.941	5000
$\mathcal{M}_{10}$	0.147	0.024	0.352	0.534	0.326	0.921	0.050	-0.016	0.546	-0.286	-2.946	1.718	4500
$\mathcal{M}_{11}$	0.302	0.027	0.478	0.484	0.413	0.877	0.020	-0.015	0.537	-0.418	-2.194	1.429	4500
$\mathcal{M}_{12}$	0.464	0.028	0.992	0.206	0.011	0.755	-0.009	-0.030	0.411	-0.732	-1.993	2.309	4500
$\mathcal{M}_{13}$	0.586	0.114	0.991	0.128	0.044	0.662	0.006	-0.028	0.167	0.199	-0.698	3.024	4500
$\mathcal{M}_{14}$	0.498	0.049	1.233	0.121	0.060	0.783	0.030	-0.012	0.327	-0.231	-1.241	0.879	4500
$\mathcal{M}_{15}$	0.387	0.145	0.589	0.204	0.156	0.471	0.037	0.012	0.126	-0.134	-1.687	1.377	5400

Note: The table summarizes the properties of the numerical simulation of the posterior densities of all the considered models. For each of the statistics, the median of all the parameters of the model, as well as minimum and maximum, are reported. The table reports the relative numerical efficiency coefficient (RNE), by Geweke (1989), autocorrelations of the MCMC draws at lags 1 and 10, as well as 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 numbers presented in this table were obtained using the package coda by Plummer et al. (2006).

DOI: 10.2870/63858

Table A.2: Properties of the simulations of the posterior densities of the models with shrinkage prior distribution

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	
$\mathcal{M}_0$	0.238	0.036	0.81	0.414	0.178	0.814	0.037	-0.007	0.38	0.056	-1.954	1.86	4500
$\mathcal{M}_1$	0.559	0.138	1.055	0.215	0.049	0.514	0.011	-0.019	0.102	-0.028	-1.323	2.231	5500
$\mathcal{M}_2$	0.084	0.013	0.270	0.757	0.577	0.908	0.132	-0.014	0.679	-0.021	-1.833	1.701	5000
$\mathcal{M}_3$	0.692	0.092	1.120	0.179	0.076	0.534	-0.005	-0.038	0.131	-0.600	-1.956	1.649	4100
$\mathcal{M}_4$	0.672	0.294	1.198	0.109	0.041	0.470	0.002	-0.034	0.023	-0.129	-1.651	1.525	5500
$\mathcal{M}_5$	0.234	0.013	0.618	0.476	0.275	0.941	0.051	-0.038	0.719	0.173	-2.022	1.691	5000
$\mathcal{M}_6$	0.686	0.235	0.987	0.196	0.089	0.528	-0.000	-0.040	0.023	-0.508	-1.898	1.668	4300
$\mathcal{M}_7$	0.071	0.016	0.137	0.765	0.700	0.940	0.167	0.043	0.618	0.512	-1.689	1.556	5500
$\mathcal{M}_8$	0.225	0.012	0.506	0.495	0.312	0.984	0.034	-0.033	0.907	-0.057	-1.588	1.329	4500
$\mathcal{M}_9$	0.427	0.175	0.735	0.196	0.124	0.578	0.005	-0.006	0.080	0.201	-1.889	2.308	5000
$\mathcal{M}_{10}$	0.456	0.013	0.886	0.162	0.067	0.966	0.026	-0.014	0.903	-0.262	-2.399	2.802	4300
$\mathcal{M}_{11}$	0.343	0.012	0.520	0.466	0.404	0.970	0.016	-0.037	0.807	-0.079	-1.954	1.706	4500
$\mathcal{M}_{12}$	0.825	0.637	1.269	0.075	0.007	0.229	-0.003	-0.031	0.031	-0.352	-2.221	2.272	4500
$\mathcal{M}_{13}$	0.430	0.091	0.954	0.244	0.112	0.775	0.017	-0.041	0.151	0.193	-1.922	2.158	5000
$\mathcal{M}_{14}$	0.748	0.015	1.115	0.069	0.038	0.955	0.008	-0.028	0.705	0.097	-0.756	1.763	4000
$\mathcal{M}_{15}$	0.143	0.010	0.501	0.465	0.235	0.974	0.058	-0.012	0.921	0.179	-1.134	1.043	4500

Note: The table summarizes the properties of the numerical simulation of the posterior densities of all the considered models. For each of the statistics, the median of all the parameters of the model, as well as minimum and maximum, are reported. The table reports the relative numerical efficiency coefficient (RNE), by Geweke (1989), autocorrelations of the MCMC draws at lags 1 and 10, as well as 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 numbers presented in this table were obtained using the package coda by Plummer et al. (2006).

DOI: 10.2870/63858

Bibliography

Bartlett, M. S. (1957). A Comment on D. V. Lindley's Statistical Paradox. *Biometrika 44*(3/4), 533–534.

Bauwens, L. and M. Lubrano (1998). Bayesian inference on GARCH models using the Gibbs sampler. *Econometrics Journal 1*(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*(3), 498–505.

Boudjellaba, H., J.-M. Dufour, and R. Roy (1992). Testing Causality Between Two Vectors in Multivariate Autoregressive Moving Average Models. *Journal of the American Statistical Association 87*(420), 1082–1090.

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

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician 49*(4), 327.

Comte, F. and O. Lieberman (2000). Second-Order Noncausality in Multivariate GARCH Processes. *Journal of Time Series Analysis 21*(5), 535–557.

Comte, F. and O. Lieberman (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis 84*(1), 61–84.

Conrad, C. and M. Karanasos (2010). Negative Volatility Spillovers in the Unrestricted ECCC-GARCH Model. *Econometric Theory 26*(3), 838–862.

Deschamps, P. J. (2006). A flexible prior distribution for Markov switching autoregressions with Student-t errors. *Journal of Econometrics 133*(1), 153–190.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Droumaguet, M. and T. Woźniak (2012). Bayesian Testing of Granger Causality in Markov-Switching VARs. Working paper series, European University Institute, Florence, Italy. Download at: http://cadmus.eui.eu/bitstream/handle/1814/20815/ECO_2012_06.pdf?sequence=1.

Dufour, J.-M., D. Pelletier, and E. Renault (2006). Short run and long run causality in time series: inference. *Journal of Econometrics 132*(2), 337–362.

Dufour, J.-M. and E. Renault (1998). Short Run and Long Run Causality in Time Series : Theory. *Econometrica 66*(5), 1099–1125.

Engle, R. F., T. Ito, and W.-l. Lin (1990). Meteor Showers or Heat Waves ? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market. *Econometrica 58*(3), 525–542.

Fiorentini, G., E. Sentana, and G. Calzolari (2003). Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroscedastic Dynamic Regression Models With Student t Innovations. *Journal of Business & Economic Statistics 21*(4), 532–546.

Florens, J. P. and M. Mouchart (1985). A Linear Theory for Noncausality. *Econometrica 53*(1), 157–176.

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *Econometrics Journal 7*(1), 143–167.

Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica 57*(6), 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, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, Volume 148. Clarendon Press, Oxford.

Geweke, J. (1995). Bayesian Comparison of Econometric Models. *Federal Reserve Bank of Minneapolis*, 1–36.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Geweke, J. (1997). *Posterior Simulators in Econometrics*, Volume 3 of *Econometric Society Monographs*, Chapter Posterior. Cambridge University Press.

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

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica 37*(3), 424–438.

Hafner, C. M. and H. Herwartz (2006). A Lagrange multiplier test for causality in variance. *Economics Letters 93*(1), 137–141.

Hafner, C. M. and H. Herwartz (2008a). Analytical quasi maximum likelihood inference in multivariate volatility models. *Metrika 67*(2), 219–239.

Hafner, C. M. and H. Herwartz (2008b). Testing for Causality in Variance Using Multivariate GARCH Models. *Annales d'Économie et de Statistique 89*, 215–241.

He, C. and T. Teräsvirta (2004). An Extended Constant Conditional Correlation Garch Model and Its Fourth-Moment Structure. *Econometric Theory 20*, 904–926.

Hoogerheide, L. F., H. K. van Dijk, and van Oest R.D. (2009). *Simulation Based Bayesian Econometric Inference: Principles and Some Recent Computational Advances*, Chapter 7, pp. 215–280. Handbook of Computational Econometrics. Wiley.

Jarociński, M. and B. Maćkowak (2011). Choice of Variables in Vector Autoregressions.

Jarque, C. M. and A. K. Bera (1980). Efficient Tests for Normality, Homoscedasticity and Serial Indipendence of Regression Residuals. *Economics Letters 6*, 255–259.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate arch models. *Econometric Theory 14*(01), 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 & Economic Statistics 13*(1), 11–25.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

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

Ling, S. and M. McAleer (2003). Asymtotic Theory for a Vector ARMA-GARCH Model. *Econometric Theory 19*(2), 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. and M. M. Burda (1997). Modified Wald tests under nonregular conditions. *Journal of Econometrics 78*(1), 315–332.

Miazhynskaia, T. and G. Dorffner (2006). A comparison of Bayesian model selection based on MCMC with an application to GARCH-type models. *Statistical Papers 47*(4), 525–549.

Nakatani, T. and T. Teräsvirta (2009). Testing for volatility interactions in the Constant Conditional Correlation GARCH model. *Econometrics Journal 12*(1), 147–163.

Newton, M. A. and A. E. Raftery (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological) 56*(1), 3–48.

Osiewalski, J. and M. Pipień (2002). Multuvariate t-GARCH Models - Bayesian Analysis for Exchange Rates. In *Modelling Economies in Transition - Proceedings of the Sixth AMFET Conference*, Łódź, pp. 151–167. Absolwent.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Osiewalski, J. and M. Pipień (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*(2), 49–73.

Plummer, M., N. Best, K. Cowles, and K. Vines (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News 6*(1), 7–11.

Robins, R. P., C. W. J. Granger, and R. F. Engle (1986). Wholesale and Retail Prices: Bivariate Time-Series Modeling with forecastable Error Variances. In *Model Reliability*, pp. 1–17. The MIT Press.

Sims, C. A. (1972). Money, Income, and Causality. *The American Economic Review 62*(4), 540 – 552.

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

Strachan, R. W. and H. K. van Dijk (2011). Divergent Priors and Well Behaved Bayes Factors.

Vrontos, I. D., P. Dellaportas, and D. N. Politis (2003). Inference for some multivariate ARCH and GARCH models. *Journal of Forecasting 22*(6-7), 427–446.

Woźniak, T. (2012). Granger causal analysis of varma-garch models. EUI Working Papers ECO 2012/19, European University Institute, Florence, Italy. Download at: http://cadmus.eui.eu/bitstream/handle/1814/23336/ECO_2012_19.pdf.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Chapter 2

Granger-causal analysis of VARMA-GARCH models

Abstract. Recent economic developments have shown the importance of spillover and contagion effects in financial markets. Such effects are not limited to relations between the levels of financial variables but also impact on their volatility. Granger causality in conditional mean and conditional variances of time series is investigated. For this purpose a VARMA-GARCH model is used. Parametric restrictions for the hypothesis of noncausality in conditional variances between two groups of variables, when there are other variables in the system as well are derived. These novel conditions are convenient for the analysis of potentially large systems of economic variables. Such systems should be considered in order to avoid the problem of omitted variable bias. Further, in order to evaluate hypotheses of noncausality, a Bayesian testing procedure is proposed. It avoids the singularity problem that may appear in the Wald test. This approach also enables the assumption of the existence of higher-order moments of the residuals required for the derivation of asymptotic results for classical tests to be relaxed. In the empirical example, the dollar-to-Euro exchange rate is found not to second-order cause the pound-to-Euro exchange rate, in the system of variables containing also the Swiss frank-to-Euro exchange rate, which confirms the *meteor shower* hypothesis of Engle, Ito and Lin (1990).

The author thanks Professors Helmut Lütkepohl, Jacek Osiewalski, William Griffiths, Timo

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

2.1 Introduction

The well-known concept of *Granger causality* (see Granger, 1969; Sims, 1972) describes relations between time series in the forecasting context. One variable does not Granger-cause the other, if adding past observations of the former to the information set with which we forecast the latter does not improve this forecast. This study investigates the Granger noncausality concept for conditional variances of the time series. For this purpose two concepts of *second-order Granger noncausality* and *Granger noncausality in variance* are discussed (see also Comte and Lieberman, 2000; Robins et al., 1986). If one variable does not second-order Granger-cause the other, then past information about the variability of the former is dispensable for conditional variance the forecasting of the conditional variances of the latter. We investigate Granger causality in conditional mean and conditional variances of time series. *Granger noncausality in variance* is established when both *Granger noncausality* and *second-order noncausality* hold.

The necessity of the joint analysis is justified for two reasons. Firstly, as Karolyi (1995) argues, in order to have a good picture of transmissions in mean between financial variables, transmissions in volatility need to be taken into account. Secondly, transmissions in volatility may be affected by transmissions in mean that have not been modeled and filtered out before, a point made by Hong (2001). The conclusion is that the combined modeling of the conditional mean and conditional variance processes increases the reliability of the inference about the transmissions. The exposition of the phenomenon in this paper is done entirely with a vector autoregressive moving average (VARMA) conditional mean process, with a generalized autoregressive conditional heteroskedasticity (GARCH) process for

Teräsvirta, Massimiliano Marcellino, Giampiero Gallo, Michael Smith, Mateusz Pipień, Russell Cooper and Jérôme Adda for useful comments and remarks on the study. The author is grateful to the participants of the seminars at the Cracow University of Economics, Monash University, the University of Queensland, the University of Melbourne and Università di Pisa. The paper was presented at the SMYE 2011 Conference at the University of Groningen, the ISBA 2012 World Meeting, Kyoto, Japan, and at Bayes on the Beach 2012 Conference. Also, the author thanks thank Michał Markun and Norbert Metiu and his colleagues from the EUI Time Series Econometrics Working Group, Pierre Guérin, Matthieu Droumaguet and Stelios Bekiros, for multiple discussions and suggestions.

conditional variances and constant conditional correlations (CCC).

Why is information about Granger-causal relations between time series important? First of all, it gives an understanding of the structure of the financial markets. More specifically, we learn about integration of the financial markets (assets) not only in returns, but also in risk, defined as time-varying volatility. Therefore, modeling transmissions in volatility may have a significant impact on volatility forecasting. If there are Granger-causal relations in conditional variances, then such modeling is potentially important in all applications based on volatility forecasting such as portfolio selection, Value at Risk estimation and option pricing.

Granger-causality relations established in conditional variances of exchange rates are in line with some economic theories. Taylor (1995) shows that they are consistent with failures of the exchange rates market efficiency. The arrival of news, in clusters and potentially with a lag, modeled with GARCH models explains the inefficiency of the market. It is also in line with a market dynamics that exhibits volatility persistence due to private information or heterogeneous beliefs (see Hong, 2001, and references therein). Finally, the *meteor showers* hypothesis for intra-daily exchange rates returns, which reflects cooperative or competitive monetary policies (see Engle et al., 1990), can be presented as a Granger second-order noncausality hypothesis.

The term *transmissions* usually represents an intuitive interpretation of the parameters, reflecting the impact of one variable on the other in dynamic systems. Karolyi (1995) and Lin et al. (1994) use the term to describe international transmissions between stock returns and their volatilities. Further, Nakatani and Teräsvirta (2009) and Koutmos and Booth (1995) use it to describe the interactions between volatilities in multivariate GARCH models. Another term, *volatility spillovers*, has been used in a similar context (see e.g. Conrad and Karanasos, 2009), as well as in others. However, parameters referred to in this way do not determine Granger causality or noncausality themselves. This study presents parameter conditions for the precisely defined Granger noncausality concept for conditional variances. In particular, the framework of the linear Granger noncausality of Florens and Mouchart (1985) is referred to, and which defined the noncausality relationship in

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

terms of the orthogonality in the Hilbert space of square integrable variables.

The contribution of this study is twofold. Firstly, conditions for second-order Granger noncausality for a family of GARCH models are derived. The conditions are applicable when the system of time series consists of a potentially large number of variables. Their novelty is that the second-order noncausality between two groups of variables is analyzed when there are other variables in the system as well. So far, such conditions have been derived when all the variables in the system were divided in two groups (e.g. Comte and Lieberman, 2000; Hafner and Herwartz, 2008; Woźniak, 2012). The introduced conditions reduce the dimensionality of the problem. They also allow the formation and testing of some hypotheses that could not be tested in the previous settings.

Secondly, a Bayesian testing procedure of the conditions for Granger noncausality in conditional mean and noncausality in conditional variance processes is proposed. It is easily applicable and solves some of the drawbacks of the classical testing. In comparison with the Wald test of Boudjellaba et al. (1992), adapted to testing noncausality relations in the VARMA-GARCH model, the Bayesian test does not have the problem of singularities. In the Wald test considered so far the singularities appear due to the construction of the asymptotic covariance matrix of the nonlinear parametric restrictions. In Bayesian analysis, on the contrary, the posterior distribution of the restrictions is available; thus, a well defined covariance matrix is available as well. Additionally, in this study the existence only of fourth-order moments of time series is assumed, which is an improvement in comparison with the assumptions of available classical tests.

The reminder of this paper is organized as follows: the notation and the parameter restrictions for Granger noncausality in VARMA models are presented in Section 2.2. The GARCH model used in the analysis is set in Section 2.3. Also this section presents the main theoretical findings of the paper, deriving the conditions for Granger noncausality in the conditional variance process. Section 2.4 discusses classical testing for noncausality in the VARMA-GARCH models, and then proposes Bayesian testing with appealing properties. Section 2.5 presents an empirical illustration, with the example of daily exchange rates of the Swiss franc, the British pound and the US dollar all denominated in Euro. Section 2.6 concludes.

2.2 Granger noncausality in VARMA models

First, we set the notation 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 $\mathbb{Z}$. Write:

$$y_t = (y_{1t}^{'}, y_{2t}^{'}, y_{3t}^{'})^{'}, \tag{2.1}$$

where y_{it} is a $N_i \times 1$ vector such that $y_{1t} = (\mathrm{y}_{1t}, \dots, \mathrm{y}_{N_1.t})^{'}$, $y_{2t} = (\mathrm{y}_{N_1+1.t}, \dots, \mathrm{y}_{N_1+N_2.t})^{'}$ and $y_{3t} = (\mathrm{y}_{N_1+N_2+1.t}, \dots, \mathrm{y}_{N_1+N_2+N_3.t})^{'}$ ($N_1, N_2 \geq 1, N_3 \geq 0$ and $N_1 + N_2 + N_3 = N$). Variables of interest are contained in vectors y_1 and y_2, between which we want to study causal relations. Vector y_3 (which for $N_3 = 0$ is empty) contains auxiliary variables that are also used for forecasting and modeling purposes. 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.

Let $I_y^2(t)$ be the Hilbert space generated by product of variables, $y_{i\tau}y_{j\tau}$, and $I_\epsilon^2(t)$ generated by products of error components, $\epsilon_{i\tau}\epsilon_{j\tau}$, where $1 \leq i, j \leq N$ and for $\tau \leq t$. $I_{-1}(t)$ is the closed subspace of $I(t)$ generated by the components of $(y_{2\tau}^{'}, y_{3\tau}^{'})'$. $I_{y.-1}$ is the closed subspace of $I_y^2(t)$ generated by variables $y_{i\tau}y_{j\tau}$ and $I_{\epsilon.-1}^2(t)$ is the closed subspace of $I^2(t)$ generated by the variables $\epsilon_{i\tau}\epsilon_{j\tau}$, where $N_1 + 1 \leq i, j \leq N$ and for $\tau \leq t$. For any subspace I_t of $I(t)$ and for $N_1 + 1 \leq i \leq N_1 + N_2$, $P(y_{it+1}|I_t)$ denotes the affine projection of y_{it+1} on I_t, i.e. the best linear prediction of y_{it+1}, based on the variables in I_t and a constant term.

For the Granger causal analysis of stochastic processes the modeling framework of the VARMA-GARCH processes is considered. This approach is practical for empirical work. Florens and Mouchart (1985) treated the problem of causality at the high level of generality, without any particular process assumed. Granger noncausality in mean from y_1 to y_2 is defined as follows.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Definition 2. y_1 does not Granger-cause y_2 in mean, given y_3, denoted by $y_1 \overset{G}{\nrightarrow} y_2|y_3$, if each component of the error vector, $y_{2t+1} - P(y_{2t+1}|I_{-1}(t))$, is orthogonal to $I(t)$ for all $t \in \mathbb{Z}$.

Definition 2, proposed by Boudjellaba et al. (1992), states simply that the forecast of y_2 cannot be improved by adding to the information set past realizations of y_1.

Suppose that y_t follows a N-dimensional VARMA(p,q) process:

$$\alpha(L)y_t = \beta(L)\epsilon_t, \tag{2.2}$$

for all $t = 1, \ldots, T$, where L is a lag operator such that $L^i y_t = y_{t-i}$, $\alpha(z) = I_N - \alpha_1 z - \cdots - \alpha_p z^p$, $\beta(z) = I_N + \beta_1 z + \cdots + \beta_q z^q$ are matrix polynomials. I_N denotes the identity matrix of order N, and $\{\epsilon_t : t \in \mathbb{Z}\}$ is a white noise process with nonsingular unconditional covariance matrix $\mathbf{V}$. Comte and Lieberman (2000) mention that all the results in this section hold also if $E[\epsilon_t \epsilon_t' | I^2(t-1)] = H_t$, i.e. if the conditional covariance matrix of ϵ_t is time-varying, provided that unconditional covariance matrix, $E[H_t] = \mathbf{V}$, is constant and nonsingular. Without the loss of generality, it is assumed in (2.2) that $E[y_t] = 0$, however any deterministic terms, such as a vector of constants, a time trend or seasonal dummies may be considered for modeling. Further assumptions for the process (2.2) are:

Assumption 6. All the roots of $|\alpha(z)| = 0$ and all the roots of $|\beta(z)| = 0$ are outside the complex unit circle.

Assumption 7. The terms $\alpha(z)$ and $\beta(z)$ are left coprime and satisfy other identifiability conditions given in Lütkepohl (2005).

These assumptions guarantee that the VARMA(p,q) process is stationary, invertible and identified. Let the vector y_t be partitioned, as in (3.13), then we can

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

write (2.2) as:

$$\begin{bmatrix} \alpha_{11}(L) & \alpha_{12}(L) & \alpha_{13}(L) \\ \alpha_{21}(L) & \alpha_{22}(L) & \alpha_{23}(L) \\ \alpha_{31}(L) & \alpha_{32}(L) & \alpha_{33}(L) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} \beta_{11}(L) & \beta_{12}(L) & \beta_{13}(L) \\ \beta_{21}(L) & \beta_{22}(L) & \beta_{23}(L) \\ \beta_{31}(L) & \beta_{32}(L) & \beta_{33}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix}. \tag{2.3}$$

Given Assumptions 6–7 and the VARMA(p,q) process in the form as in (2.3), Theorem 4 of Boudjellaba et al. (1994) states the conditions for Granger noncausality. Therefore, y_1 does not Granger-cause y_2 given y_3 ($y_1 \overset{G}{\nrightarrow} y_2|y_3$) if and only if:

$$\Gamma_{ij}(z) = \det \begin{bmatrix} \alpha_{11}^{\cdot j}(z) & \beta_{11}(z) & \beta_{13}(z) \\ \tilde{\alpha}_{N_1+i,j}(z) & \beta_{21}^{i\cdot}(z) & \beta_{23}^{i\cdot}(z) \\ \alpha_{31}^{\cdot j}(z) & \beta_{31}(z) & \beta_{33}(z) \end{bmatrix} = 0 \qquad \forall z \in \mathbb{C}, \tag{2.4}$$

for $i = 1, ..., N_2$ and $j = 1, ..., N_1$; where $\alpha_{lk}^{\cdot j}(z)$ is the jth column of $\alpha_{lk}(z)$, $\beta_{lk}^{i\cdot}(z)$ is the ith row of $\beta_{lk}(z)$, and $\tilde{\alpha}_{N_1+1,j}(z)$ is the (i, j)-element of $\alpha_{21}(z)$.

In general, condition (2.4) leads to N_1N_2 determinant conditions. Each of them can be represented in the form of a polynomial in z of degree $p+q(N_1+N_3)$: $\Gamma_{ij}(z) = \sum_{i=1}^{p+q(N_1+N_3)} a_i z^i$, where a_i are nonlinear functions of parameters of the VARMA process. Notice that $\Gamma_{ij} = 0 \Rightarrow a_i = 0$ for $i = 1, \ldots, p + q(N_1 + N_3)$, which gives restrictions for Granger noncausality.

Example 3. Let y_t be $N = 3$ dimensional VARMA(1,0) process, $N_1 = N_2 = N_3 = 1$ and let one be interested in whether y_1 Granger-causes y_2. The restriction for such a case is:

$$\mathbf{R}^I(\theta) = \alpha_{21} = 0, \tag{2.5}$$

where θ is a vector containing all the parameters of the model, $\theta \in \Theta \subset \mathbb{R}^k$, and k denotes the dimension of θ.

Example 4. Let y_t be the VARMA(1,1) process of the same dimension and partitioning as before. Determinant condition (2.4) leads to the following set of restrictions:

$$\mathbf{R}_1^{II}(\theta) = \alpha_{11}(\beta_{23}\beta_{31} - \beta_{21}\beta_{33}) + \beta_{21}(\beta_{11}\beta_{33} - \beta_{13}\beta_{31}) + \alpha_{31}(\beta_{13}\beta_{21} - \beta_{11}\beta_{23}) = 0 \quad (2.6a)$$

$$\mathbf{R}_2^{II}(\theta) = \beta_{21}(\alpha_{11} - 2\beta_{33} - \beta_{11}) + \beta_{23}(\alpha_{31} - \beta_{31}) = 0 \quad (2.6b)$$

$$\mathbf{R}_3^{II}(\theta) = \alpha_{21} - \beta_{21} = 0, \quad (2.6c)$$

and let $\mathbf{R}^{II}(\theta) = (\mathbf{R}_1^{II}(\theta), \mathbf{R}_2^{II}(\theta), \mathbf{R}_3^{II}(\theta))'$ be a vector collecting the values of the restrictions on the LHS.

The problem of testing restrictions (2.5) and (2.6) is dealt with in Section 2.4.

2.3 Parameter restrictions for second-order Granger noncausality in GARCH models

This section consists of two parts. The first presents a multivariate GARCH model with constant conditional correlations. For this model such topics as conditions for stationarity, asymptotic properties, classical and Bayesian estimation and how it was used to model and test volatility transmissions are discussed. The second part of this section presents VARMA and VAR representations of the GARCH process in order to derive parametric conditions for second-order Granger noncausality.

GARCH(r,s) model and its properties The conditional mean part of the model is described with the VARMA process (2.2) and a residual term ϵ_t following a conditional variance process:

$$\epsilon_t = D_t r_t, \quad (2.7a)$$

$$r_t \sim i.i.d.(0, \mathbf{C}), \quad (2.7b)$$

for all $t = 1, \ldots, T$, where $D_{ii.t} = [\sqrt{h_{i.t}}]$ for $i = 1, \ldots, N$ is a $N \times N$ diagonal matrix with conditional standard deviations on the diagonal, r_t is a vector of standardized residuals that follows *i.i.d.* with zero mean and a correlation matrix $\mathbf{C}$.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Conditional variances of ϵ_t follow the multivariate GARCH(r,s) process of Jeantheau (1998):

$$h_t = \omega + A(L)\epsilon_t^{(2)} + B(L)h_t, \tag{2.8}$$

for all $t = 1, \ldots, T$, where h_t is a $N \times 1$ vector of conditional variances of ϵ_t, ω is a $N \times 1$ vector of constant terms, $\epsilon_t^{(2)} = (\epsilon_{1t}^2, \ldots, \epsilon_{Nt}^2)'$ is a vector of squared residuals, $A(L) = \sum_{i=1}^{r} A_i L^i$ and $B(L) = \sum_{i=1}^{s} B_i L^i$ are matrix polynomials of ARCH and GARCH effects, respectively. All the matrices in $A(L)$ and $B(L)$ are of dimension $N \times N$ and allow for volatility transmissions from one series to another. $\mathbf{C}$ is a positive definite constant conditional correlation matrix with ones on the diagonal.

The conditional covariance matrix of the residual term ϵ_t is decomposed into $E[\epsilon_t\epsilon_t'|I_y^2(t-1)] = H_t = D_t\mathbf{C}D_t$. For the matrix H_t to be a well defined positive definite covariance matrix, h_t must be positive for all t, and $\mathbf{C}$ positive definite (see Bollerslev, 1990). Given the normality of r_t, the vector of conditional variances is $E[\epsilon_t^{(2)}|I_y^2(t-1)] = h_t$. When r_t follows a t distribution with ν degrees of freedom, the conditional variances are $E[\epsilon_t^{(2)}|I^2(t-1)] = \frac{\nu}{\nu-2}h_t$. In both cases the best linear predictor of $\epsilon_t^{(2)}$ is $h_t = P(\epsilon_t^{(2)}|I_y^2(t-1))$.

The VARMA(p,q)-GARCH(r,s) model described by (2.2), (2.7) and (2.8), which is object of the analysis in this study, has its origins in the constant conditional correlation GARCH (CCC-GARCH) model proposed by Bollerslev (1990). That model consists of N univariate GARCH equations describing the vector of conditional variances h_t. The CCC-GARCH model is equivalent to equations (2.7) and (2.8) with diagonal matrices $A(L)$ and $B(L)$. Its extended version, with non-diagonal matrices $A(L)$ and $B(L)$, was used in Karolyi (1995) and analyzed by Jeantheau (1998). He and Teräsvirta (2004) called this model extended CCC-GARCH (ECCC-GARCH).

Jeantheau (1998) proves that the GARCH(r,s) model, as in (2.8), has a unique, ergodic, weakly and strictly stationary solution when $\det[I_N - A(z) - B(z)] = 0$ has its unit roots outside the complex unit circle. He and Teräsvirta (2004) give sufficient conditions for the existence of the fourth moments and derive complete structure of fourth moments. For instance, they give the conditions for existence

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

and analytical form of $E[\epsilon_t^{(2)}\epsilon_t^{(2)\prime}]$, as well as for the nth order autocorrelation matrix of $\epsilon_t^{(2)}$, $\mathbf{R}_N(n) = D_N^{-1}\Gamma_N(n)D_N^{-1}$, where $\Gamma_N(n) = [\gamma_{ij}(n)] = E[(\epsilon_t^{(2)} - \sigma^2)(\epsilon_{t-n}^{(2)} - \sigma^2)']$ and $D_{ii.N} = [\sqrt{\sigma_i^2}]$ for $i = 1, \ldots, N$.

The VARMA-ECCC-GARCH model has well established asymptotic properties. They can be set under the following assumptions:

Assumption 8. 1. All the roots of $|I_N - A(z) - B(z)| = 0$ are outside the complex unit circle. 2. All the roots of $|I_N - B(z)| = 0$ are outside the unit circle.

Assumption 9. The multivariate GARCH(r,s) model is minimal, in the sense of Jeantheau (1998).

Under Assumptions 8.1 and 9, the GARCH(r,s) model is stationary and identifiable. Jeantheau (1998) showed that the minimum contrast estimator for the multivariate GARCH model is strongly consistent under, among others, stationarity and identifiability conditions. Ling and McAleer (2003) proved the strong consistency of the QMLE for the VARMA-GARCH model under Assumptions 6–9. Moreover, they have set the asymptotic normality of QMLE, provided that $E\|y_t\|^6 < \infty$.

It was already mentioned that for positive definiteness of conditional covariance matrix, H_t, h_t has to be positive for all t. Usual parameter conditions for h_t to be positive are $\omega > 0$ and $[A_i]_{jk}, [B_l]_{jk} \geq 0$ for $i = 1, \ldots, r$, $l = 1, \ldots, s$ and $j, k = 1, \ldots, N$. Conrad and Karanasos (2009) derived conditions such that some elements of A_i, B_l $(i = 1, \ldots, r;\ l = 1, \ldots, s)$ and even ω are allowed to be negative. Still, it is not known whether asymptotic results hold under these conditions. However, their empirical usefulness has been proven, as Conrad and Karanasos (2009) have found that some parameters of the model responsible for volatility transmissions are negative.

Classical estimation with the maximum likelihood method has been presented in Bollerslev (1990). The maximum likelihood estimator is the argument maximizing the likelihood function, $\hat{\theta} = \arg\max_{\theta\in\Theta} L(\theta; \mathbf{y})$. The likelihood functions for

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Normal and t-distributed ϵs are, respectively:

$$L_N(\theta;\mathbf{y}) = (2\pi)^{-TN/2}\prod_{t=1}^{T}|H_t|^{-1/2}\exp\left(\epsilon_t' H_t^{-1}\epsilon_t\right)^{-1/2}, \text{ and} \tag{2.9a}$$

$$L_{St}(\theta;\mathbf{y}) = \prod_{t=1}^{T}\frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}((\nu-2)\pi)^{-\frac{N}{2}}|H_t|^{-\frac{1}{2}}\left(1+\frac{1}{\nu-2}\epsilon_t' H_t^{-1}\epsilon_t\right)^{-\frac{\nu+N}{2}}, \tag{2.9b}$$

where ϵ_t is defined in equations (2.2) and (2.7). $\Gamma(.)$ is Euler's gamma and $|.|$ a matrix determinant. Algorithms maximizing the likelihood function, such as the BHHH algorithm (see Berndt et al., 1974), use analytical derivatives. Fiorentini et al. (2003) provide analytical expressions for the score, Hessian, and information matrix of multivariate GARCH models with t conditional distributions of residuals. In the Bayesian estimation of the GARCH models, numerical integration methods are used. Vrontos et al. (2003) propose Metropolis-Hastings algorithm (see Chib and Greenberg, 1995, and references therein) for the estimation of the model.

A broad family of GARCH models has already been used in the volatility spillovers literature. More specifically, the empirical works of Worthington and Higgs (2004) and Caporale et al. (2006) use the BEKK-GARCH model of Engle and Kroner (1995) to prove volatility transmissions between stock exchange indices. The issue of causality in variance or second-order causality (both defined in the next paragraph) has been treated by Comte and Lieberman (2000), who derived the conditions on parameters of the model for second-order noncausality between two vectors of variables. No testing procedure, however, was available due to the lack of asymptotic results. Comte and Lieberman (2003) filled in the gap, deriving asymptotic normal distribution for QMLE under the assumption of bounded moments of order eight for ϵ_t. Hafner and Herwartz (2008) use the results of these two papers and propose a Wald statistics for sufficient conditions for noncausality in variance hypothesis. As a consequence of using asymptotic derivations of Comte and Lieberman (2003), the test also requires finiteness of eighth-order moments of the error term. Hafner (2009) presents the conditions under which temporal aggre-

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

gation in GARCH models does not influence testing of the causality in conditional variances.

Karolyi (1995) uses the VARMA-ECCC-GARCH model to show the necessity of modeling the volatility spillovers for the inference about transmissions in returns of stock exchange indexes. The assumption of constant conditional correlation may be too strong for such data. The ECCC-GARCH model, however, proved its usefulness in modeling the volatility of the exchange rates. In a recent study, Omrane and Hafner (2009) use the trivariate model for volatility spillovers between exchange rates. Conrad and Karanasos (2009) and Nakatani and Teräsvirta (2008) show the important case that volatility transmissions may be negative, the former for the system containing inflation rate and output growth, and the latter for Japanese stock returns. A formal test for the volatility transmissions has been proposed by Nakatani and Teräsvirta (2009). Their Lagrange multiplier test statistics for the hypothesis of no volatility transmissions ($A(L)$ and $B(L)$ diagonal) versus volatility transmissions ($A(L)$ and $B(L)$ non-diagonal) assumes the existence of sixth-order moments of the residual term, $E|\epsilon_t^6| < \infty$. Woźniak (2012) introduces the notion of Granger second-order causality and causality in variance for ECCC-GARCH models for the setting similar to that of Comte and Lieberman (2000), in which the vector of variables is partitioned in two parts. The current paper extends the analysis such that an inference about causality between two (vectors of) variables is performed when there are also other variables in the system used for forecasting.

Before the notion of Granger noncausality for conditional variances is presented, the GARCH(r,s) model, (2.8), is rewritten into its VARMA and VAR representations. Define a process $\nu_t = \epsilon_t^{(2)} - h_t$. Then $\epsilon_t^{(2)}$ follows a VARMA process given by:

$$\phi(L)\epsilon_t^{(2)} = \omega + \psi(L)\nu_t, \tag{2.10}$$

for all $t = 1, \dots, T$, 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(r,s) process. Suppose $\epsilon_t^{(2)}$ and ν_t are partitioned analogously as y_t in (3.13). Then (2.10) can be written in the form:

$$\begin{bmatrix}\phi_{11}(L) & \phi_{12}(L) & \phi_{13}(L)\\ \phi_{21}(L) & \phi_{22}(L) & \phi_{23}(L)\\ \phi_{31}(L) & \phi_{32}(L) & \phi_{33}(L)\end{bmatrix}\begin{bmatrix}\epsilon_{1t}^{(2)}\\ \epsilon_{2t}^{(2)}\\ \epsilon_{3t}^{(2)}\end{bmatrix} = \begin{bmatrix}\omega_{1t}\\ \omega_{2t}\\ \omega_{3t}\end{bmatrix} + \begin{bmatrix}\psi_{11}(L) & \psi_{12}(L) & \psi_{13}(L)\\ \psi_{21}(L) & \psi_{22}(L) & \psi_{23}(L)\\ \psi_{31}(L) & \psi_{32}(L) & \psi_{33}(L)\end{bmatrix}\begin{bmatrix}\nu_{1t}\\ \nu_{2t}\\ \nu_{3t}\end{bmatrix}. \tag{2.11}$$

Given Assumption 8.2, the VARMA process (2.10) is invertible and can be written in the VAR form:

$$\Pi(L)\epsilon_t^{(2)} - \omega^* = \nu_t, \tag{2.12}$$

for all $t = 1, \dots, T$, where $\Pi(L) = \psi(L)^{-1}\phi(L) = [I_N - B(L)]^{-1}[I_N - A(L) - B(L)]$ is a matrix polynomial of potentially infinite order of the VAR representation of the GARCH(r,s) process and $\omega^* = \psi(1)^{-1}\omega$ is a constant term. Again, partitioning the vectors, rewrite (2.12) in the form:

$$\begin{bmatrix}\Pi_{11}(L) & \Pi_{12}(L) & \Pi_{13}(L)\\ \Pi_{21}(L) & \Pi_{22}(L) & \Pi_{23}(L)\\ \Pi_{31}(L) & \Pi_{32}(L) & \Pi_{33}(L)\end{bmatrix}\begin{bmatrix}\epsilon_{1t}^{(2)}\\ \epsilon_{2t}^{(2)}\\ \epsilon_{3t}^{(2)}\end{bmatrix} - \begin{bmatrix}\omega_{1t}^*\\ \omega_{2t}^*\\ \omega_{3t}^*\end{bmatrix} = \begin{bmatrix}\nu_{1t}\\ \nu_{2t}\\ \nu_{3t}\end{bmatrix}. \tag{2.13}$$

Under Assumption 8, both processes (2.10) and (2.12) are stationary.

Noncausality restrictions This paragraph presents the main theoretical findings of the paper, that is the derivation of the conditions for second-order Granger noncausality for the ECCC-GARCH model. Two concepts are defined: Granger noncausality in variance and second-order Granger noncausality. Further, the parametric conditions in Theorems 2 and 3 are derived and their novelty is discussed.

Robins et al. (1986) introduced the concept of Granger causality for conditional variances. Comte and Lieberman (2000) call this concept *second-order Granger causality* and distinguish it from *Granger causality in variance*. These noncausalities are defined slightly differently than Comte and Lieberman (2000) do. In the definition for second-order noncausality below, the Hilbert space $I_\epsilon^2(t)$ is used, whereas Comte and Lieberman use $I_y^2(t)$. The definitions are in the following forms:

Definition 3. y_1 does not second-order Granger-cause y_2 given y_3, denoted by $y_1 \overset{so}{\nrightarrow} y_2|y_3$, if:

$$P\Big([y_{2t+1} - P(y_{2t+1}|I(t))]^{(2)}|I^2_{\epsilon}(t)\Big) = P\Big([y_{2t+1} - P(y_{2t+1}|I(t))]^{(2)}|I^2_{\epsilon.-1}(t)\Big) \quad \forall t \in \mathbb{Z}.$$

Definition 4. y_1 does not Granger-cause y_2 in variance given y_3, denoted by $y_1 \overset{v}{\nrightarrow} y_2|y_3$, if:

$$P\Big([y_{2t+1} - P(y_{2t+1}|I(t))]^{(2)}|I^2_{y}(t)\Big) = P\Big([y_{2t+1} - P(y_{2t+1}|I_{-1}(t))]^{(2)}|I^2_{y.-1}(t)\Big) \quad \forall t \in \mathbb{Z}, \quad (2.14)$$

where $[.]^{(2)}$ means that we square every element of a vector. Another difference between the two definitions is in the Hilbert spaces on which y_{2t+1} is projected. On the right-hand side of Definition 3 we take the affine projection of y_{2t+1} on $I(t)$, whereas on the right-hand side of Definition 4 we take the affine projection of y_{2t+1} on $I_{-1}(t)$. In other words, before considering whether there is *second-order Granger noncausality*, one first needs to model and to filter out the Granger causality in mean. Further, an implicit assumption in the definition of *Granger noncausality in variance* is that y_1 does not Granger-cause in mean y_2, $y_1 \overset{G}{\nrightarrow} y_2|y_3$. The relation between Granger noncausality in mean, noncausality in variance and second-order noncausality have been established by Comte and Lieberman (2000) and are as follows:

$$y_1 \overset{v}{\nrightarrow} y_2|y_3 \Leftrightarrow (y_1 \overset{G}{\nrightarrow} y_2|y_3 \text{ and } y_1 \overset{so}{\nrightarrow} y_2|y_3). \quad (2.15)$$

One implication of this statement is that Definitions 3 and 4 are equivalent when y_1 does not Granger-cause y_2. And conversely, if y_1 Granger-causes y_2, then the Granger noncausality in variance is excluded, but still y_1 may not second-order cause y_2.

Under Assumptions 6–9, the VARMA-ECCC-GARCH model is stationary, identifiable and invertible in both of its parts: VARMA processes for y_t and for $\epsilon_t^{(2)}$. One more assumption is needed in order to state noncausality relations in the conditional variances process:

Assumption 10. The process v_t is covariance stationary with covariance matrix $\mathbf{V}_v$.

A theorem introduces second-order Granger noncausality relations:

Theorem 2. *Let $\epsilon_t^{(2)}$ follow a stationary vector autoregressive process, as in (2.12), partitioned, as in (2.13), that is identifiable (Assumptions 8–10). Then, y_1 does not second-order Granger-cause y_2 given y_3 (denoted by $y_1 \overset{so}{\nrightarrow} y_2|y_3$) if and only if:*

$$\Pi_{21}(z) \equiv 0 \quad \forall z \in \mathbb{C}. \tag{2.16}$$

Proof. Theorem 2 may be proved by applying Proposition 1 of Boudjellaba et al. (1992). However, since that proof is derived for the VAR models, several modifications are required to make it applicable to the GARCH model of Jeantheau (1998) in the VAR form, as in (2.12) and (2.13). Here the squared elements of the residual term, $\epsilon^{(2)}(y_{2t+1}|I(t))$, are projected on the Hilbert spaces $I_\epsilon^2(t)$ or $I_{\epsilon.-1}^2(t)$, both defined in Section 2.2. □

Theorem 2 is an adaptation of Proposition 1 of Boudjellaba et al. (1992) to the ECCC-GARCH model in the VAR representation for $\epsilon_t^{(2)}$. It sets the conditions for the second-order noncausality between two vectors of variables when in the system there are other auxiliary variables collected in vector y_{3t}.The parametric condition (2.16), however, is unfit for the practical use. This is due to the fact that $\Pi_{21}(L)$ is highly nonlinear function of parameters of the original GARCH(r,s) process (2.8). Moreover, it is a polynomial of infinite order, when $s > 0$. Therefore, evaluation of the matrix polynomial $\Pi(z)$ is further presented in Theorem 3.

Theorem 3. *Let $\epsilon_t^{(2)}$ follow a stationary vector autoregressive moving average process, as in (2.10), partitioned, as in (2.11), which is identifiable and invertible (Assumptions 8–10). Then y_1 does not second-order Granger-cause y_2 given y_3 (denoted by $y_1 \overset{so}{\nrightarrow} y_2|y_3$),*

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

if and only if:

$$\Gamma^{so}_{ij}(z) = \det \begin{bmatrix} \phi^{\cdot j}_{11}(z) & \psi_{11}(z) & \psi_{13}(z) \\ \varphi_{n_1+i,j}(z) & \psi^{i\cdot}_{21}(z) & \psi^{i\cdot}_{23}(z) \\ \phi^{\cdot j}_{31}(z) & \psi_{31}(z) & \psi_{33}(z) \end{bmatrix} = 0 \quad \forall z \in \mathbb{C}, \tag{2.17}$$

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

Proof. In order to prove the simplified conditions for second-order Granger noncausality, (2.17), apply to equation (2.16) from Theorem 2 the matrix transformations of Theorem 3 and then of Theorem 4 of Boudjellaba et al. (1994). □

As was the case for restriction (2.4), condition (2.17) leads to N_1N_2 determinant conditions. Each of them can be represented in a form of polynomial in z of degree $\max(r, s) + (N_1 + N_3)s$: $\Gamma^{so}_{ij}(z) = \sum_{i=1}^{\max(r,s)+(N_1+N_3)s} b_i z^i$, where b_i are nonlinear functions of parameters of the GARCH process. We obtain parameter restrictions for the hypothesis of second-order Granger noncausality by setting $b_i = 0$ for $i = 1, \dots, \max(r, s) + (N_1 + N_3)s$. Such restrictions are ready to be tested.

The innovation of condition (2.17) is that the second-order noncausality from y_{1t} to y_{2t} is analyzed when there are other variables in the system collected in the vector y_{3t}. Such a setting has not been considered so far in the problem of testing the second-order noncausality. The restrictions can even be used for large systems of variables. In the Granger-causality analysis, it is particularly important to consider a sufficiently large set of variables. Sims (1980), on the example of the vector moving average model, shows that the Granger causal relation may appear in the model due to the omitted variables problem. Further, Lütkepohl (1982) shows that because of the omitted variables problem a noncausality relation may arrive. The conclusions of these two papers are maintained for the second-order causality analysis in multivariate GARCH models: one should consider a sufficiently large set of relevant variables in order to avoid the omitted variables bias problem.

Condition (2.17) generalizes results from other studies. Comte and Lieberman (2000) derive similar restriction for the BEKK-GARCH model, with the difference that vector y_t is partitioned only into two sub-vectors y^*_{1t} and y^*_{2t}. Woźniak (2012) does the same for the ECCC-GARCH model. The fact that the vector of variables is partitioned in three and not only two sub-vectors has serious implications for testing Granger-causality relations in conditional variances. Notice that, under such conditions, the formulation of some hypotheses is not even possible. This is because, in general, the fact that $y^*_{1t} \stackrel{so}{\nrightarrow} y^*_{2t}$ (which can be written as $y_{1t} \stackrel{so}{\nrightarrow} (y_{2t}, y_{3t})$) does not imply that $y_{1t} \stackrel{so}{\nrightarrow} y_{2t}|y_{3t}$ or that $y_{1t} \stackrel{so}{\nrightarrow} y_{3t}|y_{2t}$. Moreover, the results of Woźniak (2012) are nested in condition (2.17) by setting $N_3 = 0$.

Two examples illustrate the derivation of the parameter restrictions for several processes that are often used in empirical works.

Example 5. Let y_t be a trivariate GARCH(1,1) process ($N = 3$ and $r = s = 1$). Then, the VARMA process for $\epsilon_t^{(2)}$ is as follows:

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

$$= \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} + \begin{bmatrix} 1-B_{11}L & -B_{12}L & -B_{13}L \\ -B_{21}L & 1-B_{22}L & -B_{23}L \\ -B_{31}L & -B_{32}L & 1-B_{33}L \end{bmatrix} \begin{bmatrix} \nu_{1t} \\ \nu_{2t} \\ \nu_{3t} \end{bmatrix}. \tag{2.18}$$

If one is interested in testing the hypothesis $y_1 \stackrel{so}{\nrightarrow} y_2|y_3$, then by applying Theorem 3 one obtains the following set of restrictions:

$$\mathbf{R}^{III}_1(\theta) = A_{11}(B_{23}B_{31} - B_{21}B_{33}) + A_{31}(B_{13}B_{21} - B_{11}B_{23}) = 0, \tag{2.19a}$$

$$\mathbf{R}^{III}_2(\theta) = A_{11}B_{21} + A_{31}B_{23} = 0, \tag{2.19b}$$

$$\mathbf{R}^{III}_3(\theta) = A_{21} = 0. \tag{2.19c}$$

If one is interested in testing the hypothesis $y_1 \stackrel{so}{\nrightarrow} (y_2, y_3)$, then from Theorem 3 the conditions are given by:

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

which results in the restrictions:

$$\mathbf{R}_1^{IV}(\theta) = A_{11}B_{21} = 0 \quad \text{and} \quad \mathbf{R}_2^{IV}(\theta) = A_{21} = 0, \tag{2.20a}$$

$$\mathbf{R}_3^{IV}(\theta) = A_{11}B_{31} = 0 \quad \text{and} \quad \mathbf{R}_4^{IV}(\theta) = A_{31} = 0. \tag{2.20b}$$

Example 6. Let $\epsilon_t^{(2)}$ follow a $N = 3$ dimensional ARCH(r) process, and let one be interested whether y_1 second-order Granger-causes y_2 (given y_3). The restrictions for this case are:

$$\mathbf{R}^V(\theta) = A_{i.21} = 0 \qquad \text{for } i = 1,\dots,r. \tag{2.21}$$

2.4 Bayesian testing of noncausality in VARMA-GARCH models

In the following section, the problem of testing restrictions imposed on the original parameters of the VARMA-GARCH model is considered. Apart from deriving separate tests for the Granger causality and second-order Granger causality hypotheses, a joint test of the parametric restrictions from conditions (2.4) and (2.17) is proposed. Thus, not only is the role of joint modeling of the transmissions in conditional mean and conditional variance processes emphasized, but also a complete set of tools for the underlying analysis is presented. Moreover, a Bayesian testing procedure is proposed as a solution for some of the drawbacks of classical tests.

The Wald test Consider the classical Wald test of Boudjellaba et al. (1992) for the parameter restrictions for Granger noncausality in the VARMA process. The Wald test has the desirable feature that it requires the estimation of only the

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

most general model. What is not required is the estimation of restricted models. Thus, estimating just one model one can do both: perform the testing procedure, and analyze the parameters responsible for the transmissions. Before a test can be performed, one should first estimate the VARMA model and derive a set of parametric restrictions from condition (2.4). The Wald statistic is given by:

$$\mathbf{W}(\hat{\theta}_m) = T\mathbf{R}(\hat{\theta}_m)'[\mathbf{T}(\hat{\theta}_m)'\mathbf{V}(\hat{\theta}_m)\mathbf{T}(\hat{\theta}_m)]^{-1}\mathbf{R}(\hat{\theta}_m), \tag{2.22}$$

where θ_m is a sub-vector of θ, containing the parameters used in $l_m \times 1$ vector of parametric restrictions $\mathbf{R}(\theta_m)$, $\mathbf{V}(\hat{\theta}_m)$ is the asymptotic covariance matrix of $\sqrt{T}(\hat{\theta}_m - \theta_m)$, and $\mathbf{T}(\hat{\theta}_m)$ is a $m \times l_m$ matrix of partial derivatives of the restrictions with respect to the parameters collected in θ_m:

$$\mathbf{T}(\hat{\theta}_m) = \frac{\partial \mathbf{R}(\theta_m)}{\partial \theta_m}\Big|_{\theta_m = \hat{\theta}_m}. \tag{2.23}$$

Under the null hypothesis of Granger noncausality $\mathbf{W}(\hat{\theta}_m)$ has asymptotic $\chi^2(l_m)$ distribution. However, in equation (2.22) $\mathbf{T}(\theta_m)$ must be of full rank. Otherwise, the asymptotic covariance matrix is singular and the asymptotic distribution is no longer $\chi^2(l_m)$. Boudjellaba et al. (1992), testing the nonlinear restrictions, as in Example (4), show that there are cases when $\mathbf{T}(\theta_m)$ is not of full rank under the null hypothesis. Several works coping with this problem have appeared (Dufour, 1989; Boudjellaba et al., 1992; Lütkepohl and Burda, 1997; Dufour et al., 2006), in the context of testing Granger noncausality for conditional mean processes.

Suppose that a $l_n \times 1$ vector θ_n contains the parameters that appear in the restrictions for second-order Granger noncausality for the multivariate GARCH model derived from condition (2.17). In order to test such restrictions the Wald test can also be used with test statistics $\mathbf{W}(\theta_n)$. Given that $\sqrt{T}(\hat{\theta}_n - \theta_n)$ has asymptotic normal distribution, the test statistic has asymptotic $\chi^2(l_n)$ distribution with l_n degrees of freedom. However, the determinant condition (2.17) results in several nonlinear restrictions on the parameters. The testing of the nonlinear restrictions leads in the problem with the asymptotic distribution of the Wald statistic. The

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

matrix of partial derivatives of the restrictions with respect to the parameters of the model, (2.23), may not be of full rank, and thus the asymptotic covariance matrix of the parametric restrictions under the null hypothesis may be singular. The asymptotic distribution of the test statistics in this case is unknown.

In fact, the Wald test was applied to test the restrictions for the second-order noncausality in the BEKK-GARCH models by Comte and Lieberman (2000) and Hafner and Herwartz (2008). The Wald statistics, proposed by Comte and Lieberman and Hafner and Herwartz, is χ^2-distributed, given the asymptotic normality of the QMLE of the parameters of the BEKK-GARCH model – the result established by Comte and Lieberman (2003). The asymptotic distribution of the test statistic, however, could only be obtained due to the simplifying approach taken. The strategy of Comte and Lieberman (2000) and Hafner and Herwartz (2008) is to derive linear zero restrictions on the original parameters of the model, which are a sufficient condition for the restrictions obtained from the determinant condition (corresponding to determinant condition (2.17) but for the BEKK-GARCH models and with $N_3 = 0$). Among the classical solutions proposed for the problem of testing the Granger noncausality in conditional means, only the modified Wald test of Lütkepohl and Burda (1997) seems applicable for testing second-order noncausality in the GARCH models. Nevertheless, further research of this topic is required.

For the VARMA-ECCC-GARCH models, Ling and McAleer (2003) proved that $\sqrt{T}(\hat{\theta}_n - \theta_n)$ has asymptotic normal distribution. For this model, the application of the Wald test meets the same obstacles as for the BEKK-GARCH model, if one is interested in the testing of the original restrictions for the second-order noncausality and not only those representing the sufficient condition.

Moreover, the asymptotic normality of the QMLE of the parameters for the VARMA-ECCC-GARCH models was derived by Ling and McAleer (2003), under the assumption of the existence of sixth-order moments of y_t. Similar result was obtained by Comte and Lieberman (2003) for the BEKK-GARCH models, under the assumption of the existence of eighth-order moments. For many financial time series analyzed with multivariate GARCH models, these assumptions may not

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

hold, as such data are often leptokurtic and the existence of higher-order moments is uncertain.

Finally, the joint test of Granger noncausality and second-order Granger noncausality is a simple generalization of the two separate tests. Suppose that θ_{m+n} stacks the parameters from restrictions derived from conditions (2.4) and (2.17). The Wald test statistics for such a hypothesis is simply $\mathbf{W}(\theta_{m+n})$ and is asymptotically $\chi^2(l_m + l_n)$ distributed, given that matrix $\mathbf{T}(\theta)$ is of full rank. It also inherits the properties and limitations of both of the separate tests.

Bayesian testing In this study an alternative approach to testing is undertaken following the Lindley type test proposed by Osiewalski and Pipień (2002) and Marzec and Osiewalski (2008). First of all, we propose the method of testing the original restrictions on the parameters for the Granger noncausality and the second-order noncausality presented in Sections 2.2 and 2.3. Secondly, the Bayesian procedure presented in the subsequent part overcomes the limitations of the Wald test. More specifically, singularities of the asymptotic covariance matrix of restrictions are excluded by construction, and the assumptions of the existence of higher-order moments of time series are relaxed.

In the context of Granger causality testing in time series models, Bayesian methods have been used in several works. Woźniak (2012) uses Bayes factors and Posteriors Odds Ratios to infer second-order noncausality between two vectors in GARCH models. Droumaguet and Woźniak (2012) use these tools to make an inference about Granger noncausality in mean and the independence of the hidden Markov process in Markov-switching VARs. Bayesian methods have also been used also in the context of testing exogeneity, a concept related to Granger noncausality . Jarociński and Maćkowiak (2011) use Savage-Dickey Ratios to test block-exogeneity in Bayesian VARs. Finally, Pajor (2011) uses Bayes factors to infer exogeneity in models with latent variables, in particular, in multivariate Stochastic Volatility models.

Consider the following set of hypotheses. The null hypothesis, $\mathcal{H}_0$, states that the $l \times 1$ vector of possibly nonlinear functions of parameters, $\mathbf{R}(\theta)$, is set to a vector

DOI: 10.2870/63858

of zeros. The alternative hypothesis, $\mathcal{H}_1$, states that it is different from a vector of zeros. The considered set of hypotheses is represented by:

$$\mathcal{H}_0 : \mathbf{R}(\theta) = 0,$$
$$\mathcal{H}_1 : \mathbf{R}(\theta) \neq 0.$$

In the context of Granger-causality, the null hypothesis states that y_1 does not cause y_2 (given that there are also other variables in the system collected in y_3). Then the alternative hypothesis states that y_1 causes y_2. The formulation of the hypotheses is general and encompasses Granger noncausality, second-order noncausality and noncausality in variance. In the following part a Bayesian procedure of evaluation of the credibility of the null hypothesis is described.

In the Bayesian approach, a complete model is specified by a prior distribution of the parameters and a likelihood function. The prior distribution, $p(\theta)$, formalizes the knowledge about the parameters that one has before seeing the data, $\mathbf{y}$. The prior beliefs are updated with information from the data that is represented by the likelihood function, $L(\theta; \mathbf{y})$. As a result of the update of the prior beliefs, a posterior distribution of the parameters of the model is obtained. The posterior distribution is proportional to the product of the likelihood function and the prior distribution:

$$p(\theta|\mathbf{y}) \propto L(\theta; \mathbf{y})p(\theta). \tag{2.24}$$

Given the posterior distribution of the parameters, the posterior distribution of the function $\mathbf{R}(\theta)$ is available, $p(\mathbf{R}(\theta)|\mathbf{y})$. Moreover, every characteristic of this distribution is available as well. For instance, the posterior mean of $\mathbf{R}(\theta)$ is calculated by definition of the expected value by integrating the product of the function and its posterior distribution over the whole parameter space:

$$E[\mathbf{R}(\theta)|\mathbf{y}] = \int_{\theta \in \Theta} \mathbf{R}(\theta)p(\theta|\mathbf{y})d\theta.$$

In order to compute such an integral, numerical methods need to be employed for

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

the VARMA-GARCH models, as analytical forms are not known.

Let $\{\theta^{(i)}\}_{i=1}^{S_1}$ be a sample of S_1 draws from the posterior distribution $p(\theta|\mathbf{y})$. Then, $\{\mathbf{R}(\theta^{(i)})\}_{i=1}^{S_1}$ appears a sample drawn from the posterior distribution $p(\mathbf{R}(\theta)|\mathbf{y})$. The posterior mean and the posterior covariance matrix of the restrictions are estimated with:

$$\hat{E}[\mathbf{R}(\theta)|\mathbf{y}] = S_1^{-1}\sum_{i=1}^{S_1}\mathbf{R}(\theta^{(i)}), \tag{2.25}$$

$$\hat{V}[\mathbf{R}(\theta)|\mathbf{y}] = S_1^{-1}\sum_{i=1}^{S_1}\left[\mathbf{R}(\theta^{(i)}) - \hat{E}[\mathbf{R}(\theta)|\mathbf{y}]\right]\left[\mathbf{R}(\theta^{(i)}) - \hat{E}[\mathbf{R}(\theta)|\mathbf{y}]\right]'. \tag{2.26}$$

Define a scalar function $\kappa : \mathbb{R}^l \to \mathbb{R}^+$ by:

$$\kappa(\mathbf{R}) = \left[\mathbf{R} - E[\mathbf{R}(\theta)|\mathbf{y}]\right]' V[\mathbf{R}(\theta)|\mathbf{y}]^{-1}\left[\mathbf{R} - E[\mathbf{R}(\theta)|\mathbf{y}]\right], \tag{2.27}$$

where $\mathbf{R}$ is the argument of the function. In order to distinguish the argument of the function $\mathbf{R} = \mathbf{R}(\theta)$, the simplified notation is used, neglecting the dependence on the vector of parameters. In place of the expected value and the covariance matrix of the vector of restrictions, $E[\mathbf{R}(\theta)|\mathbf{y}]$ and $V[\mathbf{R}(\theta)|\mathbf{y}]$, one should use their estimators, defined in equations (2.25) and (2.26).

The function κ is a positive semidefinite quadratic form of a real-valued vector. It gives a measure of the deviation of the value of the vector of restrictions from its posterior mean, $\mathbf{R} - E[\mathbf{R}(\theta)|\mathbf{y}]$, rescaled by the positive definite posterior covariance matrix, $V[\mathbf{R}(\theta)|\mathbf{y}]$. Notice that the positive definite covariance matrix is a characteristic of the posterior distribution and, by construction, cannot be singular, as long as the restrictions are linearly independent. Drawing an analogy to a Wald test, the main problem of the singularity of the asymptotic covariance matrix of the restrictions is resolved by using the posterior covariance matrix. It does not need to be constructed with the delta method and, thus, avoids the potential singularity of the asymptotic covariance matrix. Notice, however, that the function κ is not a test statistic, but a scalar function that summarizes multiple restrictions on the

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

parameters of the model.

Moreover, if $\mathbf{R}$ follows a normal density function, then $\kappa(\mathbf{R})$ would have a $\chi^2(l)$ distribution with l degrees of freedom (see e.g. Proposition B.3 (2) of Lütkepohl, 2005, pp. 678). Consider testing only the Granger noncausality in mean in the VAR model, when the covariance matrix of the innovations is assumed to be constant over time and known. Then, assuming a normal likelihood function and a normal conjugate prior distribution leads to a normal posterior distribution of the parameters. This finding still does not guarantee the χ^2-distributed κ function, as the restrictions on the parameters of the model might be nonlinear and contain sums of products of the parameters. Further, in the general setting of this study, in which the VARMA-GARCH models with Student's t likelihood function are analyzed, the posterior distribution of the parameters of neither the VARMA nor GARCH parts are in the form of some known distributions (see Bauwens and Lubrano, 1998). Therefore, the exact form of the distribution of $\kappa(\mathbf{R})$ is not known either. It is known up to a normalizing constant, as in equation (2.24). Luckily, using the Monte Carlo Markov Chain methods, the posterior distributions of the parameters of the model, θ, of the restrictions imposed on them, $\mathbf{R}$, as well as of the function $\kappa(\mathbf{R})$, may be easily simulated. The posterior distribution of the function κ is used in order to evaluate the hypothesis of noncausality.

Let $\kappa(\mathbf{0})$ be the value of function κ, evaluated at the vector of zeros, representing the null hypothesis. Then, a negligible part of the posterior probability mass of $\kappa(\mathbf{R})$ attached to the values greater than $\kappa(\mathbf{0})$ is an argument against the null hypothesis. Therefore, the credibility of the null hypothesis can be assessed by computing the posterior probability of the condition $\kappa(\mathbf{R}) > \kappa(\mathbf{0})$:

$$p_0 = Pr\left(\kappa(\mathbf{R}) > \kappa(\mathbf{0})|\mathbf{y}\right) = \int_{\kappa(\mathbf{0})}^{\infty} p(\kappa(\mathbf{R})|\mathbf{y})d\kappa(\mathbf{R}). \tag{2.28}$$

Estimation of the probability, p_0, has to be performed using numerical integration methods. Let $\{\mathbf{R}^{(i)}\}_{i=S_1+1}^{S_2}$ be a sample of S_2 draws from the stationary posterior distribution $p(\mathbf{R}(\theta)|\mathbf{y})$, where $\mathbf{R}^{(i)} = \mathbf{R}(\theta^{(i)})$. Using the transformation κ of the restrictions $\mathbf{R}$, one obtains a sample of S_2 draws, $\{\kappa(\mathbf{R}^{(i)})\}_{i=1}^{S_2}$, from the posterior

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

distribution, $p(\kappa(\mathbf{R})|\mathbf{y})$. Then the probability, p_0, is simply estimated by the fraction of the draws from the posterior distribution of $\kappa(\mathbf{R})$, for which the inequality $\kappa(\mathbf{R}) > \kappa(\mathbf{0})$ holds:

$$\hat{p}_0 = \frac{\#\{\kappa(\mathbf{R^{(i)}}) > \kappa(\mathbf{0})\}}{S_2}. \tag{2.29}$$

The probability, p_0, should be compared to a probability, π_0, that represents a confidence level of the test. The usual values used in many statistical works are 0.05 or 0.1.

The procedure is summarized in five steps:

Step 1 Draw $\{\theta^{(i)}\}_{i=1}^{S_1}$ from the posterior distribution $p(\theta|\mathbf{y})$.

Step 2 Compute $\{\mathbf{R}(\theta^{(i)})\}_{i=1}^{S_1}$, as well as the estimators of the posterior mean, $\hat{E}[\mathbf{R}(\theta)|\mathbf{y}]$, and the posterior covariance matrix, $\hat{V}[\mathbf{R}(\theta)|\mathbf{y}]$, for the vector of restrictions on the parameters.

Step 3 Draw $\{\theta^{(i)}\}_{i=S_1+1}^{S_2}$ from the posterior distribution $p(\theta|\mathbf{y})$, and compute $\{\kappa(\mathbf{R}^{(i)})\}_{i=S_1+1}^{S_2}$ using the estimated posterior mean and covariance matrix from **Step 2** to compute $\kappa(.)$.

Step 4 Compute $\kappa(\mathbf{0})$ and $\hat{p}_0$.

Step 5 If $\hat{p}_0 < \pi_0$, then reject the null hypothesis, $\mathcal{H}_0$. Otherwise, do not reject the null hypothesis.

Osiewalski and Pipień (2002) and Marzec and Osiewalski (2008) use the quadratic form as in (2.27) in order to assess the restrictions imposed on parameters of the analyzed models. Osiewalski and Pipień (2002) proposed the *Lindley-type Bayesian counterpart of the usual F test* to compare a bivariate BEKK-GARCH specification with the vech-GARCH formulation in which BEKK-GARCH is nested. They tested zero restrictions imposed on nonlinear functions of parameters. In order to assess the null hypothesis they use the fact that the posterior distribution of the parameters is approximately asymptotically normal, and therefore the quadratic form

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

is approximately χ^2-distributed. Therefore the null hypothesis is rejected if the probability:

$$Pr(\chi^2 \leq \kappa(\mathbf{0}))$$

is greater than $1-\alpha$, a confidence level of the classical test, where χ^2 is a chi-square random variable with appropriate degrees of freedom parameter. Marzec and Osiewalski (2008) test zero restrictions imposed on the parameters of a stochastic frontier model in order to compare different specifications of the cost function and the distributions for inefficiency effects nested in the estimated general formulation of the model. The null hypothesis is assessed by checking the probability content of the shortest Highest Posterior Density interval that includes value $\kappa(\mathbf{0})$. This procedure practically corresponds to the the one used in the current study.

Discussion The proposed Bayesian procedure allows testing of the noncausality restrictions resulting directly from the determinant condition (2.16). There is no need to derive the simplified zero restrictions on the parameters of the model in order to test the noncausality hypothesis, as proposed by Comte and Lieberman (2000) and Hafner and Herwartz (2008). Second, the procedure requires the estimation of only one unrestricted model for the purpose of testing the noncausality hypotheses. Given the time required to estimate the multivariate VARMA-GARCH models, this is a significant gain in comparison to the procedure proposed by Woźniak (2012). He used Bayes factors to test the second-order noncausality hypotheses between two vectors of variables in ECCC-GARCH models. Consequently, his method requires the estimation of multiple models: the unrestricted and the restricted models representing the hypotheses of interest.

Note, however, that only the inference based on Bayes factors attaches posterior probabilities to hypotheses, and therefore constitutes formal Bayesian assessment of the hypotheses. Lindley type inference based on Highest Posterior Density regions provides an informal summary of the Bayesian evidence for or against hypotheses. Note also that probability p_0 is the probability of the event that the condition $\kappa(\mathbf{R}) > \kappa(\mathbf{0})$ holds, and it can be helpful in the decision-making

 DOI: 10.2870/63858

process. Therefore, the two procedures, Bayes factors and Lindley type tests, are not substitutes, but are rather complimentary. The choice to use the Lindley type test in this study is practical and motivated with the reasons stated above*.

Further, the posterior distribution of function κ is a finite sample distribution. Therefore, the test is also based on the exact finite sample distribution. On the contrary, in the classical inference on VARMA-GARCH models only the asymptotic distribution of the QML estimator of the parameters is available. Since there is no need to refer to asymptotic theory in this study, there is also no need to keep its strict assumptions. As a result, the Bayesian test relaxes the assumptions of the existence of higher-order moments. Only the existence of fourth-order moments is assumed (see Assumption 10), in comparison to the assumption of the existence of sixth-order moments in a classical derivation of the asymptotic distribution of the QMLE (see Ling and McAleer, 2003). Moreover, this testing procedure could be employed for the restrictions of Comte and Lieberman (2000) for testing the noncausality in variance in the BEKK-GARCH models. The asymptotic normality of the QMLE established by Comte and Lieberman (2003) requires the existence of the eighth-order moments, an assumption that can now be relaxed.

These improvements are particularly important in the context of the analysis of financial high-frequency data. Many empirical studies have proved that the empirical distribution of such data is leptokurtic, and that the existence of higher-order moments is questionable. Therefore, the relaxed assumptions may give an advantage on the applicability of the proposed testing procedure over the applicability of classical tests.

2.5 Granger causal analysis of exchange rates

Data In order to illustrate the use of the methods presented in previous sections three time series of daily exchange rates are chosen. The series, all denominated in Euro, are the Swiss franc (CHF/EUR), the British pound (GBP/EUR) and the United

* The author thanks Professor Jacek Osiewalski for pointing the difference between inferences based on Bayes factors and Lindley type tests.

Table 2.1: Data: summary statistics

	CHF/EUR	GBP/EUR	USD/EUR
Mean	-0.034	0.012	-0.006
Median	-0.033	0.011	0.016
Standard Deviation	0.704	0.707	0.819
Minimum	-3.250	-2.657	-4.735
Maximum	7.997	3.461	4.038
Excess kurtosis	25.557	2.430	2.683
Excess kurtosis (robust)	0.785	0.060	0.085
Skewness	2.220	0.344	-0.091
Skewness (robust)	-0.038	0.010	-0.016
LJB test	21784.921	206.525	234.063
LJB p-value	0.000	0.000	0.000
T	777.0	777.0	777.0
Correlations	.	.	.
GBP/EUR	0.079	.	.
USD/EUR	0.301	0.368	.

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

States dollar (USD/EUR). Logarithmic rates of return expressed in percentage points are used, $y_{it} = 100(\ln x_{it} - \ln x_{it-1})$ for $i = 1, 2, 3$, where x_{it} are levels of the assets. The data spans the period from September 16, 2008 to September 22, 2011, which gives $T = 777$ prices. It was downloaded from the European Central Bank website (`http://sdw.ecb.int/browse.do?node=2018794`). The analyzed period starts the day after Lehman Brothers filed for Chapter 11 bankruptcy protection.

The motivation behind this choice of variables and the period of analysis is its usefulness for the institutions for which the forecast of the exchange rates is a crucial element of financial planning. For instance, suppose that the government of a country participating in the Eurozone is indebted in currencies, and therefore its future public debt depends on the exchange rates. Or, suppose that a financial

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

institution settled in the Eurozone keeps assets bought on the New York or London stock exchanges, or simply keeps currencies. In these and many other examples, the performance of an institution depends on the forecast of the returns, but even more important is the forecast of the future volatility of exchange rates. The knowledge that the past information about one exchange rate has an impact on the forecast of the variability of some other exchange rate may be crucial for the analysis of the risk of a portfolio of assets. The two exchange rates, GBP/EUR and USD/EUR, were analyzed for the same period in Woźniak (2012).

Figure B.1 from B.1 plots the three time series. The clustering of the volatility of the data is evident. Two of the exchange rates, GBP/EUR and USD/EUR, during the first year of the sample period were characterized by higher volatility than in the subsequent years. The Swiss franc is characterized by more periods of different volatility. The first year of high variability was followed by nearly a year of low volatility. After that period, again there was a period of high volatility. As the volatility clustering seems to be present in the data, the GARCH models that are capable of modeling this feature are chosen for the subsequent analysis.

Table 2.1 presents the summary statistics of the time series. While the sample means are very similar to one another and close to zero, the variability measured with the sample standard deviation seems to be a bit higher for the US dollar than for other currencies. All the series are leptokurtic, as the kurtosis coefficients are high. This is especially the case for the Swiss franc. The Swiss franc and the British pound are positively skewed. None of the variables follows a normal distribution, as shown by the results of the Lomnicki-Jarque-Bera test.

Estimation of the model The Bayesian estimation of the VARMA-GARCH models consists of the numerical simulation of the posterior distribution of the parameters, which is proportional to the product of the likelihood function and the prior distribution of the parameters of the model, as in equation (2.24).

The parameters of the VAR-GARCH model follow the subsequent prior specification. For the parameters of the vector autoregressive process of order one and of the GARCH(1,1) model, the prior distribution proportional to a constant

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

and constrained to a parameter space bounded according to Assumptions 6–10 is assumed. Each of the parameters of the correlation matrix, **C**, collected in a $N(N-1)/2 \times 1$ vector $\rho = \text{vecl}(\mathbf{C})$, follows a uniform distribution on the interval $[-1, 1]$, where a *vecl* operator stacks lower-diagonal elements of a matrix in a vector. Finally, the degrees of freedom parameter follows the prior distribution proposed by Deschamps (2006). Such a prior specification, with diffuse distributions for all the parameters but the degrees of freedom parameter ν, guarantees the existence of the posterior distribution understood as integrability of the product of the likelihood function and the prior distribution (see Bauwens and Lubrano, 1998). It does not discriminate any of the values of the parameters from within the parameter space. The prior distribution for the parameter ν is a proper density function, and it gives as much as a 32 percent chance that its value is greater than 30. For such values of this parameter, the likelihood function given by equation (2.9b), is a close approximation of the normal likelihood function.

Summarizing, the prior specification for the considered model has the detailed form of:

$$p(\theta) = p\left(\alpha_0', \text{vec}(\alpha_1)\right) p(\omega', \text{vec}(A)', \text{vec}(B)')\, p(\nu) \prod_{i=1}^{N(N-1)/2} p(\rho_i), \tag{2.30}$$

where each prior distribution is specified by:

$$\begin{aligned} p\left(\alpha_0', \text{vec}(\alpha_1)'\right)' &\propto \mathcal{I}(\theta \in \Theta) \\ p(\omega', \text{vec}(A)', \text{vec}(B)')' &\propto \mathcal{I}(\theta \in \Theta) \\ \nu &\sim .04 \exp\left[-.04(\nu - 2)\right] \mathcal{I}(\nu \geq 2) \\ \rho_i &\sim \mathcal{U}(-1, 1) \quad \text{for } i = 1, \ldots, N(N-1)/2, \end{aligned}$$

where $\mathcal{I}(.)$ is an indictor function, taking a value equal to 1 if the condition in brackets holds and 0 otherwise.

The kernel of the posterior distribution of the parameters of the model, given by equation (2.24), is a complicated function of the parameters. It is not given

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

by the kernel of any known distribution function. In consequence, the analytical forms are known neither for the posterior distribution nor for full conditional distributions. Therefore, numerical methods need to be employed in order to simulate the posterior distribution. The Metropolis-Hastings algorithm adapted for the GARCH models by Vrontos et al. (2003) and used in Osiewalski and Pipień (2002, 2004) is used also in this study. At each s^{th} step of the algorithm, a candidate draw, θ^*, is made from the candidate density. The candidate generating density is a multivariate t distribution with the location parameter set to the previous state of the Markov chain, $\theta^{(s-1)}$, the scale matrix $c\Omega$ and the degrees of freedom parameter set to five. The scale matrix, Ω, should be a close approximation of the posterior covariance matrix of the parameters, and a constant c is set in order to obtain the desirable acceptance rate of the candidate draws. A new candidate draw, θ^*, is accepted with the probability:

$$\alpha\left(\theta^{(s-1)}, \theta^*|\mathbf{y}\right) = \min\left[1, \frac{L(\theta^*; \mathbf{y})p(\theta^*)}{L(\theta^{(s-1)}; \mathbf{y})p(\theta^{(s-1)})}\right].$$

Every 100^{th} state of the Markov Chain is kept in the final sample of draws from the posterior distribution of the parameters. The rationale behind this strategy is that, at the cost of decreasing the length of the MCMC, the chain of desirable properties according to several criteria (see Geweke, 1989, 1992; Plummer et al., 2006) is obtained. The summary of the properties of the final sample of draws from the posterior distribution is presented in Table B.1 in B.2.

Estimation results Table 2.2 presents the results of the posterior estimation of the VAR(1)- ECCC-GARCH(1,1) model chosen for the analysis of causality relations in the system of three exchange rates: CHF/EUR, GBP/EUR and USD/EUR. Plots of marginal posterior densities of the parameters are presented in B.2.

Considering posterior means and standard deviations of the parameters of the VAR(1) process, one sees that none of the parameters but $\alpha_{1.13}$ is significantly different from zero. The graphs, however, show that the 90 percent highest posterior density regions of parameters $\alpha_{0.1}$, $\alpha_{1.11}$, $\alpha_{1.22}$ and $\alpha_{1.13}$ do not contain the value zero.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 2.2: Summary of the estimation of the VAR(1)-ECCC-GARCH(1,1) model

	VAR(1)						
	α_0		α_1				
CHF/EUR	-0.022	-0.068	0.003	0.041			
	(0.011)	(0.037)	(0.019)	(0.018)			
GBP/EUR	0.014	-0.016	0.077	-0.019			
	(0.021)	(0.030)	(0.040)	(0.033)			
USD/EUR	0.027	-0.027	0.050	0.006			
	(0.025)	(0.041)	(0.045)	(0.041)			
	GARCH(1,1)						
	ω		A			B	
CHF/EUR	0.001	0.117	0.002	0.002	0.873	0.001	0.001
	(0.001)	(0.029)	(0.002)	(0.002)	(0.030)	(0.001)	(0.001)
GBP/EUR	0.011	0.002	0.062	0.017	0.002	0.808	0.063
	(0.009)	(0.002)	(0.024)	(0.011)	(0.003)	(0.158)	(0.098)
USD/EUR	0.086	0.018	0.117	0.051	0.031	0.787	0.164
	(0.059)	(0.018)	(0.062)	(0.034)	(0.028)	(0.215)	(0.147)
	Degrees of freedom and correlations						
	ν	ρ_{12}	ρ_{13}	ρ_{23}			
	6.267	0.145	0.356	0.400			
	(0.743)	(0.038)	(0.035)	(0.034)			

The table summarises the estimation of the VAR(1)-ECCC-GARCH(1,1) model described by the equations (2.2), (2.7), (2.8) and the likelihood function (2.9b). The prior distributions are specified in equation (2.30). The posterior means and the posterior standard deviations (in brackets) are reported. For graphs of the marginal posterior distributions of the parameters, as well as for the summary of characteristics of the MCMC simulation of the posterior distribution, refer to B.2.

The parameter $\alpha_{1.13}$ is responsible for the interaction of the lagged value for US dollar on the current value of the Swiss frank. This finding has its consequences in testing the Granger causality in mean hypothesis.

All the parameters of the GARCH(1,1) process are constrained to be non-negative. However, a significant part of the posterior probability mass concentrated at the bound given by zero is an argument for a lack of the statistical significance of the parameter. For most of the parameters of the GARCH process reported in Table 2.2, this is the case (see also graphs in B.2). The posterior probability mass of several of the parameters, however, is distant from zero. All the diagonal parameters of matrices A and Bm, beside parameters A_{33} and B_{33}, are different from zero. This finding is common for multivariate GARCH models and reflects the persistence of volatility.

Nevertheless, it is the value of the posterior mean of parameter B_{32} equal to 0.787 that is interesting in this model. This parameter models the impact of the lagged conditional variance of British pound on the current conditional variance of the US dollar. This effect is significant. Moreover, estimates of the parameters for the system of variables that would include only GBP/EUR and USD/EUR are very similar to the values of the parameters of the bivariate VAR-ECCC-GARCH model estimated by Woźniak (2012) for the same period. To conclude, the estimate of this parameter in particular may be considered robust to including an additional variable to the model, namely the CHF/EUR, as well as to the prior distribution specification. Woźniak (2012) estimates two models with truncated-normal priors with two different variance parameters: 100 and 0.1.

Finally, Figure B.7 proves that the parameter of the degrees of freedom, ν, of the t-distributed residuals cannot be considered greater than 6. This value lies in the high posterior probability mass of this parameter. Therefore, the existence of moments of order 6 and higher of the error term is questionable. In effect, classical testing of the VARMA-ECCC-GARCH model has limited use in this case. This statement is justified by the requirement of the existence of sixth-order moments for the asymptotic normality of the QML estimator (see Ling and McAleer, 2003).

Granger-causality testing results Table 2.3 presents the results of the Granger noncausality in mean testing. The values of $\kappa(\mathbf{0})$ and of the estimate of the probability p_0 are reported. B.3 presents plots of the posterior distribution of $\kappa(\mathbf{R})$ for each of the hypotheses.

Only a few of the hypotheses of noncausality in mean are rejected at the confidence levels 0.05 or 0.1. All the rejected hypotheses relate to two of the exchange

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 2.3: Results of testing: Granger causality hypothesis

$\mathcal{H}_0$:	$\kappa(\mathbf{0})$	$\hat{p}_0$	Figure Ref.
$y_1 \overset{G}{\nrightarrow} y_2 \vert y_3$	0.294	0.586	B.9.1
$y_1 \overset{G}{\nrightarrow} y_3 \vert y_2$	0.427	0.522	B.9.2
$y_2 \overset{G}{\nrightarrow} y_1 \vert y_3$	0.022	0.883	B.9.3
$y_2 \overset{G}{\nrightarrow} y_3 \vert y_1$	1.226	0.270	B.9.4
$y_3 \overset{G}{\nrightarrow} y_1 \vert y_2$	5.013	0.023	B.9.5
$y_3 \overset{G}{\nrightarrow} y_2 \vert y_1$	0.336	0.561	B.9.6
$(y_1, y_2) \overset{G}{\nrightarrow} y_3$	1.580	0.455	B.10.1
$(y_1, y_3) \overset{G}{\nrightarrow} y_2$	0.884	0.642	B.10.2
$(y_2, y_3) \overset{G}{\nrightarrow} y_1$	5.520	0.063	B.10.3
$y_1 \overset{G}{\nrightarrow} (y_2, y_3)$	0.530	0.765	B.10.4
$y_2 \overset{G}{\nrightarrow} (y_1, y_3)$	1.252	0.543	B.10.5
$y_3 \overset{G}{\nrightarrow} (y_1, y_2)$	5.776	0.059	B.10.6
$y_1 \overset{G}{\nrightarrow} y_2 \vert y_3$ & $y_2 \overset{G}{\nrightarrow} y_1 \vert y_3$	0.315	0.858	B.11.1
$y_1 \overset{G}{\nrightarrow} y_3 \vert y_2$ & $y_3 \overset{G}{\nrightarrow} y_1 \vert y_2$	5.249	0.072	B.11.2
$y_2 \overset{G}{\nrightarrow} y_3 \vert y_1$ & $y_3 \overset{G}{\nrightarrow} y_2 \vert y_1$	1.402	0.490	B.11.3

Note: The table presents the considered null hypotheses, $\mathcal{H}_0$, of Granger noncausality, as in Definition 2. The values of function κ associated with the null hypotheses, $\kappa(\mathbf{0})$, are reported in the second column. $\hat{p}_0$ is the posterior probability of the condition for not rejecting the null hypothesis, as defined in (2.28). For a graphical presentation of the posterior densities of $\kappa(\mathbf{R})$ and the values $\kappa(\mathbf{0})$, see the figure references given in the last column. The figures may be found in B.3.
Description of the variables: y_1 = CHF/EUR, y_2 = GBP/EUR, y_3 = USD/EUR.

rates: CHF/EUR and USD/EUR. First, the US dollar has a significant effect on the Swiss frank, a result established at the level of confidence equal to 0.05. These two exchange rates impact on each other as well. Further, the US dollar has a significant effect on the Swiss frank and the British pound taken jointly. Finally, the frank is significantly affected by both the pound and the dollar, taken jointly.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 2.4: Results of testing: second-order Granger causality hypothesis

$\mathcal{H}_0$:	$\kappa(\mathbf{0})$	$\hat{p}_0$	Figure Ref.
$y_1 \overset{so}{\nrightarrow} y_2 \vert y_3$	1.521	0.279	B.12.1
$y_1 \overset{so}{\nrightarrow} y_3 \vert y_2$	3.120	0.221	B.12.2
$y_2 \overset{so}{\nrightarrow} y_1 \vert y_3$	2.550	0.247	B.12.3
$y_2 \overset{so}{\nrightarrow} y_3 \vert y_1$	9.811	0.039	B.12.4
$y_3 \overset{so}{\nrightarrow} y_1 \vert y_2$	2.379	0.231	B.12.5
$y_3 \overset{so}{\nrightarrow} y_2 \vert y_1$	2.354	0.193	B.12.6
$(y_1, y_2) \overset{so}{\nrightarrow} y_3$	11.539	0.113	B.13.1
$(y_1, y_3) \overset{so}{\nrightarrow} y_2$	2.733	0.372	B.13.2
$(y_2, y_3) \overset{so}{\nrightarrow} y_1$	3.926	0.386	B.13.3
$y_1 \overset{so}{\nrightarrow} (y_2, y_3)$	3.491	0.246	B.13.4
$y_2 \overset{so}{\nrightarrow} (y_1, y_3)$	10.714	0.061	B.13.5
$y_3 \overset{so}{\nrightarrow} (y_1, y_2)$	4.084	0.227	B.13.6
$y_1 \overset{so}{\nrightarrow} y_2 \vert y_3$ & $y_2 \overset{so}{\nrightarrow} y_1 \vert y_3$	3.481	0.386	B.14.1
$y_1 \overset{so}{\nrightarrow} y_3 \vert y_2$ & $y_3 \overset{so}{\nrightarrow} y_1 \vert y_2$	4.741	0.324	B.14.2
$y_2 \overset{so}{\nrightarrow} y_3 \vert y_1$ & $y_3 \overset{so}{\nrightarrow} y_2 \vert y_1$	11.633	0.099	B.14.3

Note: The table presents the considered null hypotheses of second-order Granger causality, as in Definition 3. For a description of the notation, see the note to Table 2.3.
Description of the variables: y_1 = CHF/EUR, y_2 = GBP/EUR, y_3 = USD/EUR.

All the three last results are established at the level of confidence 0.1.

Consideration of the results of testing the second-order noncausality hypotheses, reported in Table 2.4, brings new findings. The pattern of connections between the exchange rates is different for second-order causality than for causality in mean. The rejected hypotheses of second-order noncausality relate to the British pound and the US dollar. Information about the history of volatility of GBP/EUR has a significant effect on the current conditional variance of variable USD/EUR at the level of confidence equal to 0.05. It has also a significant effect on CHF/EUR and USD/EUR, taken jointly at the level of confidence equal to 0.1. The same

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 2.5: Results of testing: Granger causality in variance hypothesis

$\mathcal{H}_0$:	$\kappa(\mathbf{0})$	$\hat{p}_0$	Figure Ref.
$y_1 \overset{V}{\nrightarrow} y_2 \vert y_3$	1.755	0.502	B.15.1
$y_1 \overset{V}{\nrightarrow} y_3 \vert y_2$	3.791	0.297	B.15.2
$y_2 \overset{V}{\nrightarrow} y_1 \vert y_3$	2.591	0.430	B.15.3
$y_2 \overset{V}{\nrightarrow} y_3 \vert y_1$	11.177	0.040	B.15.4
$y_3 \overset{V}{\nrightarrow} y_1 \vert y_2$	8.007	0.095	B.15.5
$y_3 \overset{V}{\nrightarrow} y_2 \vert y_1$	2.795	0.324	B.15.6
$(y_1, y_2) \overset{V}{\nrightarrow} y_3$	13.412	0.128	B.16.1
$(y_1, y_3) \overset{V}{\nrightarrow} y_2$	3.728	0.545	B.16.2
$(y_2, y_3) \overset{V}{\nrightarrow} y_1$	10.350	0.205	B.16.3
$y_1 \overset{V}{\nrightarrow} (y_2, y_3)$	4.229	0.440	B.16.4
$y_2 \overset{V}{\nrightarrow} (y_1, y_3)$	12.098	0.083	B.16.5
$y_3 \overset{V}{\nrightarrow} (y_1, y_2)$	10.599	0.118	B.16.6
$y_1 \overset{V}{\nrightarrow} y_2 \vert y_3$ & $y_2 \overset{V}{\nrightarrow} y_1 \vert y_3$	3.826	0.629	B.17.1
$y_1 \overset{V}{\nrightarrow} y_3 \vert y_2$ & $y_3 \overset{V}{\nrightarrow} y_1 \vert y_2$	10.900	0.184	B.17.2
$y_2 \overset{V}{\nrightarrow} y_3 \vert y_1$ & $y_3 \overset{V}{\nrightarrow} y_2 \vert y_1$	13.310	0.119	B.17.3

Note: The table presents the considered null hypotheses of Granger causality in variance, as in Definition 4. For a description of the notation, see the note to Table 2.3.
Description of the variables: y_1 = CHF/EUR, y_2 = GBP/EUR, y_3 = USD/EUR.

conclusions are found in Woźniak (2012). This finding is particularly interesting, as Woźniak uses Bayes factors and Posterior probabilities in order to assess the hypotheses. These conclusions are, therefore, robust to the choice of the testing procedure.

The following interpretation of the testing results of second-order noncausality hypothesis is proposed. The Swiss frank does not have any significant effect on the volatility of the British pound or the US dollar, which proves its minor role in modeling volatility in comparison to the other two exchange rates. The impact of the pound-to-Euro exchange rate on the volatility of the dollar-to-Euro exchange rate, is most probably related to the the *meteor showers* hypothesis of Engle et al. (1990). The proper conclusion seems to be that the spillovers in volatility are due to the activity of traders on the exchange rates market. Although the market is open 24 hours a day, there exist periods of higher activity of trading of particular currencies. Therefore, the behavior of traders in Europe, reflected in the exchange rate prices and their volatility, affects the decisions of traders in North America. Such a pattern can be captured by the dataset and the model considered in this study.

One more hypothesis is rejected at the confidence level equal to 0.1: the pound is found to second-order cause the dollar, *and* the dollar second-order causes pound, which is mainly driven by parameter B_{32}.

Finally, the results of testing hypotheses of noncausality in variance are reported in Table 2.5. These results are not just a simple intersection of the results for Granger-causality in mean and second-order noncausality testing, as one could deduce from equation (2.15). The parameters of the VAR process are not independent of the parameters of the GARCH process. The posterior covariance matrix is not block-diagonal. Therefore, the results of noncausality in variance should be discussed separately. One of the hypotheses is rejected at the confidence level equal to 0.05: the hypothesis of noncausality in variance from the British pound to the US dollar. The other three hypotheses are rejected at the confidence level equal to 0.1. The following relations are found: the dollar causes the frank in variance; and the pound causes the frank and the dollar in variance, taken jointly.

2.6 Conclusions

This study first of all proposes the parameter restrictions for second-order noncausality between two vectors of variables, when there are also other variables in the considered system used for modeling and forecasting. The derivations are made within the framework of the popular VARMA-GARCH model. The novelty

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

of these conditions is that, contrary to the developments of Comte and Lieberman (2000) and Woźniak (2012), they allow the finding of restrictions for a hypothesis of noncausality between chosen variables from the system. The two cited works use a setting in which all the variables are split into two vectors, which imposes a kind of a rigidity in forming hypotheses.

The conditions may result in several nonlinear restrictions on the parameters of the model, which results in a conclusion that the available classical tests have limited use. As a solution to this testing problem, the Bayesian procedure based on the posterior distribution of a function summarizing all the restrictions is proposed. This procedure allows for testing of the hypotheses of Granger noncausality in mean and second-order noncausality jointly, forming a hypothesis of noncausality in variance as well as separately. The procedure requires the estimation of only one model, the unrestricted. Note that the procedure proposed by Woźniak (2012), based on Bayes factors, required the estimation of several models representing different hypotheses. Further, the restrictions of the existence of the higher-order moments of the processes required in the classical tests are relaxed. Similarly to the test of Woźniak (2012), the existence of fourth-order moments is required in the proposed analysis, whereas the asymptotic derivations of Ling and McAleer (2003) require the existence of the sixth-order moments for the VARMA-GARCH models.

The main limitation of the noncausality analysis in this work, is that the conditions only for one-period-ahead noncausality are presented. In the works of Comte and Lieberman (2000) and Woźniak (2012), due to the specific setting of the vectors of variables from the system, these conditions imply noncausality at all the future horizons. In this work, however, when the third vector of variables, y_3, is non-empty, then the conditions from Theorem 3 are useful only for the analysis one period ahead.

This limitation forms a motivation for future research that would aim at derivation of the restrictions for h-period-ahead noncausality within the flexible framework of splitting the variables into three vectors, and where $h = 1, 2, 3, \dots$. Such conditions would be informative of the non-direct causality that is, a situation in

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

which, despite the fact that one variable does not Granger-cause the other one period ahead, it may still be causal several periods ahead through the channel of the third variable (see Dufour et al., 2006).

Another direction of possible research is a derivation of the conditions for second-order noncausality for GARCH models, when the data have specific features. Some financial data are proven to have persistent volatility that is modeled with integrated GARCH processes. Such processes are defined by the fact that the polynomial $|I_N - A(z) - B(z)| = 0$ has a unit root. This case is excluded from the analysis in this study. Further, the analysis of some financial time series conducted by Diebold and Yilmaz (2009), has proved that the values of financial assets as well as their volatility spillover at different rates in different periods. This finding might result in the parameters of the GARCH process changing values over time. Such nonlinearities may be modeled, e.g. with the GARCH processes with a regime change, or when the parameters change their values according to a latent hidden Markov process, as in the Markov-switching models. For such data, the analysis of Granger causality is of interest as well.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute
DOI: 10.2870/63858

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

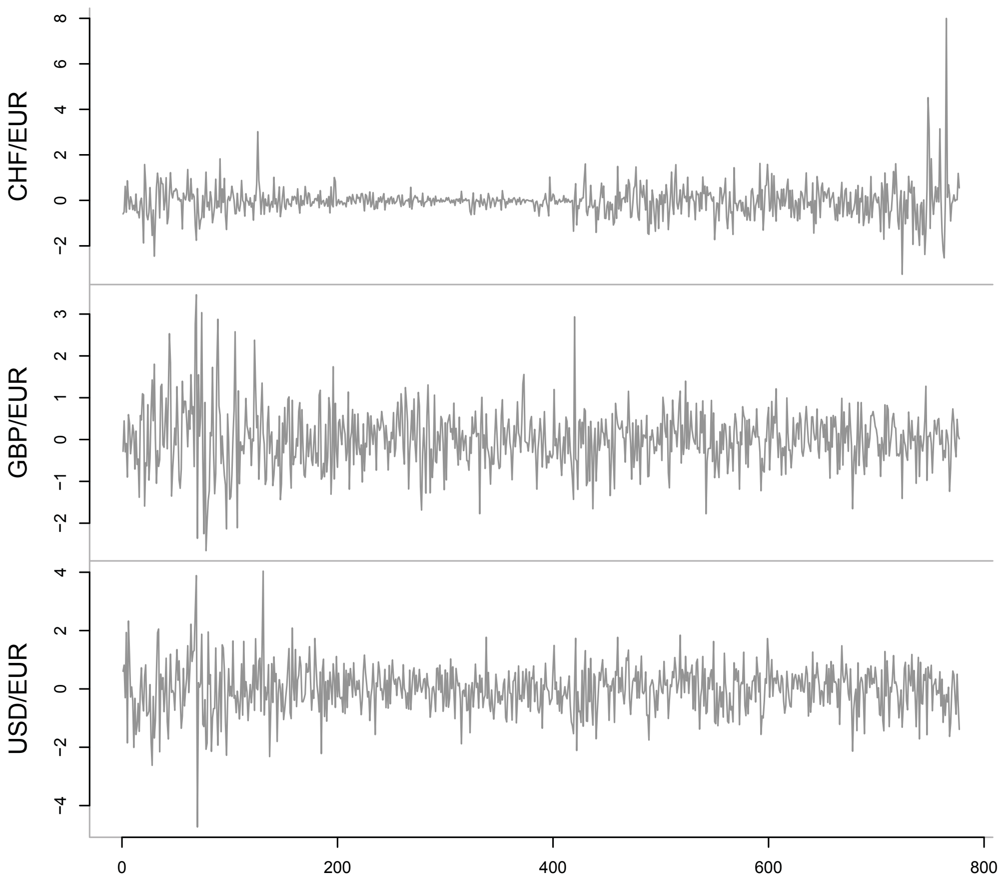
Appendix B

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

B.1 Data

Figure B.1: Data plot: (CHF/EUR, GBP/EUR, USD/EUR)



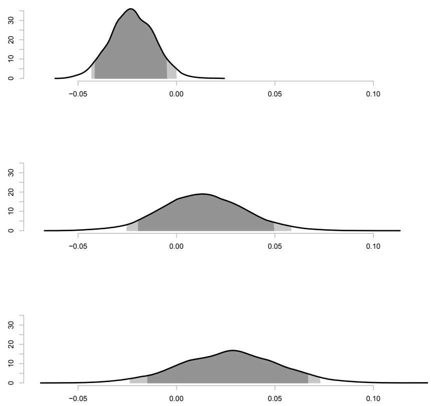
The graph presents daily logarithmic rates of return expressed in percentage points: $y_{it} = 100(\ln x_{it} - \ln x_{it-1})$, for $i = 1, 2, 3$, where x_{it} denotes the level of an asset of three exchange rates: the Swiss franc, the British pound and the US dollar, all denominated in Euro. The data spans the period from September 16, 2008 to September 22, 2011, which gives $T = 777$ observations. The data was downloaded from the European Central Bank website (http://sdw.ecb.int/browse.do?node=2018794).

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute
DOI: 10.2870/63858

B.2 Summary of the posterior density simulation

Note: Figures B.2–B.8 present the marginal posterior distribution of the parameters with the 95% and 90% highest posterior density regions represented by light-grey and dark-grey areas respectively.

Figure B.2: Summary of the posterior distribution: α_0



Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Figure B.3: Summary of the posterior distribution: α_1

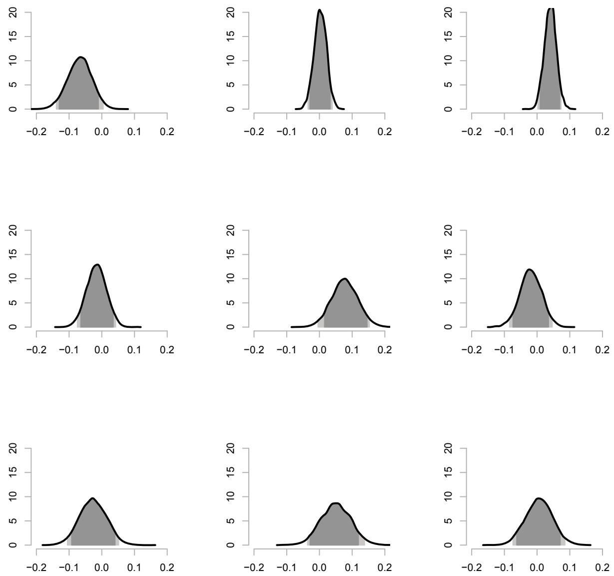
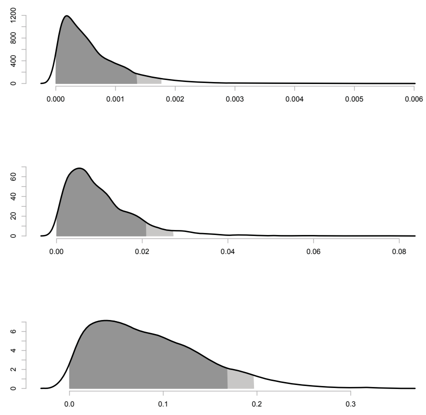
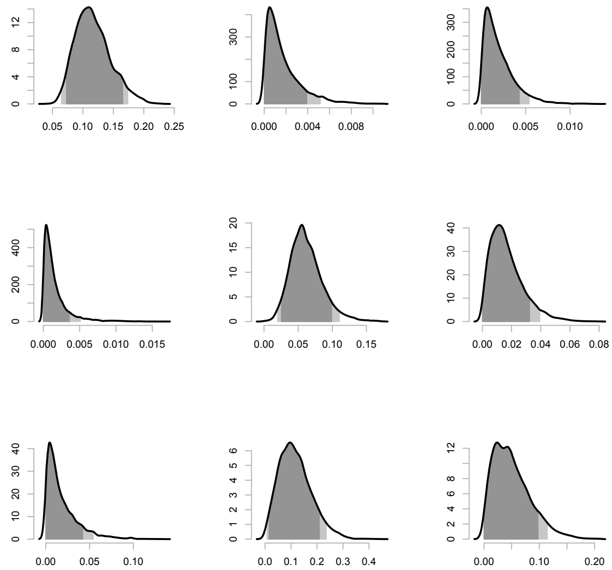


Figure B.4: Summary of the posterior distribution: ω



Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Figure B.5: Summary of the posterior distribution: A



Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Figure B.6: Summary of the posterior distribution: B

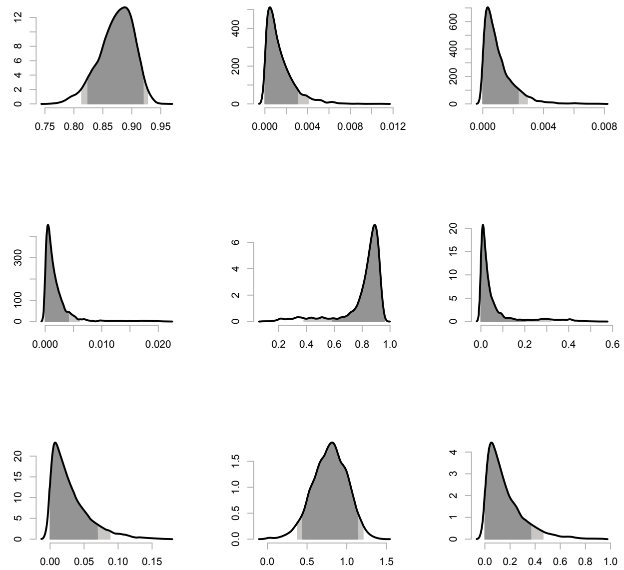
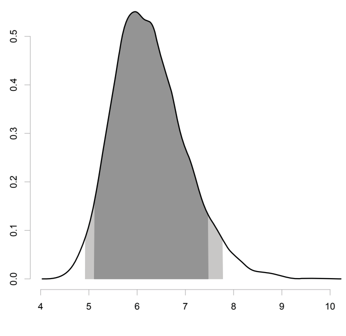
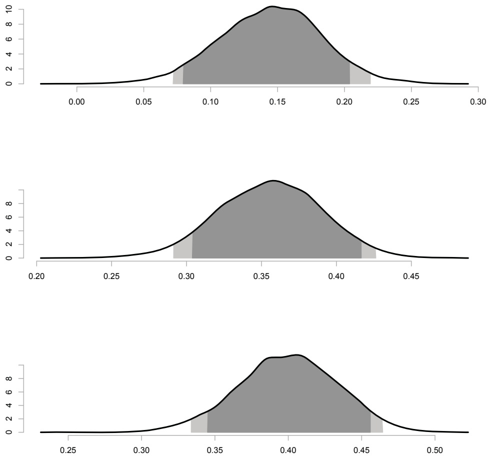


Figure B.7: Summary of the posterior distribution: ν



Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Figure B.8: Summary of the posterior distribution: ρ



Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table B.1: Summary of the posterior distribution simulation

		Mean	SD	Autocorrelations at lag 1	Autocorrelations at lag 50	RNE	Geweke's z
	Vector Autoregression						
α_0	$\alpha_{0.1}$	-0.022	0.011	0.316	-0.012	0.554	-0.475
	$\alpha_{0.2}$	0.014	0.021	0.341	-0.017	0.436	-0.089
	$\alpha_{0.3}$	0.027	0.025	0.335	-0.024	0.355	1.701
α_1	$\alpha_{1.11}$	-0.068	0.037	0.338	0.001	0.429	0.568
	$\alpha_{1.21}$	-0.016	0.030	0.403	0.015	0.303	0.143
	$\alpha_{1.31}$	-0.027	0.041	0.378	-0.007	0.315	-0.658
	$\alpha_{1.12}$	0.003	0.019	0.290	0.002	0.322	1.350
	$\alpha_{1.22}$	0.077	0.040	0.330	-0.012	0.435	0.166
	$\alpha_{1.32}$	0.050	0.045	0.233	0.022	0.511	-0.982
	$\alpha_{1.13}$	0.041	0.018	0.338	-0.008	0.415	-0.467
	$\alpha_{1.23}$	-0.019	0.033	0.371	-0.020	0.284	-2.032
	$\alpha_{1.33}$	0.006	0.041	0.347	0.013	0.404	0.471
	GARCH(1,1)						
ω	ω_1	0.001	0.001	0.677	0.032	0.108	-0.773
	ω_2	0.011	0.009	0.600	0.023	0.082	-1.154
	ω_3	0.086	0.059	0.392	0.058	0.145	1.818
A	A_{11}	0.117	0.029	0.460	-0.019	0.186	0.176
	A_{21}	0.002	0.002	0.639	0.066	0.058	-1.143
	A_{31}	0.018	0.018	0.581	0.042	0.109	0.987
	A_{12}	0.002	0.002	0.630	-0.055	0.085	-0.364
	A_{22}	0.062	0.024	0.575	0.005	0.134	0.606
	A_{32}	0.117	0.062	0.258	-0.001	0.468	-0.074
	A_{13}	0.002	0.002	0.533	-0.054	0.173	0.609
	A_{23}	0.017	0.011	0.483	0.058	0.121	-1.442
	A_{33}	0.051	0.034	0.425	-0.004	0.226	1.117
B	B_{11}	0.873	0.030	0.480	-0.023	0.171	-0.021
	B_{21}	0.002	0.003	0.878	0.317	0.020	-0.731
	B_{31}	0.031	0.028	0.207	0.031	0.141	1.718
	B_{12}	0.001	0.001	0.619	-0.039	0.171	-0.354
	B_{22}	0.808	0.158	0.956	0.569	0.015	2.401
	B_{32}	0.787	0.215	0.518	-0.006	0.122	-0.438
	B_{13}	0.001	0.001	0.575	-0.034	0.277	2.530
	B_{23}	0.063	0.098	0.968	0.585	0.015	-2.357
	B_{33}	0.164	0.147	0.659	0.056	0.064	-0.885
	Degrees of freedom and correlations						
	ν	6.267	0.743	0.415	0.017	0.447	0.123
C	ρ_{12}	0.145	0.038	0.335	0.036	0.481	1.444
	ρ_{13}	0.356	0.035	0.359	-0.021	0.481	2.820
	ρ_{23}	0.400	0.034	0.328	-0.006	0.501	0.048

Note: The table reports posterior means and posterior standard deviations of the parameters of the model. Also, autocorrelations at lag 1 and 50 are given. The relative numerical efficiency coefficient (RNE) was introduced by Geweke (1989). Geweke's z scores test the stationarity of the draws from the posterior distribution, comparing the mean of the first 50% of the draws to the mean of the last 35% of the draws. z scores follow the standard normal distribution (see Geweke, 1992). The numbers presented in this table were computed using the package coda by Plummer et al. (2006).

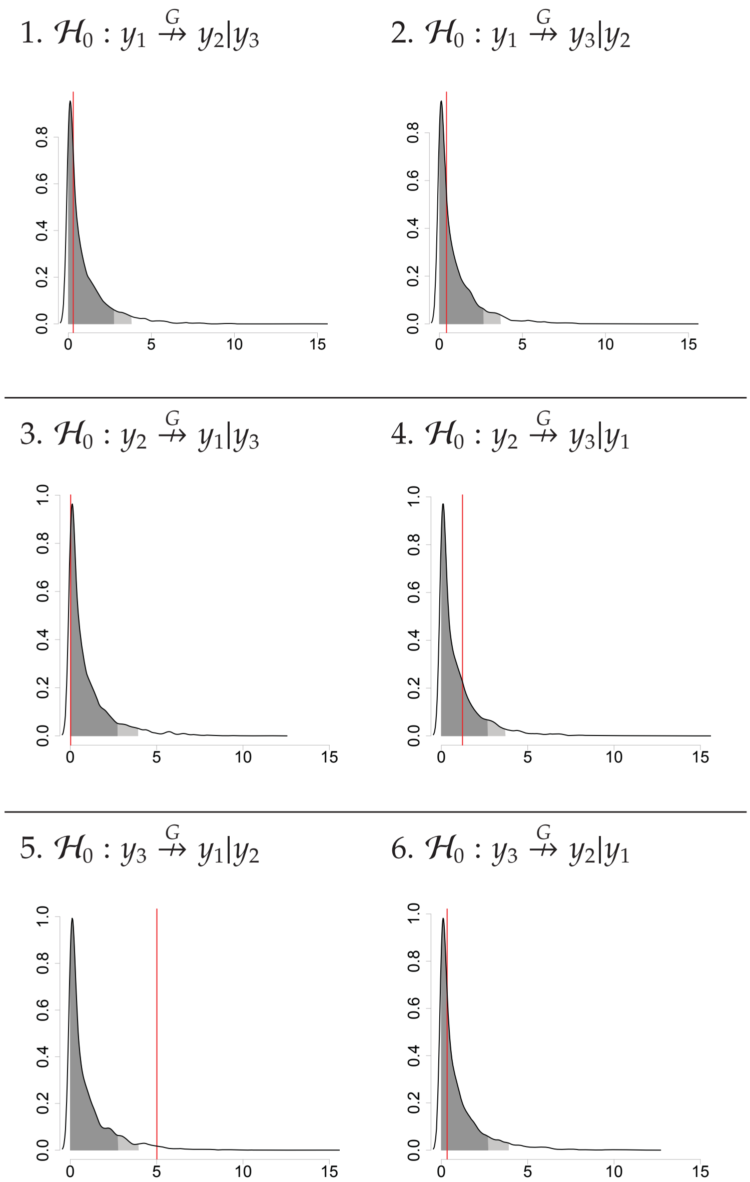
Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

B.3 Graphs summarising testing of the noncausality hypotheses

Graphs B.9–B.17 present simulated posterior distributions of function κ for different hypotheses. The shaded areas denote the 95% and 90% highest posterior density regions of the distributions. For more detailed results, refer to Tables 2.3–2.5. Description of the variables: y_1 = CHF/EUR, y_2 = GBP/EUR, y_3 = USD/EUR.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

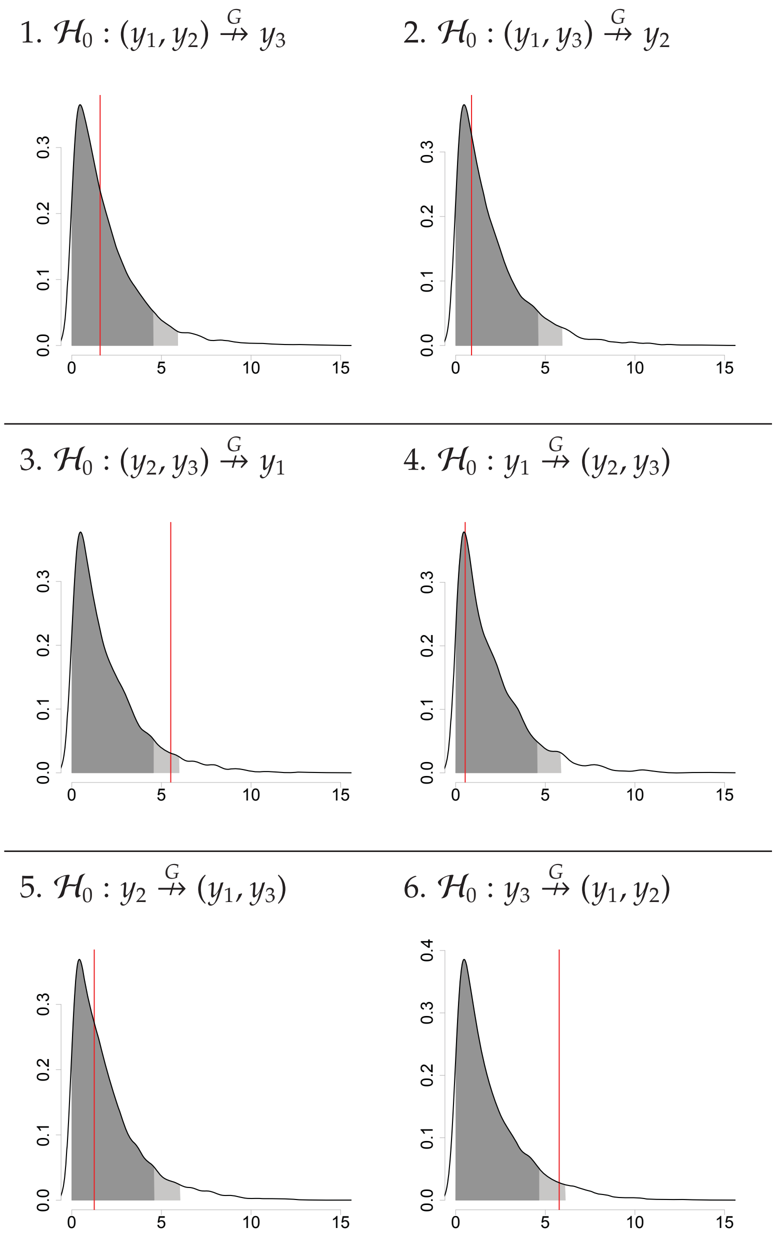
Figure B.9: Results of testing: Granger causality hypotheses I



Visual representation of results from Table 2.3.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Figure B.10: Results of testing: Granger causality hypotheses II

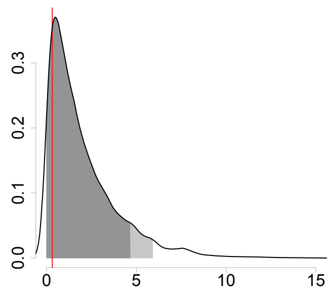


Visual representation of results from Table 2.3.

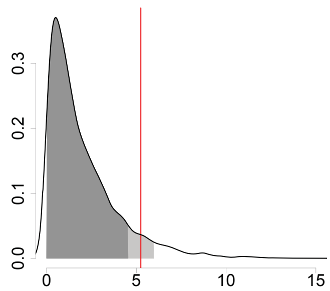
Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Figure B.11: Results of testing: Granger causality hypotheses III

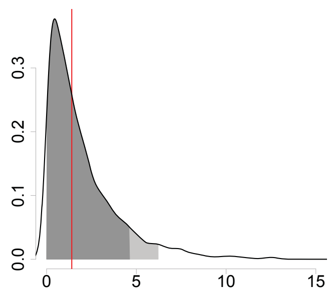
1. $\mathcal{H}_0 : y_1 \overset{G}{\nrightarrow} y_2|y_3$ and $y_2 \overset{G}{\nrightarrow} y_1|y_3$



2. $\mathcal{H}_0 : y_1 \overset{G}{\nrightarrow} y_3|y_2$ and $y_3 \overset{G}{\nrightarrow} y_1|y_2$



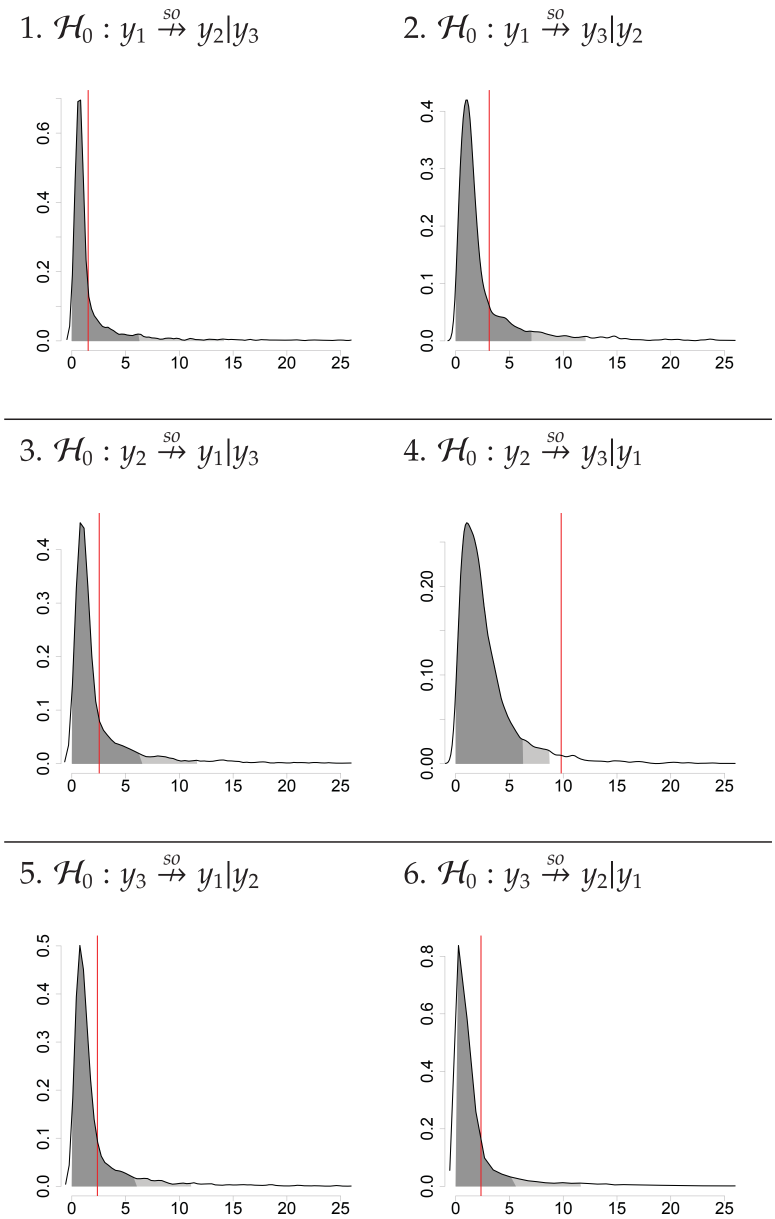
3. $\mathcal{H}_0 : y_2 \overset{G}{\nrightarrow} y_3|y_1$ and $y_3 \overset{G}{\nrightarrow} y_2|y_1$



Visual representation of results from Table 2.3.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

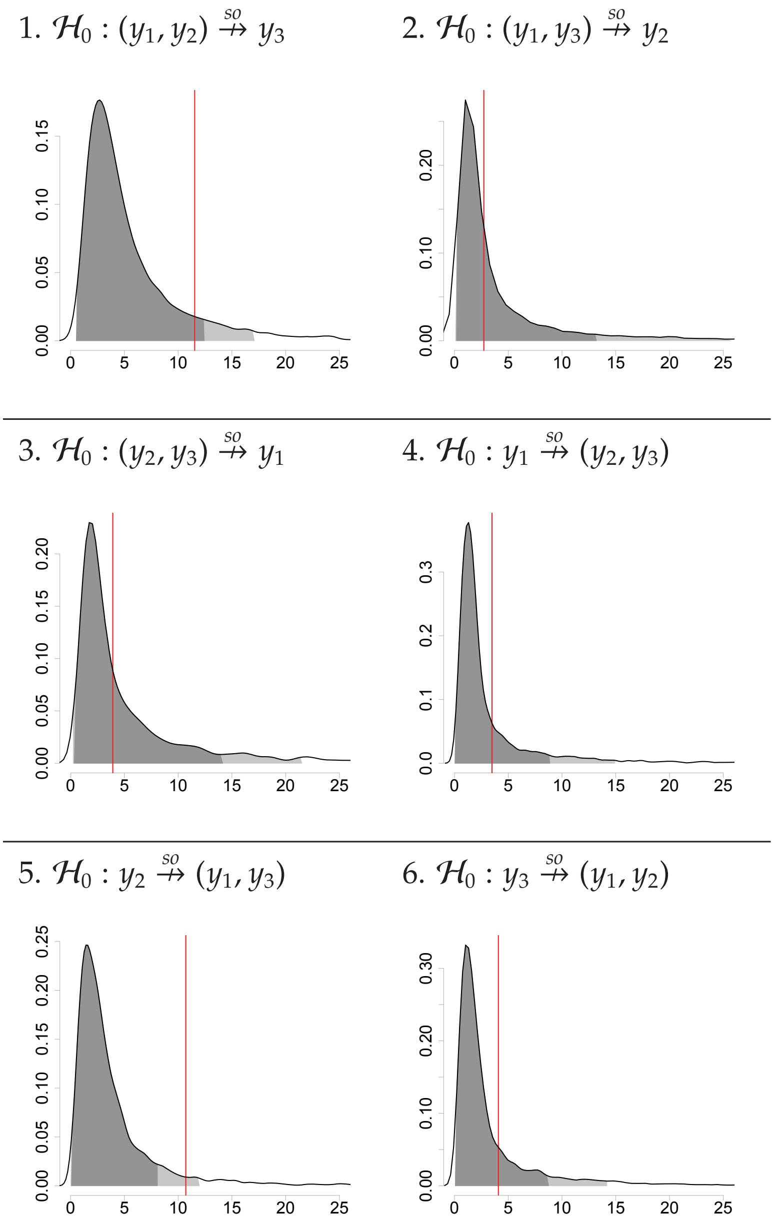
Figure B.12: Results of testing: second-order Granger causality hypotheses I



Visual representation of results from Table 2.4.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute
DOI: 10.2870/63858

Figure B.13: Results of testing: second-order Granger causality hypotheses II

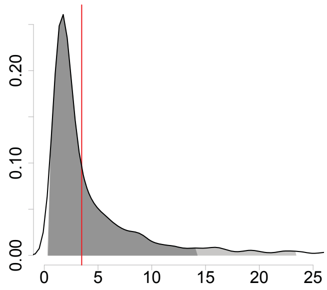


Visual representation of results from Table 2.4.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

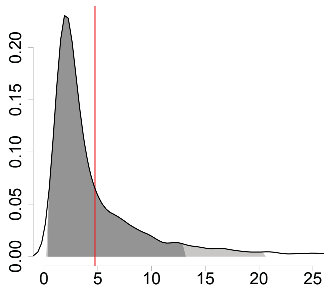
Figure B.14: Results of testing: second-order Granger causality hypotheses III

1. $\mathcal{H}_0 : y_1 \overset{so}{\nrightarrow} y_2|y_3$ and $y_2 \overset{so}{\nrightarrow} y_1|y_3$



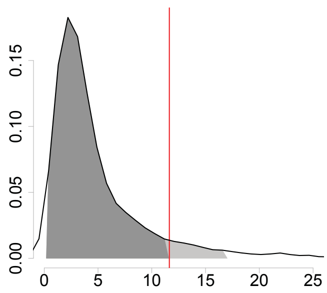
---

2. $\mathcal{H}_0 : y_1 \overset{so}{\nrightarrow} y_3|y_2$ and $y_3 \overset{so}{\nrightarrow} y_1|y_2$



---

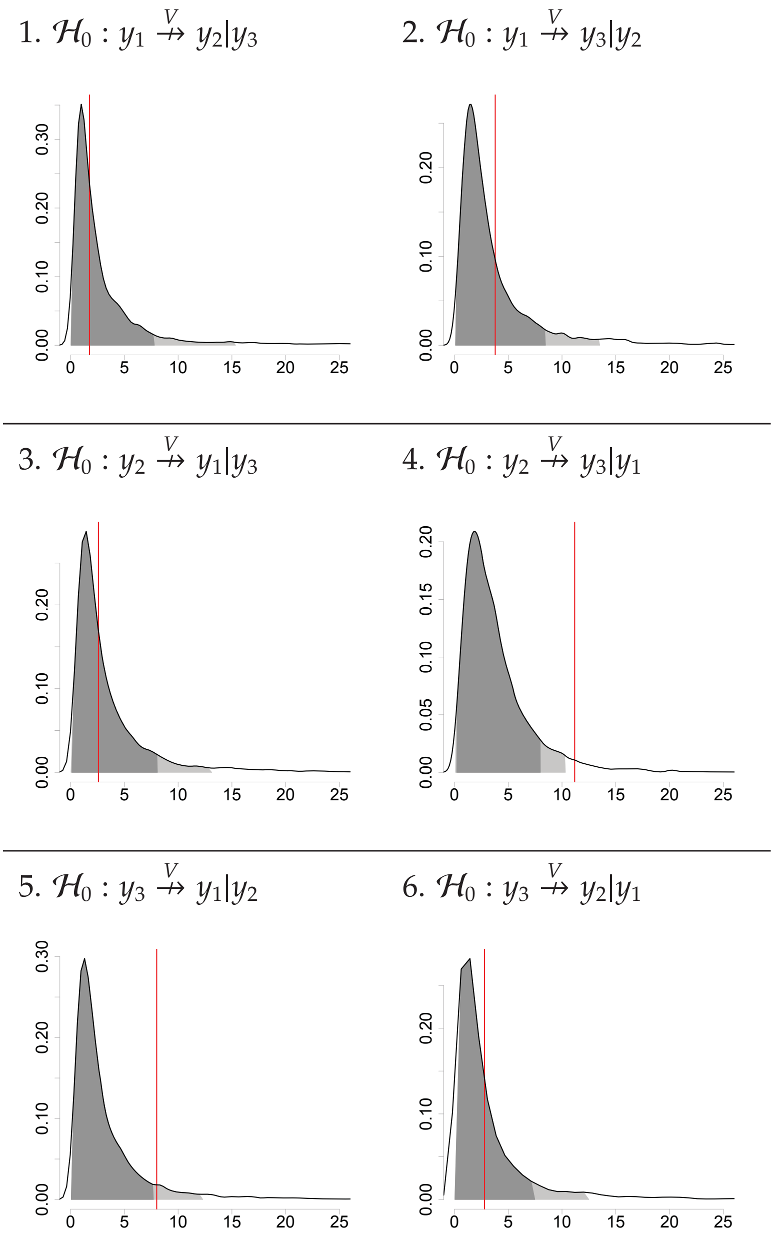
3. $\mathcal{H}_0 : y_2 \overset{so}{\nrightarrow} y_3|y_1$ and $y_3 \overset{so}{\nrightarrow} y_2|y_1$



Visual representation of results from Table 2.4.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute
DOI: 10.2870/63858

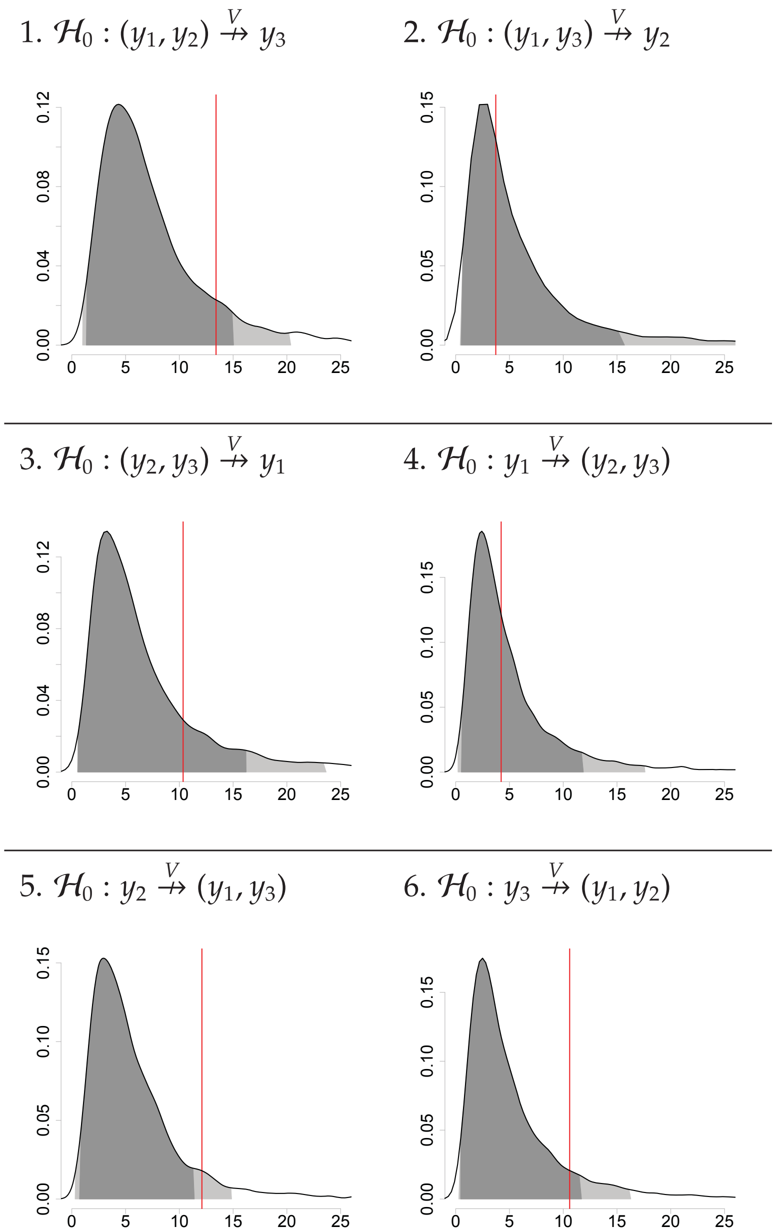
Figure B.15: Results of testing: Granger causality in variance hypotheses I



Visual representation of results from Table 2.5.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute
DOI: 10.2870/63858

Figure B.16: Results of testing: Granger causality in variance hypotheses II

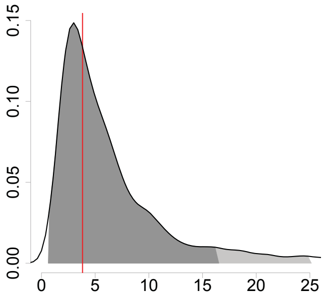


Visual representation of results from Table 2.5.

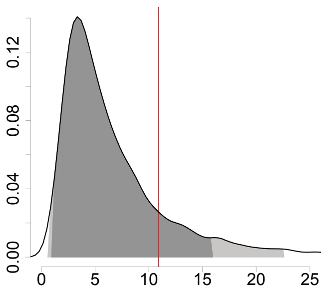
Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute
DOI: 10.2870/63858

Figure B.17: Results of testing: Granger causality in variance hypotheses III

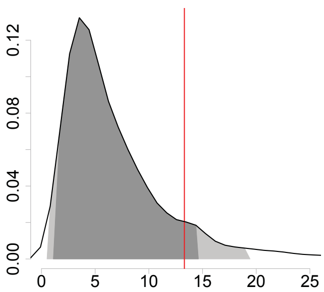
1. $\mathcal{H}_0 : y_1 \overset{V}{\nrightarrow} y_2|y_3$ and $y_2 \overset{V}{\nrightarrow} y_1|y_3$



2. $\mathcal{H}_0 : y_1 \overset{V}{\nrightarrow} y_3|y_2$ and $y_3 \overset{V}{\nrightarrow} y_1|y_2$



3. $\mathcal{H}_0 : y_2 \overset{V}{\nrightarrow} y_3|y_1$ and $y_3 \overset{V}{\nrightarrow} y_2|y_1$



Visual representation of results from Table 2.5.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Bibliography

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

Berndt, E. K., B. H. Hall, R. E. Hall, and A. Hausman (1974). Estimation and Inference in Nonlinear Structural Models. *Annals of Economic and Social Measurement 3*(4), 653–665.

Bollerslev, T. (1990, August). Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model. *The Review of Economics and Statistics 72*(3), 498–505.

Boudjellaba, H., J.-M. Dufour, and R. Roy (1992). Testing Causality Between Two Vectors in Multivariate Autoregressive Moving Average Models. *Journal of the American Statistical Association 87*(420), 1082– 1090.

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

Caporale, G. M., N. Pittis, and N. Spagnolo (2006, September). Volatility transmission and financial crises. *Journal of Economics and Finance 30*(3), 376–390.

Chib, S. and E. Greenberg (1995, November). Understanding the Metropolis-Hastings Algorithm. *The American Statistician 49*(4), 327.

Comte, F. and O. Lieberman (2000). Second-Order Noncausality in Multivariate GARCH Processes. *Journal of Time Series Analysis 21*(5), 535–557.

Comte, F. and O. Lieberman (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis 84*(1), 61–84.

Conrad, C. and M. Karanasos (2009). Negative Volatility Spillovers in the Unrestricted Eccc-Garch Model. *Econometric Theory*, 1–25.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Deschamps, P. J. (2006). A flexible prior distribution for Markov switching autoregressions with Student-t errors. *Journal of Econometrics 133*(1), 153–190.

Diebold, F. X. and K. Yilmaz (2009). Measuring Financial Asset Return and Volatility Spillovers, with Application to Global Equity Markets. *The Economic Journal 119*(534), 158–171.

Droumaguet, M. and T. Woźniak (2012). Bayesian Testing of Granger Causality in Markov-Switching VARs. Working paper series, European University Institute, Florence, Italy. Download at: http://cadmus.eui.eu/bitstream/handle/1814/20815/ECO_2012_06.pdf?sequence=1.

Dufour, J.-M. (1989). Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions. *Econometrica 54*(2), 335–355.

Dufour, J.-M., D. Pelletier, and E. Renault (2006, June). Short run and long run causality in time series: inference. *Journal of Econometrics 132*(2), 337–362.

Engle, R. F., T. Ito, and W.-l. Lin (1990). Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market. *Econometrica 58*(3), 525–542.

Engle, R. F. and K. F. Kroner (1995). Multivariate Simultaneous Generalized ARCH. *Econometric Theory 11*(1), 122–150.

Fiorentini, G., E. Sentana, and G. Calzolari (2003). Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroscedastic Dynamic Regression Models With Student t Innovations. *Journal of Business and Economic Statistics 21*(4), 532–546.

Florens, J. P. and M. Mouchart (1985). A Linear Theory for Noncausality. *Econometrica 53*(1), 157–176.

Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica 57*(6), 1317–1339.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

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, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, Volume 148. Clarendon Press, Oxford.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica 37*(3), 424–438.

Hafner, C. M. (2009, March). Causality and forecasting in temporally aggregated multivariate GARCH processes. *Econometrics Journal 12*(1), 127–146.

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

He, C. and T. Teräsvirta (2004). An Extended Constant Conditional Correlation Garch Model and Its Fourth-Moment Structure. *Econometric Theory 20*, 904–926.

Hong, Y. (2001, July). A test for volatility spillover with application to exchange rates. *Journal of Econometrics 103*, 183–224.

Jarociński, M. and B. Maćkowiak (2011). Choice of Variables in Vector Autoregressions.

Jarque, C. M. and A. K. Bera (1980). Efficient Tests for Normality, Homoscedasticity and Serial Indipendence of Regression Residuals. *Economics Letters 6*, 255–259.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate arch models. *Econometric Theory 14*(01), 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 & Economic Statistics 13*(1), 11–25.

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

Koutmos, G. and G. Booth (1995, December). Asymmetric volatility transmission in international stock markets. *Journal of International Money and Finance 14*(6), 747–762.

Lin, W.-L., R. F. Engle, and T. Ito (1994). Do Bulls and Bears Move across Borders? International Transmission of Stock Returns and Volatility. *The Review of Financial Studies 7*, 507–538.

Ling, S. and M. McAleer (2003). Asymptotic Theory for a Vector {Arma-Garch} Model. *Econometric Theory 19*(02), 280–310.

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

Lütkepohl, H. (1982). Non-causality due to omitted variables. *Journal of Econometrics 19*(2-3), 367–378.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

Lütkepohl, H. and M. M. Burda (1997, June). Modified Wald tests under nonregular conditions. *Journal of Econometrics 78*(1), 315–332.

Marzec, J. and J. Osiewalski (2008). Bayesian Inference on Technology and Cost Efficiency. *Bank i Kredyt 39*(9), 1–32.

Nakatani, T. and T. Teräsvirta (2008, June). Positivity constraints on the conditional variances in the family of conditional correlation GARCH models. *Finance Research Letters 5*, 88–95.

Nakatani, T. and T. Teräsvirta (2009). Testing for volatility interactions in the Constant Conditional Correlation GARCH model. *Econometrics Journal 12*(1), 147–163.

Omrane, W. B. and C. M. Hafner (2009). Information Spillover, Volatility and the Currency Markets. *International Econometric Review 1*(1), 47–59.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Osiewalski, J. and M. Pipień (2002). Multuvariate t-GARCH Models - Bayesian Analysis for Exchange Rates. In *Modelling Economies in Transition - Proceedings of the Sixth AMFET Conference*, Łódź, pp. 151–167. Absolwent.

Osiewalski, J. and M. Pipień (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*(2), 49–73.

Plummer, M., N. Best, K. Cowles, and K. Vines (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News 6*(1), 7–11.

Robins, R. P., C. W. J. Granger, and R. F. Engle (1986). *Wholesale and Retail Prices: Bivariate Time-Series Modeling with forecastable Error Variances*, pp. 1–17. The MIT Press.

Sims, C. A. (1972). Money, Income, and Causality. *The American Economic Review 62*(4), 540 – 552.

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

Taylor, M. P. (1995). The Economics of Exchange Rates. *Journal of Economic Literature 33*(1), 13–47.

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

Worthington, A. and H. Higgs (2004, January). Transmission of equity returns and volatility in Asian developed and emerging markets: a multivariate GARCH analysis. *International Journal of Finance & Economics 9*(1), 71–80.

Woźniak, T. (2012). Testing Causality Between Two Vectors in Multivariate GARCH Models. EUI Working Papers ECO 2012/20, European University Institute, Flo-

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

rence, Italy. Download at: http://cadmus.eui.eu/bitstream/handle/1814/23337/ECO_2012_20.pdf.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Chapter 3

Bayesian Testing of Granger Causality in Markov Switching VARs

with Matthieu Droumaguet

Abstract. Recent economic developments have shown the importance of spillover and contagion effects in financial markets as well as in macroeconomic reality. Such effects are not limited to relations between the levels of variables but also impact on the volatility and the distributions. We propose a method of testing restrictions for Granger noncausality on all these levels in the framework of Markov-switching Vector Autoregressive Models. The conditions for Granger noncausality for these models were derived by Warne (2000). Due to the nonlinearity of the restrictions, classical tests have limited use. We, therefore, choose a Bayesian approach to testing. The computational tools for posterior inference consists of a novel Block Metropolis-Hastings sampling algorithm for estimation of the restricted models, and of standard methods of computing the Posterior Odds Ratio. The analysis may be applied to financial and macroeconomic time series with complicated properties, such as changes of parameter values over time and heteroskedasticity.

Matthieu Droumaguet is a doctorate candidate at the Department of Economics at the European University Institute. Email: matthieu.droumaguet@eui.eu

This paper was presented at the Econometric Society Australasian Meeting 2012 in Melbourne

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

3.1 Introduction

The concept of Granger causality was introduced by Granger (1969) and Sims (1972). One variable does not Granger-cause some other variable, if past and current information about the former cannot improve the forecast of the latter. Note that this concept refers to the forecasting of variables, in contrast to the causality concept based on *ceteris paribus* effects attributed to Rubin (1974) (for the comparison of the two concepts used in econometrics, see e.g. Lechner, 2011). Knowledge of Granger causal relations allows a researcher to formulate an appropriate model and obtain a good forecast of values of interest. But what is even more important, a Granger-causal relation, once established, informs us that past observations of one variable have a significant effect on the forecast value of the other, delivering crucial information about the relations between economic variables.

The original Granger causality concept refers to forecasts of conditional means. There are, however, extensions referring to the forecasts of higher-order conditional moments or to distributions. We present and discuss these in Section 3.3. Again, information that they deliver not only helps in performing good forecasts of the variables, but is crucial for decision-making in economic and financial applications as well.

Among the time series models that have been analyzed for Granger causality of different types are: a family of Vector Autoregressive Moving Average (VARMA) models (see Boudjellaba et al., 1994, and references therein), the Logistic Smooth Transition Vector Autoregressive (LST-VAR) model (Christopoulos and León-Ledesma, 2008), some models from the family of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models (Comte and Lieberman, 2000; Woźniak, 2012; Woźniak, 2012). Finally, Warne (2000) derived conditions for different types of Granger noncausality for the Markov-switching VAR models on

and at the 22nd EC^2 Conference: *Econometrics for Policy Analysis: after the Crisis and Beyond* in Florence in December 2011, as well as during seminars at the Australian National University, the National Bank of Poland and Deutsche Bundesbank. The authors thank Anders Warne, Joshua Chan, Peter R. Hansen, Andrzej Kocięcki, Helmut Lütkepohl, Massimiliano Marcellino, Mateusz Pipień, John Stachurski, Rodney Strachan for their useful comments on the paper.

which we focus in this study. We present the model and its estimation in Section 3.2, while in Section 3.3 the definitions for different types of noncausality and restrictions on parameters are given. Note that all these works analyzed *one period ahead* Granger noncausality (see Lütkepohl, 1993; Lütkepohl and Burda, 1997; Dufour et al., 2006, for h periods ahead inference in VAR models).

The testing of the restrictions meets multiple problems. The most important limitation in the classical approach is that neither the asymptotic nor finite-sample distribution of the estimator has been derived so far. Consequently, the asymptotic distributions of the Wald, Likelihood Ratio and Lagrange Multiplier tests are not known. Further, the restrictions on the parameters derived by Warne (2000) may result in several sets of restrictions associated with one hypothesis. Therefore, a hypothesis of noncausality may be represented by several restricted models. Finally, some of the restrictions are nonlinear functions of parameters. All these features of the Granger noncausality analysis for Markov-switching VARs makes classical testing of hypotheses difficult, if possible at all.

The contribution of this work is a Bayesian testing procedure that allows the testing of all the restrictions derived by Warne (2000) for different kinds of Granger noncausality, as well as for the inference on the hidden Markov process. None of the existing classical solutions to the problem of testing nonlinear restrictions on parameters that we describe in Section 3.4 is easily applicable to Markov-switching VAR models. The proposed approach consists of a Bayesian estimation of the unrestricted model, allowing for Granger causality, and of the restricted models, where the restrictions represent hypotheses of noncausality. For this purpose, we construct a novel Block Metropolis-Hastings sampling algorithm that allows for restricting the models. The algorithm is discussed in Section 3.4 and presented in Section 3.5. Having estimated the models, we compare competing hypotheses, represented by the unrestricted and the restricted models, with standard Bayesian methods using Posterior Odds Ratios and Bayes factors.

The main advantage of our approach is that we can test the nonlinear restrictions. The restrictions of all the considered types of noncausality may be tested. Thus, the analysis of causal relations between variables is profound and

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

potentially informative. Other advantages include an effect of adopting Bayesian inference. First, the Posterior Odds Ratio method gives arguments *in favour of* the hypotheses, as posterior probabilities of the competing hypotheses are compared. In consequence, all the hypotheses are treated symmetrically. Finally, our estimation procedure combines and improves the existing algorithms restricting the models, but it also preserves the possibility of using different methods for computing the marginal density of data necessary to compute the Posterior Odds Ratio. We discuss further the benefits and costs of our approach at the end of Section 3.4.

As potential applications of the testing procedure, we indicate macroeconomic as well as financial time series. In particular, recent financial turmoil and the following global recession are interesting periods for analysis. There exist many applied studies presenting evidence that these events have the nature of switching the regime. Taylor and Williams (2009), on the example of Libor-OIS and Libor-Repo spreads, being an approximation for counterpart risk, present how different the perception of the risk by agents on the financial market was, first, starting from August 2007 and then, even more, from October 2008. Further, Diebold and Yilmaz (2009) show how different behaviors characterize return spillovers and volatility spillovers for stock exchange markets. These two studies clearly indicate that the financial data should be analyzed in terms of Granger causality with a model that allows for changes in regimes, such as a Markov-switching model.

For macroeconomic time series, the motivation for using Markov-switching models comes mainly from the business cycle analysis, as in Hamilton (1989). It is important to know whether variables have different impacts on other variables during the expansion and recession periods. Still, allowing for higher number of states than two may allow a more detailed analysis of the interactions between variables within the cycles. For example, Psaradakis et al. (2005) used the Markov-switching VAR models to analyze, the so called temporary Granger causality within the Money-Output system. They condition their causality analysis on realizations of the Hidden-Markov process. They proposed a restricted MS-VAR specification that assumed four states of the economy: 1. both variables cause

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

each other; 2. money does not cause output; 3. output does not cause money; 4. none of the variables causes another. Our approach consists of choosing a Markov-switching VAR model specification which is best supported by the data, and then restricting it according to the restrictions derived by Warne (2000). This approach takes into account the two sources of relations between the variables: first, having a source in linear relations modeled with the VAR model, and second, taking into consideration the fact that all of the variables are used to forecast the future probabilities of the states. In the setting analyzed by Warne (2000) Granger noncausality is not conditioned on the past realizations of the hidden Markov process.

The remaining part of the paper is organized as follows. In Section 3.2 we present the model and the Bayesian estimation of the unrestricted model. The definitions for Granger noncausality, noncausality in variance and noncausality in distribution are presented in Section 3.3, together with parameter restrictions representing them. Section 3.4 presents discussion and critique of classical methods of testing restrictions for Granger noncausality in different multivariate models. The discussion is followed by a proposal of solution of the testing problem. First, the computation of the Posterior Odds Ratio is shown, and then the algorithm for estimating the restricted models is discussed. It is described in detail in Section 3.5. Section 3.6 gives empirical illustration of the methodology, using the example of the money-income system of variables in the USA. The data support the hypothesis of Granger noncausality (in mean) from money to income, as well as the hypotheses of causality in variance and distribution. Section 3.7 concludes.

3.2 A Markov-Switching Vector Autoregressive Model

Model Let $\mathbf{y} = (y_1, \ldots, y_T)'$ denote a time series of T observations, where each y_t is a N-variate vector for $t \in \{1, \ldots, T\}$, taking values in a sampling space $\mathbf{Y} \subset \mathbb{R}^N$. $\mathbf{y}$ is a realization of a stochastic process $\{Y_t\}_{t=1}^{T}$. We consider a class of parametric finite Markov mixture distribution models in which the stochastic process Y_t depends

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

on the realizations, s_t, of a hidden discrete stochastic process S_t with finite state space $\{1,\dots,M\}$. Such a class of models has been introduced in time series analysis by Hamilton (1989). Conditioned on the state, s_t, and realizations of $\mathbf{y}$ up to time $t-1$, $\mathbf{y}_{t-1}$, y_t follows an independent identical normal distribution. A conditional mean process is a Vector Autoregression (VAR) model in which an intercept, μ_{s_t}, as well as lag polynomial matrices, $A_{s_t}^{(i)}$, for $i = 1,\dots,p$, and covariance matrices, Σ_{s_t}, depend on the state $s_t = 1,\dots,M$.

$$y_t = \mu_{s_t} + \sum_{i=1}^{p} A_{s_t}^{(i)} y_{t-i} + \epsilon_t, \tag{3.1}$$

$$\epsilon_t \sim i.i.\mathcal{N}(\mathbf{0}, \Sigma_{s_t}), \tag{3.2}$$

for $t = 1,\dots,T$. We set the vector of initial values $\mathbf{y}_0 = (y_{p-1},\dots,y_0)'$ to the first p observations of the available data.

S_t is assumed to be an irreducible aperiodic Markov chain starting from its ergodic distribution $\pi = (\pi_1,\dots,\pi_M)$, such that $\Pr(S_0 = i|\mathbf{P}) = \pi_i$. Its properties are sufficiently described by the $(M \times M)$ transition probabilities matrix:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1M} \\ p_{21} & p_{22} & \dots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \dots & p_{MM} \end{bmatrix},$$

in which an element, p_{ij}, denotes the probability of transition from state i to state j, $p_{ij} = \Pr(s_{t+1} = j|s_t = i)$. All the transition probabilities are positive, $p_{ij} > 0$, for all $i, j \in \{1,\dots,M\}$, and the elements of each row of matrix $\mathbf{P}$ sum to one, $\sum_{j=1}^{M} p_{ij} = 1$.

Such a formulation of the model is called, according to the taxonomy of Krolzig (1997), MSIAH-VAR(p). Conditioned on the state s_t, it models a current vector of observations, y_t, with an intercept, μ_{s_t}, and a linear function of its lagged values up to p periods backwards. The linear relation is captured by matrices of the lag polynomial $A_{s_t}^{(i)}$, for $i = 1,\dots,p$. The parameters of the VAR process, as well as the covariance matrix Σ_{s_t}, change with time, t, according to discrete valued hidden Markov process, s_t. These changes in parameter values introduce nonlinear relationships between variables. Consequently, the inference about interactions between variables must consider the linear and nonlinear relations; this is the subject of the analysis in Section 3.3.

Complete-data likelihood function Let $\theta \in \Theta \subset \mathbb{R}^k$ be a vector of size k, collecting parameters of the transition probabilities matrix $\mathbf{P}$ and all the state-dependent parameters of the VAR process, θ_{s_t}: μ_{s_t}, $A_{s_t}^{(i)}$, Σ_{s_t}, for $s_t = 1,\dots,M$ and $i = 1,\dots,p$. As stated by Frühwirth-Schnatter (2006), the complete-data likelihood function is equal to the joint sampling distribution $p(\mathbf{S},\mathbf{y}|\theta)$ for the complete data $(\mathbf{S},\mathbf{y})$ given θ, where $\mathbf{S} = (s_1,\dots,s_T)'$. This distribution is now considered to be a function of θ for the purpose of estimating the unknown parameter vector θ. It is further decomposed into a product of a conditional distribution of $\mathbf{y}$ given $\mathbf{S}$ and θ, and a conditional distribution of $\mathbf{S}$ given θ:

$$p(\mathbf{S},\mathbf{y}|\theta) = p(\mathbf{y}|\mathbf{S},\theta)p(\mathbf{S}|\theta). \tag{3.3}$$

The former is assumed to be a conditional normal distribution function of ϵ_t, for $t = 1,\dots,T$, given the states, s_t, with the mean equal to a vector of zeros and Σ_{s_t} as the covariance matrix:

$$p(\mathbf{y}|\mathbf{S},\theta) = \prod_{t=1}^{T} p(y_t|\mathbf{S},\mathbf{y}^{t-1},\theta) = \prod_{t=1}^{T} (2\pi)^{-K/2}|\Sigma_{s_t}|^{-1/2}\exp\left\{-\frac{1}{2}\epsilon_t'\Sigma_{s_t}^{-1}\epsilon_t\right\}. \tag{3.4}$$

The form of the latter comes from the assumptions about the Markov process and is given by:

$$p(\mathbf{S}|\theta) = p(s_0|\mathbf{P})\prod_{i=1}^{M}\prod_{j=1}^{M} p_{ij}^{N_{ij}(\mathbf{S})}, \tag{3.5}$$

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

where $N_{ij}(\mathbf{S}) = \#\{s_{t-1} = j, s_t = i\}$ is a number of transitions from state i to state j, $\forall i, j \in \{1, \ldots, M\}$.

A convenient form of the complete-data likelihood function (3.3) results from representing it as a product of $M + 1$ factors. The first M factors depend on the state-specific parameters, θ_{s_t}, and the remaining one depends on the transition probabilities matrix, $\mathbf{P}$:

$$p(\mathbf{y}, \mathbf{S}|\theta) = \prod_{i=1}^{M} \left(\prod_{t:s_t=i} p(y_t|\mathbf{y}^{t-1}, \theta_i) \right) \prod_{i=1}^{M} \prod_{j=1}^{M} p_{ij}^{N_{ij}(\mathbf{S})} p(s_0|\mathbf{P}). \tag{3.6}$$

Classical estimation of the model consists of the maximization of the likelihood function with e.g. the EM algorithm (see Krolzig, 1997; Kim and Nelson, 1999b). For the purpose of testing Granger-causal relations between variables, we propose, however, the Bayesian inference, which is based on the posterior distribution of the model parameters θ. (For details of a standard Bayesian estimation and inference on Markov-switching models, the reader is referred to Frühwirth-Schnatter, 2006). The complete-data posterior distribution is proportional to the product of the complete-data likelihood function (3.6) and the prior distribution:

$$p(\theta|\mathbf{y}, \mathbf{S}) \propto p(\mathbf{y}, \mathbf{S}|\theta)p(\theta). \tag{3.7}$$

Prior distribution The convenient factorization of the likelihood function (3.6) is maintained by the choice of the prior distribution in the following form:

$$p(\theta) = \prod_{i=1}^{M} p(\theta_i)p(\mathbf{P}_{i\cdot}). \tag{3.8}$$

The independence of the prior distribution of the state-specific parameters for each state and the transition probabilities matrix is assumed. This allows the possibility to incorporate prior knowledge of the researcher about the state-specific parameters of the model, θ_{s_t}, separately for each state.

For the unrestricted MSIAH-VAR(p) model, we assume the following prior

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

specification. Each row of the transition probabilities matrix, $\mathbf{P}$, *a priori* follows an M variate Dirichlet distribution, with parameters set to 1 for all the transition probabilities except the diagonal elements $\mathbf{P}_{ii}$, for $i = 1, \ldots, M$, for which it is set to 10. Therefore, we assume that the states of an economy are persistent over time (see e.g. Kim and Nelson, 1999a). Further, the state-dependent parameters of the VAR process are collected in vectors $\beta_{s_t} = (\mu_{s_t}', \text{vec}(A_{s_t}^{(1)})', \ldots, \text{vec}(A_{s_t}^{(p)})')'$, for $s_t = 1, \ldots, M$. These parameters follow a $(N + pN^2)$-variate Normal distribution, with mean equal to a vector of zeros and a diagonal covariance matrix with 100s on the diagonal. Note that the means of the prior distribution for the off-diagonal elements of matrices A_{s_t} are set to zero. If we condition our analysis on the states, this would mean that we assume *a priori* the Granger noncausality hypothesis. However, in Section 3.3 we show that, when the states are unknown, the inference about Granger noncausality involves many other parameters of the model. Moreover, huge values of the variances of the prior distribution are assumed. Consequently, no values from the interior of the parameters space are, in fact, discriminated *a priori*.

We model the state-dependent covariance matrices of the MSIAH-VAR process, decomposing each to a $N \times 1$ vector of standard deviations, σ_{s_t}, and a $N \times N$ correlation matrix, $\mathbf{R}_{s_t}$, according to the decomposition:

$$\Sigma_{s_t} = \text{diag}(\sigma_{s_t})\mathbf{R}_{s_t}\text{diag}(\sigma_{s_t}).$$

Modeling covariance matrices using such a decomposition was proposed in Bayesian inference by Barnard et al. (2000). We adapt this approach to Markov-switching models, since the algorithm easily enables the imposing of restrictions on the covariance matrix (see the details of the Gibbs sampling algorithm for the unrestricted and the restricted models in Section 3.5). We model the unrestricted model in the same manner, because we want to keep the prior distributions for the unrestricted and the restricted models comparable. Thus, each standard deviation $\sigma_{s_t.j}$ for $s_t = 1, \ldots, M$ and $j = 1, \ldots, N$, follows a log-Normal distribution, with a mean parameter equal to 0 and the standard deviation parameter set to 2.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Finally, we assume that the prior distributions of each of correlation coefficient $\mathbf{R}_{s_t.jk}$ is uniformly-distributed at interval (a, b). The bounds a and b are set such that sampling individual correlations one by one results in positive definite correlation matrix, $\mathbf{R}_{s_t}$. For the implications of such a prior specification for the matrix of correlations and for the algorithm for setting values a and b the reader is referred to the original paper of Barnard et al. (2000).

To summarize, the prior specification (3.8) now takes the detailed form of:

$$p(\theta) = \prod_{i=1}^{M} p(\mathbf{P}_i)p(\beta_i)p(\mathbf{R}_i)\left(\prod_{j=1}^{N} p(\sigma_{i.j})\right), \tag{3.9}$$

where each of the prior distributions is as assumed:

$$\begin{aligned}
\mathbf{P}_{i\cdot} &\sim \mathcal{D}_M(\imath_M' + 9I_{M.i\cdot}) \\
\beta_i &\sim \mathcal{N}(\mathbf{0}, 100 I_{N+pN^2}) \\
\sigma_{i.j} &\sim log\mathcal{N}(0, 2) \\
\mathbf{R}_{i.jk} &\sim \mathcal{U}(a, b)
\end{aligned}$$

for $i = 1, \ldots, M$ and $j, k = 1, \ldots, N$, where $\imath_M$ is a $M \times 1$ vector of ones and $I_{M.i\cdot}$ is i^{th} row of an identity matrix I_M.

Posterior distribution The structure of the likelihood function (3.6) and the prior distribution (3.9) have an effect on the form of the posterior distribution that is proportional to the product of the two densities. The form of the posterior distribution (3.7), resulting from the assumed specification, is as follows:

$$p(\theta|\mathbf{y}, \mathbf{S}) \propto \prod_{i=1}^{M} p(\theta_i|\mathbf{y}, \mathbf{S})p(\mathbf{P}|\mathbf{y}, \mathbf{S}). \tag{3.10}$$

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

It is now easily decomposed into a posterior density of the transition probabilities matrix:

$$p(\mathbf{P}|\mathbf{S}) \propto p(s_0|\mathbf{P}) \prod_{i=1}^{M} \prod_{j=1}^{M} p_{ij}^{N_{ij}(\mathbf{S})} p(\mathbf{P}), \tag{3.11}$$

and the posterior density of the state-dependent parameters:

$$p(\theta_i|\mathbf{y}, \mathbf{S}) \propto \prod_{t:S_t=i} p(y_t|\theta_i, \mathbf{y}_{t-1},)p(\theta_i). \tag{3.12}$$

Since the form of the posterior density for all the parameters is not standard, the commonly used strategy is to simulate the posterior distribution with numerical methods. A Monte Carlo Markov Chain (MCMC) algorithm, the Gibbs sampler (see Casella and George, 1992, and references therein), enables us to simulate the joint posterior distribution of all the parameters of the model by sampling from the full conditional distributions. Such an algorithm has also been adapted to Markov-switching models by Albert and Chib (1993) and McCulloch and Tsay (1994). However, the model specification considered in this study results in full conditional distributions that are not in a form of any standard distribution functions. Therefore, the algorithm that samples from such full conditional distributions belongs to a broader class of Block Metropolis-Hastings algorithms. The algorithm is presented in detail in Section 3.5.

3.3 Granger Causality - Following Warne (2000)

Notation Let $\{y_t : t \in \mathbb{Z}\}$ be a $N \times 1$ multivariate square integrable stochastic process on the integers $\mathbb{Z}$. Write:

$$y_t = (y_{1t}', y_{2t}', y_{3t}', y_{4t}')', \tag{3.13}$$

for $t = 1, \ldots, T$, where y_{it} is a $N_i \times 1$ vector such that $y_{1t} = (\mathrm{y}_{1t}, \ldots, \mathrm{y}_{N_1.t})'$, $y_{2t} = (\mathrm{y}_{N_1+1.t}, \ldots, \mathrm{y}_{N_1+N_2.t})'$, $y_{3t} = (\mathrm{y}_{N_1+N_2+1.t}, \ldots, \mathrm{y}_{N_1+N_2+N_3.t})'$, and $y_{4t} = (\mathrm{y}_{N_1+N_2+N_3+1.t}, \ldots, \mathrm{y}_{N_1+N_2+N_3+N_4.t})'$ $(N_1, N_4 \geq 1, N_2, N_3 \geq 0$ and $N_1 + N_2 + N_3 + N_4 = N)$. Variables of interest are con-

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

tained in vectors y_1 and y_4, between which we want to study causal relations. Vectors y_2 and y_3 (that for $N_2 = 0$ and $N_3 = 0$ are empty) contain auxiliary variables that are also used for forecasting and modeling purposes. Finally, define two vectors: first $(N_1 + N_2)$-dimensional, $v_{1t} = (y'_{1t}, y'_{2t})'$, and second $(N_3 + N_4)$-dimensional, $v_{2t} = (y'_{3t}, y'_{4t})'$, such that:

$$y_t = \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}.$$

Suppose that there exists a proper probability density function $f_t(y_{t+1}|\mathbf{y}_t;\theta)$ for each $t \in \{1, 2, \ldots, T\}$. Suppose that the conditional mean $E[y_{t+1}|\mathbf{y}_t]$ is finite and that the conditional covariance matrix:

$$E\left[(y_{t+1} - E[y_{t+1}|\mathbf{y}_t])(y_{t+1} - E[y_{t+1}|\mathbf{y}_t])'|\mathbf{y}_t\right]$$

positive definite for all finite t. Further, let u_{t+1} denote 1-step ahead forecast error for y_{t+1}, conditional on $\mathbf{y}_t$ when the predictor is given by the conditional expectations, i.e.:

$$u_{t+1} = y_{t+1} - E[y_{t+1}|\mathbf{y}_t]. \tag{3.14}$$

By construction, u_{t+1} has conditional mean zero and positive-definite conditional covariance matrix. And let $\tilde{u}_{t+1} = y_{t+1} - E[y_{t+1}|\mathbf{v}_{1t}, \mathbf{y}_{3t}]$ be 1-step ahead forecast error for y_{t+1}, conditional on $\mathbf{v}_{1t}$ and $\mathbf{y}_{3t}$ with analogous properties.

Definitions We focus on the Granger-causal relations between variables y_1 and y_4. The first definition of *Granger causality*, originally given by Granger (1969), states simply that y_4 is not causal for y_1 when the past and current information about, $\mathbf{y}_{4.t}$ cannot improve mean square forecast error of $y_{1.t+1}$.

Definition 5. y_4 does not Granger-cause y_1, denoted by $y_4 \overset{G}{\nrightarrow} y_1$, if and only if:

$$E\left[u_{t+1}^2\right] = E\left[\tilde{u}_{t+1}^2\right] < \infty \quad \forall t = 1, \ldots, T. \tag{3.15}$$

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

This definition refers to the conditional mean process, and holds if and only if the two means conditioned on the full set of variables, $\mathbf{y}_t$, and on the restricted set, $(\mathbf{v}_{1t}, \mathbf{y}_{3t})$, are the same (see Boudjellaba et al., 1992). It is argued, however, that this definition cannot give a full insight into relations between variables under changing economic circumstances: if the series is heteroskedastic, then it is useful to refer to a different concept of causality, namely *Granger causality in variance*, introduced by Robins et al. (1986). It states the noncausality condition for conditional second-order moments of the series. Note that this definition states noncausality in conditional covariance as well as in conditional mean processes. Therefore, this condition is stricter than (3.15).

Definition 6. y_4 does not Granger-cause in variance y_1, denoted by $y_4 \overset{V}{\nrightarrow} y_1$, if and only if:

$$E\left[u_{t+1}^2|\mathbf{y}_t\right] = E\left[\tilde{u}_{t+1}^2|\mathbf{v}_{1t}, \mathbf{y}_{3t}\right] < \infty \quad \forall t. \tag{3.16}$$

Finally, we define the third concept of Granger noncausality, *Granger noncausality in distribution*.

Definition 7. y_4 does not Granger-cause in distribution y_1, denoted by $y_4 \overset{D}{\nrightarrow} y_1$, if and only if:

$$g_{t+1}\left(u_{t+1}^2|\mathbf{y}_t, \theta\right) = h_{t+1}\left(\tilde{u}_{t+1}^2|\mathbf{v}_{1t}, \mathbf{y}_{3t}, \theta\right) \quad \forall t, \tag{3.17}$$

where g_{t+1} and h_{t+1} are probability distribution functions with properties as for f_{t+1}.

All the definitions are given in the form following Warne (2000). Note that the definition of Granger noncausality in variance (3.16) is stricter than the definition of Granger noncausality (3.15); Granger noncausality in variance implies Granger noncausality. The definition of Granger noncausality in distribution (3.17) is defined for conditional distributions. It applies also to these distributions that have their moments undefined. All three definitions are, however, identical in linear Gaussian models.

Comte and Lieberman (2000) introduce a new definition of *second-order Granger noncausality* and distinguish it from *Granger noncausality in variance* of Robins et al.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

(1986). For the second-order noncausality, if there exists Granger causality (in mean), then it needs to be modeled and filtered out; only then may the causal relations in conditional second moments be established. The definition of non-causality in variance assumes Granger noncausality (in mean) and second-order noncausality, and therefore is stricter than second-order noncausality. In effect, once Granger noncausality is established, the two definitions, noncausality in variance and second-order noncausality, are equivalent. The consequences of testing these different concepts are presented in Woźniak (2012).

MSIAH-VARs for Granger causality testing We now present the parameter restrictions for different definitions of Granger noncausality for Markov-switching vector autoregressions. Before that, however, we introduce the more convenient formulation of the model specified in Section 3.2. Firstly, we use the decomposition of the vector of observations into two sub-vectors, $y_t = (v'_{1t}, v'_{2t})'$, and appropriate decomposition of the parameter matrices, μ_{s_t}, $A^{(l)}_{s_t}$, and vector of residuals, ϵ_t, which has covariance matrix specified in (3.19). Also, the hidden Markov process is decomposed for the purpose of setting the Granger causality relations into two sub-processes, $s_t = (s_{1t}, s_{2t})$. The sub-processes have M_1 and M_2 states that are characterized by transition probability matrices, $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ (and ergodic probabilities, $\pi^{(1)}$ and $\pi^{(2)}$) respectively, such that $M = M_1 \cdot M_2$. The construction of the transition probabilities matrix, $\mathbf{P}$, is not specified for the moment and will be the subject of further analysis. Parameters of the equation for v_{1t} change in time with the Markov process s_{1t}, whereas the parameters of the equation for v_{2t} change with process s_{2t}:

$$\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1.s_{1t}} \\ \mu_{2.s_{2t}} \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} A^{(i)}_{11.s_{1t}} & A^{(i)}_{12.s_{1t}} \\ A^{(i)}_{21.s_{2t}} & A^{(i)}_{22.s_{2t}} \end{bmatrix} \begin{bmatrix} v_{1t-i} \\ v_{2t-i} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}. \tag{3.18}$$

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

The residual term in (3.18) has zero conditional mean and conditional covariance matrix decomposed into sub-matrices as on the left-hand side of (3.19):

$$\text{Var}\left(\begin{bmatrix}\epsilon_{1t}\\ \epsilon_{2t}\end{bmatrix}\right)=\begin{bmatrix}\Sigma_{11.s_{1t}} & \Sigma'_{21.s_t}\\ \Sigma_{21.s_t} & \Sigma_{22.s_{2t}}\end{bmatrix},\qquad \text{Var}\left(\begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\\ \varepsilon_{3t}\\ \varepsilon_{4t}\end{bmatrix}\right)=\begin{bmatrix}\Omega_{11.s_{1t}} & \Omega'_{21.s_t} & \Omega'_{31.s_t} & \Omega'_{41.s_t}\\ \Omega_{21.s_t} & \Omega_{22.s_{1t}} & \Omega'_{32.s_t} & \Omega'_{42.s_t}\\ \Omega_{31.s_t} & \Omega_{32.s_t} & \Omega_{33.s_{2t}} & \Omega'_{43.s_t}\\ \Omega_{41.s_t} & \Omega_{42.s_t} & \Omega_{43.s_t} & \Omega_{44.s_{2t}}\end{bmatrix},\tag{3.19}$$

where covariance matrices may be decomposed respectively into:

$$\Sigma_{ij.s_t}=\text{diag}(\sigma_{i.s_t})\mathbf{R}_{ij.s_t}\text{diag}(\sigma_{j.s_t}),\qquad \Omega_{ij.s_t}=\text{diag}(\omega_{i.s_t})R_{ij.s_t}\text{diag}(\omega_{j.s_t}).\tag{3.20}$$

We further decompose vectors of observations, $v_{1t}=(y'_{1t},y'_{2t})'$ and $v_{2t}=(y'_{3t},y'_{4t})'$, matrices of model parameters with the covariance matrix of the residual term specified on the right-hand side of (3.19). The decomposition of the Markov process is maintained, as in (3.18):

$$\begin{bmatrix}y_{1t}\\ y_{2t}\\ y_{3t}\\ y_{4t}\end{bmatrix}=\begin{bmatrix}m_{1.s_{1t}}\\ m_{2.s_{1t}}\\ m_{3.s_{2t}}\\ m_{4.s_{2t}}\end{bmatrix}+\sum_{i=1}^{p}\begin{bmatrix}a^{(i)}_{11.s_{1t}} & a^{(i)}_{12.s_{1t}} & a^{(i)}_{13.s_{1t}} & a^{(i)}_{14.s_{1t}}\\ a^{(i)}_{21.s_{1t}} & a^{(i)}_{22.s_{1t}} & a^{(i)}_{23.s_{1t}} & a^{(i)}_{24.s_{1t}}\\ a^{(i)}_{31.s_{2t}} & a^{(i)}_{32.s_{2t}} & a^{(i)}_{33.s_{2t}} & a^{(i)}_{34.s_{2t}}\\ a^{(i)}_{41.s_{2t}} & a^{(i)}_{42.s_{2t}} & a^{(i)}_{43.s_{2t}} & a^{(i)}_{44.s_{2t}}\end{bmatrix}\begin{bmatrix}y_{1t-i}\\ y_{2t-i}\\ y_{3t-i}\\ y_{4t-i}\end{bmatrix}+\begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\\ \varepsilon_{3t}\\ \varepsilon_{4t}\end{bmatrix}.\tag{3.21}$$

Parameter restrictions The parameter restrictions for Markov-switching vector autoregressions for the three definitions of Granger noncausality presented in this section have been derived by Warne (2000). Firstly, we present the restrictions that are specific for the Markov-switching models. The Restriction 1 states the relations between the two Markov processes , s_{1t} and s_{2t}.

Restriction 1. The regime forecast of $s_{1.t+1}$ is independent, and there is no information in $\mathbf{v}_{2t}$ for predicting $s_{1.t+1}$, i.e.:

$$Pr\left[(s_{1.t+1},s_{2.t+1})=(j_1,j_2)|\mathbf{y}_t,\theta\right]=Pr\left[s_{1.t+1}=j_1|\mathbf{v}_{1t},\theta\right]\cdot Pr\left[s_{2.t+1}=j_2|\mathbf{v}_{2t},\theta\right],$$

DOI: 10.2870/63858

for all $j_1 = 1, \ldots, M_1$ with $M_1 \geq 2$, $j_2 = 1, \ldots, M_2$ and $t = 1, \ldots, T$, if and only if either:

(A1): (i) $\mathbf{P} = (\mathbf{P}^{(1)} \otimes \mathbf{P}^{(2)})$,

(ii) $\mu_{i.s_t} = \mu_{i.s_{i.t}}$,

(iii) $A^{(k)}_{ij.s_t} = A^{(k)}_{ij.s_{i.t}}$,

(iv) $\Sigma_{ii.s_t} = \Sigma_{ii.s_{i.t}}$ and

(v) $\Sigma_{12.s_t} = 0$

for all $i, j \in \{1, 2\}$, $k \in \{1, \ldots, p\}$ and $s_{1.t} \in \{1, \ldots, q_1\}$, and

(vi) $A^{(k)}_{12.s_{1.t}} = 0$ for all $k \in \{1, \ldots, p\}$ and $s_{1.t} \in \{1, \ldots, q_1\}$; or

(A2): $\mathbf{P} = (\iota_{M_1} \pi^{(1)\prime} \otimes \mathbf{P}^{(2)})$,

is satisfied.

Note that if we change the restriction (A1)(vi) into $A^{(k)}_{21.s_{1.t}} = 0$, then there is no information in $\mathbf{v}_{1t}$ for predicting $s_{2.t+1}$.

Restriction (A1)(i) gives the condition for independence of the transition probabilities. Restrictions (A1)(ii)-(A1)(iv) state simply that the parameters of the equation for $\mathbf{v}_{1t}$ change only according to the process s_{1t}, and the parameters of the equation for $\mathbf{v}_{2t}$ change only according to the process s_{2t}. Consequently, the decomposition of the hidden Markov process s_t into two independent subprocesses (s_{1t}, s_{2t}) is fully respected. Further, restriction (A1)(v) states the instantaneous noncausality between the two vectors of variables, $\mathbf{v}_{1t}$ and $\mathbf{v}_{2t}$, defined as zero correlation condition. Finally, restriction (A1)(vi) states the Granger noncausality condition for the VAR process. According to condition (A2), all the states of process s_{1t} have the same probability of appearance for all t equal to the ergodic probability, $\pi^{(1)}$, which is a condition for s_{1t} to be an independent hidden Markov chain.

Before we go on to the conditions for different types of Granger noncausality, we define the conditional expected values of the parameters of the VAR process

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

for one period ahead forecast:

$$\bar{m}_{1t} \equiv E\left[m_{1.s_{t+1}}|\mathbf{y}_t, \theta\right], \tag{3.22a}$$

$$\bar{a}^{(k)}_{1rt} \equiv E\left[a^{(k)}_{1r.s_{t+1}}|\mathbf{y}_t, \theta\right], \tag{3.22b}$$

for all $r = 1, \ldots, N$ and $k = 1, \ldots, p$. These parameters are used for forecasting of variable $y_{1.t+1}$ (see equation (16) of Warne, 2000), as well as for the purpose of setting noncausality conditions. Restriction 2 states the conditions for Granger noncausality.

Restriction 2. y_4 does not Granger-cause y_1 if and only if either:

(A1) or

(A3): (i) $\sum_{j=1}^{M} m_{1.j} p_{ij} = \bar{m}_1$,

(ii) $\sum_{j=1}^{M} a^{(k)}_{1r.j} p_{ij} = \bar{a}^{(k)}_{1r}$, and

(iii) $\bar{a}^{(k)}_{14} = 0$

for all $i \in \{1, \ldots, M\}$, $r \in \{1, \ldots, N\}$, and $k \in \{1, \ldots, p\}$,

is satisfied.

Contrary to conditions (A1) and (A2), the condition (A3) is not linear in parameters. Still, conditions (A3)(i) and (A3)(ii) have equivalent form, $\sum_{j=1}^{M} m_{1,j}(p_{ij} - p_{kj}) = 0$ for $i, k = 1, \ldots, M$ and $i \neq k$, which for some special cases may give restrictions linear in parameters. The condition (A3)(iii) does not have such a form and thus stays nonlinear. Further, in Section 3.4 we discuss consequences of the nonlinearity of the restrictions for testing them. Restriction 3 for noncausality in variance contains highly nonlinear conditions as well.

Restriction 3. y_4 does not Granger-cause in variance y_1 if and only if either:

(A1) or

(A4): (i) (A2),

(ii) $\sum_{j=1}^{M}\left[(m_{1.j}-\bar{m}_1)\otimes(m_{1.j}-\bar{m}_1)\right]p_{ij}=\varsigma_m$,

(iii) $\sum_{j=1}^{M}\left[(a_{1r.j}^{(k)}-\bar{a}_{1r}^{(k)})\otimes(a_{1s.j}^{(l)}-\bar{a}_{1s}^{(l)})\right]p_{ij}=\varsigma_{r.s}^{(k.l)}$,

(iv) $\sum_{j=1}^{M}\left[(m_{1.j}-\bar{m}_1)\otimes(a_{1r.j}^{(k)}-\bar{a}_{1r}^{(k)})\right]p_{i.j}=\varsigma_{\mu.r}^{(k)}$,

(v) $\sum_{j=1}^{M}\sigma_{1.j}p_{ij}=\varsigma_\sigma$, and

(vi) $a_{14.j}^{(k)}=0$,

for all $i,j\in\{1,\dots,M\}$, $r,s\in\{1,2,3\}$, and $k,l\in\{1,\dots,p\}$,

is satisfied.

In condition (A4), ς_m, $\varsigma_{r.s}^{(k.l)}$, $\varsigma_{\mu.r}^{(k)}$ and ς_σ, are time-invariant covariance matrices of the conditional expected value of the one period ahead forecast of the state-dependent parameters (see Warne, 2000, for the exact definition). Some of these restrictions may be simplified using the algebraically equivalent form: $\sum_{j=1}^{M}(m_{1.j}\otimes m_{1.j})p_{ij}=\varsigma_m+(\bar{m}_1\otimes\bar{m}_1)$.

Finally, we present Restriction 4, which states the conditions for noncausality in distribution.

Restriction 4. y_4 does not Granger-cause in distribution y_1 if and only if either:

(A1) or

(A5): **(i)** (A2)

(ii) $m_{1.j}=m_{1.j_1}$,

(iii) $a_{1r.j}^{(k)}=a_{1r.j_1}^{(k)}$,

(iv) $a_{14.j}^{(k)}=0$, and

(v) $\sigma_{1.j}=\sigma_{1.j_1}$

for all $j\in\{1,\dots,M\}$, $r\in\{1,2,3\}$, and $k\in\{1,\dots,p\}$

is satisfied.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

All the Restrictions 4 are linear in parameters and can be easily tested. Conditions (A5)(ii)–(A5)(v) state simply that the parameters of the equation for y_{1t} cannot vary in time according to process s_{1t} , but should instead be s_{1t}-invariant.

Warne (2000) sets additional and simplified forms of restrictions (A3)–(A5), given the condition (A2) and that rank($\mathbf{P}^{(2)}$) = M_2. We present these in C.1.

3.4 Bayesian Testing

Restrictions 1–4 can be tested. We first consider classical tests and their limitations and then present the Bayesian testing procedure as a solution. The obstacles in using classical tests are threefold:

- The asymptotic distribution of the parameters of the MS-VAR is unknown;
- The conditions for noncausality may result in several sets of restrictions on parameters. Consequently, one hypothesis may be represented by several restricted models;
- Some of the restrictions are in the form of nonlinear functions of parameters of the model.

The proposed solution consists of a new Block Metropolis-Hastings sampling algorithm for the estimation of the restricted models, and of the application of a standard Bayesian test to compare the restricted models to the unrestricted one.

Classical testing In the general case, all the mentioned problems with classical testing are difficult to cope with. While, the lack of the asymptotic distribution of the parameters could be solved using simulation methods, the problem of testing a hypothesis represented by several restricted models seems unsolvable with existing classical methods.

The problem of the nonlinearity of the restrictions, however, is well known in the studies on testing parameter conditions for Granger noncausality in multivariate models. In the general case, nonlinear restrictions on parameters of the model

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

may result in the matrix of partial derivatives of the restrictions with respect to the parameters not having a full rank. Consequently, the asymptotic distribution of test statistic is not known.

This problem was met in several studies on Granger noncausality testing in time series models. Boudjellaba et al. (1992) derive conditions for Granger noncausality for VARMA models that result in multiple nonlinear restrictions on original parameters of the model. As a solution to the problem of testing the restrictions, they propose a sequential testing procedure. There are two main drawbacks of this method. First, despite properly performed procedure, the test may still appear inconclusive, and second, the confidence level is given in the form of inequalities. The problem of testing non-linear restrictions was examined for h-periods ahead Granger causality for VAR models. Dufour et al. (2006) propose the solution based on formulating a new model for each h, and obtain linear restrictions on the parameters on the model. These restrictions can be easily tested with standard tests. In another work by Dufour (1989) the approach is based on the linear regression theory; its solutions would require separate proofs in order to apply it to Markov-switching VARs. Finally, Lütkepohl and Burda (1997) propose a solution for testing nonlinear hypotheses based on a modification of the Wald test statistic. Given the asymptotic normality of the estimator of the parameters, the method uses a modification that, together with standard asymptotic derivations, overcomes the singularity problem.

Finally, the problem of testing the nonlinear restrictions was faced by Warne (2000), who derives the restrictions for Granger noncausality, noncausality in variance and noncausality in distribution for Markov-switching VAR models. Among the solutions reviewed in this Section, only that proposed by Lütkepohl and Burda (1997) seems applicable to this particular problem. This finding should, however, be followed with further studies proving its applicability.

Bayesian testing In this study we propose a method of solving the problems of testing the parameter restrictions based on Bayesian inference. This approach to testing the noncausality conditions was used by Woźniak (2012, 2012). Both of the

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

papers work on the Extended CCC-GARCH model of Jeantheau (1998). Two other works use the Bayesian approach to make inference about a concept somehow related to Granger noncausality, namely exogeneity. Jarociński and Maćkowak (2011) use Savage-Dickey's Ratio to test block-exonegeneity in the VAR model, while Pajor (2011) uses Bayes factors to assess exogeneity conditions for models with latent variables, and in particular in multivariate Stochastic Volatility models.

In order to compare the unrestricted model, denoted by $\mathcal{M}_i$, and the restricted model, $\mathcal{M}_j$ and $= j \neq i$, we use the Posterior Odds Ratio (POR), which is a ratio of the posterior probabilities, $Pr(\mathcal{M}|\mathbf{y})$, attached to each of these models representing the hypotheses:

$$\text{POR} = \frac{Pr(\mathcal{M}_i|\mathbf{y})}{Pr(\mathcal{M}_j|\mathbf{y})} = \frac{p(\mathbf{y}|\mathcal{M}_i)}{p(\mathbf{y}|\mathcal{M}_j)}\frac{Pr(\mathcal{M}_i)}{Pr(\mathcal{M}_j)}, \tag{3.23}$$

where $p(\mathbf{y}|\mathcal{M})$ is the marginal density of data and $Pr(\mathcal{M})$ is the prior probability of a model. In order to compare two competing models, one might also consider using Bayes factors, defined by:

$$\mathcal{B}_{ij} = \frac{p(\mathbf{y}|\mathcal{M}_i)}{p(\mathbf{y}|\mathcal{M}_j)}. \tag{3.24}$$

Note that if one chooses not to discriminate any of the models *a priori*, setting equal prior probabilities for both of the models ($Pr(\mathcal{M}_i)/Pr(\mathcal{M}_j) = 1$), the Posterior Odds Ratio is then equal to a Bayes factor. This method of testing does not have any of the drawbacks of the Likelihood Ratio test, once samples of draws from the posterior distributions of parameters for both the models are available (see Geweke, 1995; Kass and Raftery, 1995).

In this work, in order to asses the credibility of the hypotheses, each of which is represented by several sets of restrictions – and thus several models – we compute Posterior Odds Ratios. The results of this analysis are reported in Table 3.6 in Section 3.6. Suppose that a hypothesis is represented by several models. Let $\mathcal{H}_i$ denote the set of indicators of the models that represent this hypothesis, $\mathcal{H}_i = \{j : \mathcal{M}_j \text{ represents } i^{th} \text{ hypothesis}\}$. For instance, in our example, the hypothesis of Granger noncausality in mean is represented by four models, such that $\mathcal{H}_2 =$

$\{1, 2, 4, 5\}$. Further, suppose that one is interested in comparing the posterior probability of this hypothesis to the hypothesis $\mathcal{H}_0$, represented by the unrestricted model $\mathcal{M}_0$. Then the credibility of the hypothesis $\mathcal{H}_i$ compared to the hypothesis $\mathcal{H}_0$ may be assessed with the Posterior Odds Ratio given by:

$$\text{POR} = \frac{Pr(\mathcal{H}_i|\mathbf{y})}{Pr(\mathcal{H}_0|\mathbf{y})} = \frac{\sum_{j\in\mathcal{H}_i} Pr(\mathbf{y}|\mathcal{M}_j)Pr(\mathcal{M}_j)}{Pr(\mathbf{y}|\mathcal{M}_0)Pr(\mathcal{M}_0)}. \tag{3.25}$$

We set equal prior probabilities for all the models, which has the effect that none of the models is preferred *a priori*.

Testing the noncausality restrictions in MS-VARs Taking into account the complicated structure of the restrictions, we choose Posterior Odds Ratio (3.23) to assess the hypotheses. The crucial element of this method the is computation of marginal data densities, $p(\mathbf{y}|\mathcal{M})$, for the unrestricted and the restricted models. There are several available methods of computing this value. In this study we choose the Modified Harmonic Mean (MHM) method of Geweke (1999). For a chosen model, given the sample of draws, $\{\theta^{(i)}\}_{i=1}^{S}$, from the posterior distribution of the parameters, $p(\theta|\mathbf{y}, \mathcal{M})$, the marginal density of data is computed using:

$$p(\mathbf{y}|\mathcal{M}) = \left(S^{-1} \sum_{i=1}^{S} \frac{h(\theta^{(i)})}{L(\mathbf{y}; \theta^{(i)}, \mathcal{M})p(\theta^{(i)}|\mathcal{M})} \right)^{-1}, \tag{3.26}$$

where $L(\mathbf{y}; \theta^{(i)}, \mathcal{M})$ is a likelihood function od model $\mathcal{M}$. $h(\theta^{(i)})$, as specified in Geweke (1999), is a k-variate truncated normal distribution with mean parameter equal to the posterior mean and covariance matrix set to the posterior covariance matrix of θ. The truncation must be such that $h(\theta)$ had thinner tails than the posterior distribution.

Other methods of computing the marginal density of data may also be employed. Several estimators were derived, taking into account the characteristics of Markov-switching models. The reader is referred to the original papers by Frühwirth-Schnatter (2004), Sims et al. (2008) and Chib and Jeliazkov (2001).

 DOI: 10.2870/63858

Moreover, Frühwirth-Schnatter (2004) rises the problem of the bias of the estimators when the label permutation mechanism is missing in the algorithm sampling from the posterior distribution of the parameters. The bias appears to be due to the invariance of the likelihood function and the prior distribution of the parameters, with respect to permutations of the regimes' labels. Then the model is not globally identified. The identification can be insured by the ordering restrictions on parameters, and can also be implemented within the Gibbs sampler. Simply, it is sufficient that the values taken by one of the parameters of the model in different regimes can be ordered, and that the ordering holds for all the draws from the Gibbs algorithm to assure global identification (see Frühwirth-Schnatter, 2004). We assure that this is the case, i.e. that the MS-VAR models considered for causality inference are globally identified by the ordering imposed on some parameter.

Another element of the testing procedure is the estimation of the unrestricted model and the restricted models representing hypotheses of interest. We present a new Block Metropolis-Hastings sampling algorithm specially constructed for the purpose of testing noncausality hypotheses in the MS-VAR models in Section 3.5. It enables the imposing of restrictions on parameters resulting from conditions (A1) - (A7), and in effect testing different hypotheses of Granger noncausality between variables. In the algorithm, the restrictions are imposed on different groups of the parameters of the model. First, linear restrictions on the parameters of the VAR process, β, are implemented according to Frühwirth-Schnatter (2006). Next, parameters of the covariance matrices are decomposed into standard deviations, σ, and correlation parameters, $\mathbf{R}$. To these parameter groups we apply the Griddy-Gibbs sampler of Ritter and Tanner (1992), as in Barnard et al. (2000). Such a form of the sampling algorithm easily allows to restrict any of the parameters. Note that the algorithm of Barnard et al. (2000) has not yet been applied to Markov-switching models. Finally, we restrict the matrix of transition probabilities, $\mathbf{P}$, joining the approach of Sims et al. (2008) with the Metropolis-Hastings algorithm of Frühwirth-Schnatter (2006). The Metropolis-Hastings step needs to be implemented, as we require the hidden Markov process to be irreducible. Moreover,

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

additional parts of the algorithm are constructed in order to impose nonlinear restrictions on the parameters of the VAR process and the decomposed covariance matrix.

To summarize, we propose the following procedure in order to test different Granger noncausality hypotheses in Markov-switching VAR models.

Step 1: Specify the MS-VAR model. Choose the order of VAR process, $p \in \{0, 1, \ldots, p_{\max}\}$, and the number of states, $M \in \{1, \ldots, M_{\max}\}$, using marginal densities of data (estimation of all the models is required).

Step 2: Set the restrictions. For the chosen model, derive restrictions on parameters.

Step 3: Test restrictions (A1) and (A2). Estimate the restricted models and compute for them marginal densities of data. Compare the restricted models to the the unrestricted one using the Posterior Odds Ratio, e.g. according to the scale proposed by Kass and Raftery (1995).

Step 4: Test hypotheses of noncausality. If the model restricted according to (A1) is preferred to the unrestricted model, then noncausality of all kinds is established. In the other case, if the model restricted according to (A2) is preferred to the unrestricted model, in order to test different noncausality hypotheses use conditions (A6)–(A7). In the opposite case use conditions (A3)–(A5). For testing, use the Posterior Odds Ratio as in Step 3.

Advantages and costs of the proposed approach We start by naming the main advantages of the proposed Bayesian approach to testing the restrictions for Granger noncausality. First, using the Posterior Odds Ratio testing principle, we avoid all the problems of testing nonlinear restrictions on the parameters of the model that appear in classical tests. Secondly, in the context of the controversies concerning the choice of number of states for Markov-switching models in the classical approach (see Psaradakis and Spagnolo, 2003; Psaradakis and Sola, 1998), the Bayesian model selection proposed in Step 1 is a proper method free of

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

such problems. Next, as emphasized in Hoogerheide et al. (2009), the Bayesian Posterior Odds Ratio procedure gives arguments *in favour of* hypotheses. Accordingly, the hypothesis preferred by the data is not only *rejected* or *not rejected*, but is actually *accepted* with some probability. Finally, Bayesian estimation is a basic estimation procedure proposed for the MS-VAR models and is broadly described and used in many applied publications.

However, this approach has also its costs. First of all, in order to specify the complete model, prior distributions for the parameters of the model and the prior probabilities of models need to be specified. This necessity gives way to subjective interpretation of the inference, on the one hand, but on the others it may ensure economic interpretation of the model. The other cost of the implementation of the Bayesian approach is the time required for simulation of all the models, first in the model selection procedure, and second in testing the restrictions of the parameters.

3.5 The Block Metropolis-Hastings sampler for restricted MS-VAR models

This section scrutinizes the MCMC sampler set up for sampling from the full conditional distributions. Each step describes the full conditional distribution of one element of the partitioned parameter vector. The parameter vector is broken up into five blocks: the vector of the latent states of the economy **S**, the transition probabilities **P**, the regime-dependent covariance matrices (themselves decomposed into standard deviations σ and correlations **R**), and finally the regime-dependent vector of constants plus autoregressive parameters β. For each block of parameters – conditionally on the parameter draws from the four other blocks – we describe how we sample from the posterior distribution. The symbols, l and $l-1$, refer to the iteration of the MCMC sampler. For the first iteration of a MCMC run, $l = 1$, initial parameter values come from an EM algorithm. The rest of this section describes all the constituting blocks that form the MCMC sampler.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

3.5.1 Sampling the vector of the states of the economy

The first drawn parameter is the vector representing the states of the economy, $\mathbf{S}$. Being a latent variable, there are no priors nor restrictions on $\mathbf{S}$. We first use a filter (see Section 11.2 of Frühwirth-Schnatter, 2006, and references therein) and obtain the probabilities $Pr(s_t = i|\mathbf{y}, \theta^{(l-1)})$, for $t = 1, \ldots, T$ and $i = 1, \ldots, M$, and then draw $\mathbf{S}^{(l)}$, for l^{th} iteration of the algorithm. For the full description of the algorithm used in this work the reader is referred to Droumaguet and Woźniak (2012).

3.5.2 Sampling the transition probabilities

In this step of the MCMC sampler, we draw from the posterior distribution of the transition probabilities matrix, conditioning on the states drawn in the previous step of the current iteration, $\mathbf{P}^{(l)} \sim p(\mathbf{P}|\mathbf{S}^{(l)})$. For the purpose of testing, we impose restrictions of identical rows of $\mathbf{P}$. Sims et al. (2008) provide a flexible analytical framework for working with restricted transition probabilities, and the reader is invited to consult Section 3 of that work for an exhaustive description of the possibilities provided by the framework. We however limit the latitude given by the reparametrization in order to ensure the stationarity of Markov chain $\mathbf{S}$.

Reparametrization The transitions probabilities matrix $\mathbf{P}$ is modeled with Q vectors w_j, $j = 1, \cdots, Q$ and each of size d_j. Let all the elements of w_j belong to the $(0, 1)$ interval and sum up to one, and stack all of them into the column vector $\mathbf{w} = (w_1', \ldots, w_Q')'$ of dimension $d = \sum_{j=1}^{Q} d_j$. Writing $p = vec(\mathbf{P}')$ as a M^2 dimensional column vector, and introducing the $(M^2 \times d)$ matrix $\mathbf{M}$, the transition matrix is decomposed as:

$$p = \mathbf{M}\mathbf{w}, \tag{3.27}$$

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

where the $\mathbf{M}$ matrix is composed of the M_{ij} sub-matrices of dimension $(M \times d_j)$, where $i = 1, \dots, M$, and $j = 1, \dots, Q$:

$$\mathbf{M} = \begin{bmatrix} M_{11} & \dots & M_{1Q} \\ \vdots & \ddots & \\ M_{M1} & & M_{MQ} \end{bmatrix},$$

where each M_{ij} satisfies the following conditions:

1. For each (i, j), all elements of M_{ij} are non-negative.
2. $\imath'_M M_{ij} = \Lambda_{ij}\imath'_{d_j}$, where Λ_{ij} is the sum of the elements in any column of M_{ij}.
3. Each row of $\mathbf{M}$ has, at most, one non-zero element.
4. M is such that $\mathbf{P}$ is irreducible: for all $j, d_j \geq 2$.

The first three conditions are inherited from Sims et al. (2008), whereas the last condition assures that $\mathbf{P}$ is irreducible, forbidding the presence of an absorbing state that would render the Markov chain $\mathbf{S}$ non-stationary. The non-independence of the rows of $\mathbf{P}$ is described in Frühwirth-Schnatter (2006, Section 11.5.5). Once the initial state s_0 is drawn from the ergodic distribution π of $\mathbf{P}$, direct MCMC sampling from the conditional posterior distribution becomes impossible. However, a Metropolis-Hastings algorithm can be set up to circumvent this issue, since a kernel of joint posterior density of all rows is known: $p(\mathbf{P}|\mathbf{S}) \propto \prod_{j=1}^{Q} \mathcal{D}_{d_j}(w_j)\pi$. Hence, the proposal for transition probabilities is obtained by sampling each w_j from the convenient Dirichlet distribution. The priors for w_j follow a Dirichlet distribution, $w_j \sim \mathcal{D}_{d_j}(b_{1,j}, \dots, b_{d_j,j})$. We then transform the column vector $\mathbf{w}$ into our candidate matrix of transitions probabilities using equation (3.27). Finally, we compute the acceptance rate before retaining or discarding the draw.

Algorithm 1. *Metropolis-Hastings for the restricted transition matrix.*

1. $s_0 \sim \pi$. The initial state is drawn from the ergodic distribution of $\mathbf{P}$.

2. $w_j \sim \mathcal{D}_{d_j}(n_{1,j} + b_{1,j}, \ldots, n_{d_j,j} + b_{d_j,j})$ for $j = 1, \ldots, Q$. $n_{i,j}$ corresponds to the number of transitions from state i to state j, counted from $\mathbf{S}$. The candidate transition probabilities matrix – in the transformed notation – are sampled from a Dirichlet distribution.

3. $\mathbf{P}^{new} = \mathbf{Mw}$. The proposal for the transitions probabilities matrix is reconstructed.

4. Accept $\mathbf{P}^{new}$ if $u \leq \frac{\pi^{new}}{\pi^{l-1}}$, where $u \sim \mathcal{U}[0,1]$. π^{new} and π^{l-1} are the vectors of the ergodic probabilities resulting from the draws of the transition probabilities matrix $\mathbf{P}^{new}$ and $\mathbf{P}^{l-1}$ respectively.

3.5.3 Sampling a second and independent hidden Markov process

Regime inference from proposition (A1) involves two independent Markov processes. Equation (3.18) decomposes the vector of observations into two subvectors. Equations contained within each sub-vector are subject to switches from a different and independent Markov process. Sims et al. (2008, section 3.3.3) cover a similar decomposition.

Adding a Markov process is trivial in the sense it involves repeating the steps of Section 3.5.1 and algorithm 1 subsequently for a second process, yielding two distinct transition probabilities matrices $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$. The transition probabilities matrix for the whole system is formed out of the transition probabilities matrices of two independent hidden Markov processes, $\mathbf{P} = (\mathbf{P}^{(1)} \otimes \mathbf{P}^{(2)})$.

3.5.4 Sampling the covariance matrices

Adapting the approach proposed by Barnard et al. (2000) to Markov-switching models, we sample from the full conditional distribution of non-restricted and restricted covariance matrices. We thus decompose each covariance matrix of the MSIAH-VAR process into a vector of standard deviations (σ_{s_t}) and a correlation

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

matrix ($\mathbf{R}_{s_t}$) using the equality:

$$\Sigma_{s_t} = \text{diag}(\sigma_{s_t})\mathbf{R}_{s_t}\text{diag}(\sigma_{s_t}).$$

This decomposition – statistically motivated – enables the partition of the covariance matrix parameters into two categories that are well suited for the restrictions we want to impose on the matrices. In a standard covariance matrix, restricting a variance parameter to some value has some impact on the depending covariances, whereas here variances and covariances (correlations) are treated as separate entities. The second and not the least advantage of the approach of Barnard et al. (2000) lies in the employed estimation procedure, the griddy-Gibbs sampler. The method introduced in Ritter and Tanner (1992) is well suited for sampling from an unknown univariate density $p(\mathbf{X}_i|\mathbf{X}_j, i \neq j)$. This is done by approximating the inverse conditional density function, which is done by evaluating $p(\mathbf{X}_i|\mathbf{X}_j, i \neq j)$ thanks to a grid of points. Imposing the desired restrictions on the parameters, and afterwards iterating a sampler for every standard deviation $\sigma_{i.s_t}$ and every correlation $\mathbf{R}_{j.s_t}$, we are able to simulate desired posteriors of the covariance matrices. While adding to the overall computational burden, the griddy-Gibbs sampler gives us full latitude to estimate restricted covariance matrices of the desired form.

Algorithm 2. *Griddy-Gibbs for the standard deviations.* The algorithm iterates on all the standard deviation parameters $\sigma_{i.s_t}$ for $i = 1, \ldots, N$ and $s_t = 1, \ldots, M$. Similarly to Barnard et al. (2000) we assume log-normal priors, $\log(\sigma_{i.s_t}) \sim \mathcal{N}(0, 2)$. The grid is centered on the residuals' sample standard deviation $\hat{\sigma}_{i.s_t}$ and divides the interval $(\hat{\sigma}_{i.s_t} - 2\hat{\sigma}_{\hat{\sigma}_{i.s_t}}, \hat{\sigma}_{i.s_t} + 2\hat{\sigma}_{\hat{\sigma}_{i.s_t}})$ into G grid points. $\hat{\sigma}_{\hat{\sigma}_{i.s_t}}$ is an estimator of the standard error of the estimator of the sample standard deviation.

1. Regime-invariant standard deviations: Draw from the unknown univariate density $p(\sigma_i|\mathbf{y}, \mathbf{S}, \mathbf{P}, \beta, \sigma_{-i}, \mathbf{R})$. This is done by evaluating a kernel on a grid of points, using the proportionality relation, with the likelihood function times the prior: $\sigma_i|\mathbf{y}, \mathbf{S}, \mathbf{P}, \beta, \sigma_{-i}, \mathbf{R} \propto p(\mathbf{y}|\mathbf{S}, \theta) \cdot p(\sigma_i)$. Reconstruct the c.d.f. from the grid through deterministic integration and sample from it.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

2. Regime-varying standard deviations: For all regimes $s_t = 1, \dots, M$, draw from the univariate density $p(\sigma_{i.s_t}|\mathbf{y}, \mathbf{S}, \mathbf{P}, \beta, \sigma_{-i.s_t}, \mathbf{R})$, evaluating a kernel thanks to the proportionality relation, with the likelihood function times the prior: $\sigma_{i.s_t}|\mathbf{y}, \mathbf{S}, \mathbf{P}, \beta, \sigma_{-i.s_t}, \mathbf{R} \propto p(\mathbf{y}|\mathbf{S}, \theta) \cdot p(\sigma_{i.s_t})$.

Algorithm 3. *Griddy-Gibbs for the correlations* The algorithm iterates on all the correlation parameters $\mathbf{R}_{i.s_t}$ for $i = 1, \dots, \frac{(N-1)N}{2}$ and $s_t = 1, \dots, M$. Similarly to Barnard et al. (2000), we assume uniform distribution on the feasible set of correlations, $\mathbf{R}_{i.s_t} \sim \mathcal{U}(a, b)$, with a and b being the bounds that keep the implied covariance matrix positive definite; see the aforementioned reference for details of setting a and b. The grid divides (a, b) into G grid points.

1. Depending on the restriction scheme, set correlations parameters to 0.

2. Regime-invariant correlations: Draw from the univariate density $p(\mathbf{R}_i|\mathbf{y}, \mathbf{S}, \mathbf{P}, \beta, \sigma, \mathbf{R}_{-i})$, evaluating a kernel thanks to the proportionality relation, with the likelihood function times the prior: $\mathbf{R}_i|\mathbf{y}, \mathbf{S}, \mathbf{P}, \beta, \sigma, \mathbf{R}_{-i} \propto p(\mathbf{y}|\mathbf{S}, \theta) \cdot p(\mathbf{R}_i)$.

3. Regime-varying correlations: For all regimes $s_t = 1, \dots, M$, draw from the univariate density $p(\mathbf{R}_{i.s_t}|\mathbf{y}, \mathbf{S}, \mathbf{P}, \beta, \sigma, \mathbf{R}_{-i.s_t})$, evaluating a kernel thanks to the proportionality relation, with the likelihood function times the prior: $\mathbf{R}_{i.s_t}|\mathbf{y}, \mathbf{S}, \mathbf{P}, \beta, \sigma, \mathbf{R}_{-\mathbf{R}_{i.s_t}} \propto p(\mathbf{y}|\mathbf{S}, \theta) \cdot p(\mathbf{R}_{i.s_t})$.

3.5.5 Sampling the vector autoregressive parameters

Finally, we draw the state-dependent autoregressive parameters, β_{s_t} for $s_t = 1, \dots, M$. The Bayesian parameter estimation of finite mixtures of regression models when the realizations of states is known has been precisely covered in Frühwirth-Schnatter (2006, Section 8.4.3). The procedure consists of estimating all the regression coefficients simultaneously by stacking them into $\beta = (\beta_0, \beta_1, \dots, \beta_M)$, where β_0 is a common regression parameter for each regime, and hence is useful for the imposing of restrictions of state invariance for the autoregressive parameters.

 DOI: 10.2870/63858

The regression model becomes:

$$y_t = Z_t\beta_0 + Z_t D_{i.1}\beta_1 + \cdots + Z_t D_{i.M}\beta_M + \epsilon_t, \tag{3.28}$$

$$\epsilon_t \sim i.i.\mathcal{N}(\mathbf{0}, \Sigma_{s_t}). \tag{3.29}$$

We have here introduced the $D_{i.s_t}$, which are M dummies taking the value 1 when the regime occurs and set to 0 otherwise. A transformation of the regressors Z_T also has to be performed in order to allow for different coefficients on the dependent variables, for instance to impose zero restrictions on parameters. In the context of VARs, Koop and Korobilis (2010, Section 2.2.3) detail a convenient notation that stacks all the regression coefficients on a diagonal matrix for every equation. We adapt this notation by stacking all the regression coefficients for all the states on diagonal matrix. If $z_{n.s_t.t}$ corresponds to the row vector of $1 + Np$ independent variables for equation n, state s_t (starting at 0 for regime-invariant parameters), and at time t, the stacked regressor Z_t will be of the following form:

$$Z_t = \text{diag}(z_{1.0.t}, \ldots, z_{N.0.t}, z_{1.1.t}, \ldots, z_{N.1.t}, \ldots, z_{1.M.t}, \ldots, z_{N.M.t}).$$

This notation enables the restriction of each parameter, by simply setting $z_{n.s_t.t}$ to 0 where desired.

Algorithm 4. *Sampling the autoregressive parameters.* We assume normal prior for β, i.e. $\beta \sim \mathcal{N}(\mathbf{0}, \underline{V_\beta})$.

1. For all Z_ts, impose restrictions by setting $z_{n,s_t,t}$ to zero accordingly.

2. $\beta|\mathbf{y}, \mathbf{S}, \mathbf{P}, \sigma, \mathbf{R} \sim \mathcal{N}(\overline{\beta}, \overline{V_\beta})$. Sample β from the conditional normal posterior distribution, with the following parameters:

$$\overline{V_\beta} = \left(\underline{V_\beta^{-1}} + \sum_{t=1}^{T} Z_t' \Sigma_{s_t}^{-1} Z_t \right)^{-1}$$

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

and

$$\overline{\beta} = \overline{V_\beta}\left(\sum_{t=1}^{T} Z_t' \Sigma_{s_t}^{-1} y_t\right).$$

3.5.6 Simulating restrictions in the form of functions of the parameters.

Some of the restrictions for Granger noncausality presented in Section 3.3 will be in the form of complicated functions of parameters. Suppose some restriction is in the form:

$$\theta_i = g(\theta_{-i}),$$

where $g(.)$ is a scalar function of all the parameters of the model but θ_i. The restricted parameter, θ_i, in this study may be one of the parameters from the autoregressive parameters, β, or standard deviations, σ. In such a case, β or σ are no longer conditionally independent and need to be simulated with a Metropolis-Hastings algorithm.

Restriction on the vector autoregressive parameters β In this case, the deterministic function restricting parameter β_i will be of the following form:

$$\beta_i = g(\beta_{-i}, \sigma, \mathbf{R}, \mathbf{P}).$$

We draw from the full conditional distribution of the vector autoregressive parameters, $p(\beta|\mathbf{y}, \mathbf{S}, \mathbf{P}, \sigma, \mathbf{R})$, using the Metropolis-Hastings algorithm:

Algorithm 5. *Metropolis-Hastings for the restricted vector autoregressive parameters β.*

1. Form a candidate draw, β^{new}, using Algorithm 6.

2. Compute the probability of acceptance of a draw:

$$\alpha(\beta^{l-1}, \beta^{new}) = \min\left[\frac{p(\mathbf{y}|\mathbf{S}, \mathbf{P}, \beta^{new}, \sigma, \mathbf{R})p(\beta^{new})}{p(\mathbf{y}|\mathbf{S}, \mathbf{P}, \beta^{l-1}, \sigma, \mathbf{R})p(\beta^{l-1})}, 1\right]. \tag{3.30}$$

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

3. Accept β^{new} if $u \leq \alpha(\beta^{l-1}, \beta^{new})$, where $u \sim \mathcal{U}[0, 1]$.

The algorithm has its justification in the block Metropolis-Hastings algorithm of Greenberg and Chib (1995). The formula for computing the acceptance probability from equation (3.30) is a consequence of the choice of the candidate generating distributions. For the parameters β_{-i}, it is a symmetric normal distribution, as in step 2 of Algorithm 4, whereas β_i is determined by a deterministic function.

Algorithm 6. *Generating a candidate draw β.*

1. Restrict parameter β_i to zero. Draw all the parameters $(\beta_1, \dots, \beta_{i-1}, \mathbf{0}, \beta_{i+1}, \dots, \beta_k)'$ according to the algorithms described in Section 3.5.5.

2. Compute $\beta_i = g(\beta_{-i}, \sigma, \mathbf{R}, \mathbf{P})$.

3. Return the vector $(\beta_1, \dots, \beta_{i-1}, \mathbf{g}(\beta_{-\mathbf{i}}, \sigma, \mathbf{R}, \mathbf{P}), \beta_{i+1}, \dots, \beta_k)'$

3.6 Granger causal analysis of US money-income data

In both studies focusing on Granger causality analysis within Markov-switching vector autoregressive models, Warne (2000) and Psaradakis et al. (2005), the focus of study is the causality relationship between U.S. money and income. At the heart of this issue is the empirical analysis conducted in Friedman and Schwartz (1971) asserting that money changes led income changes. The methodology was rejected by Tobin (1970) as a *post hoc ergo propter hoc* fallacy, arguing that the timing implications from money to income could be generated not only by monetarists' macroeconomic models but also by Keynesian models. Sims (1972) initiated the econometric analysis of the causal relationship from the Granger causality perspective. While a Granger causality study concentrates on forecasting outcomes, macroeconomic theoretical modeling tries to remove the question mark over the

The total US economic activity is approached from two different perspectives in these papers: Warne (2000) uses monthly income data, whereas Psaradakis et al. (2005) use quarterly output data.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

neutrality of monetary policy for the business cycle. The causal relationship between money and income is, however, of particular interest to the econometric debate, since over the past forty years researchers have not reached a consensus.

This historical debate between econometricians is well narrated by Psaradakis et al. (2005), and the interested reader is advised to consult this paper for a depiction of events. Without detailing the references of the aforementioned paper, there is a problem in the instability of the empirical results found for the causality between money and output. Depending on the samples considered (postwar onwards data, 1970s onwards data, 1980s onwards, 1980s excluded, etc.), the existence and intensity of the causal effect of money on output are subject to different conclusions. Hence, the strategy of Psaradakis et al. (2005): to set up a Markov-switching VAR model in which the parameters responsible for noncausality in VAR models are subject to regime switches, with some regimes in which they are set to zero (noncausality for VARs) and others in which they are allowed to be different from zero. MS-VAR models are convenient tools because the switches in regimes are endogenous and can occur as many times as the data impose.

As outlined in the introduction, with the approach of Warne (2000) which we follow, the MS-VAR models are 'standard' ones, and we perform Bayesian model selection through the comparison of their marginal densities of data, to determine the number of states as well as the number of autoregressive lags. Moreover, we perform an analysis with precisely stated definitions of Granger causality for Markov-switching models. In this section, we use the Bayesian testing apparatus to investigate this relationship once again.

Data The data are identical to those estimated by Warne (2000) and cover the same time period as in the original paper. Two monthly series are included, the US money stock M1 and the industrial production, both containing 434 observations covering the period, from 1959:1 to 1995:2, and both were extracted from the Citibase database. As in the original paper, the data are seasonally adjusted, transformed into log levels, and multiplied by 1200. Warne (2000) performed Johansen tests for cointegration, and – unlike for level series – trace statistics

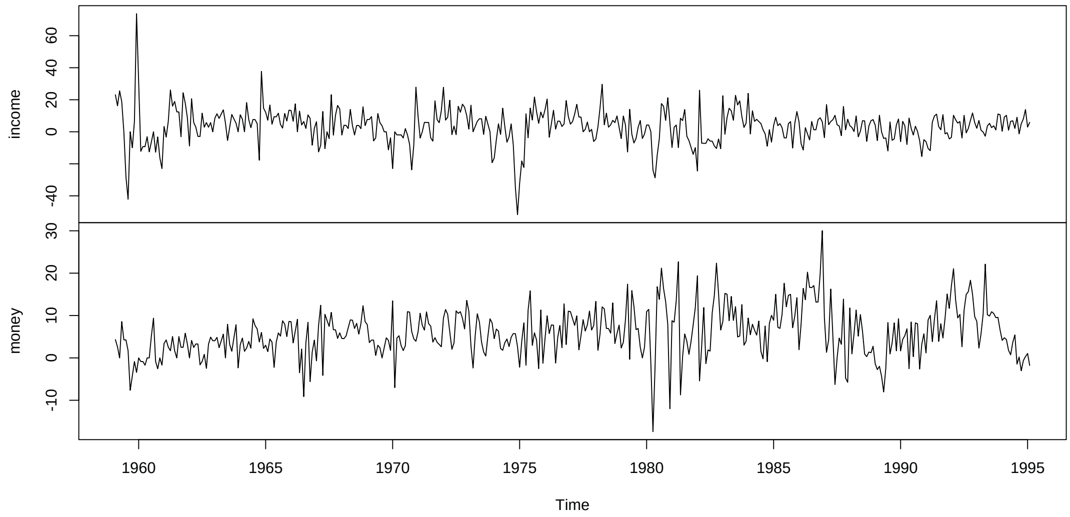


Figure 3.1: Log-differentiated series of money and income.

Table 3.1: Summary statistics

Variable	Mean	Median	Standard Deviation	Minimum	Maximum
Δy	3.396	4.18	10.99	-51.73	73.72
Δm	5.851	5.24	5.79	-17.39	30.03

Data Source: Citibase.

indicated no cointegration for differentiated series. Similarly, we work with the first difference of the series.

The summary statistics of both series are presented in Table 3.1. Income grows yearly by 3% on average, with a standard deviation of 11%, which seems a lot, but one has to note that we manipulate the monthly series for which the rates are annualized. Money has a stronger growth rate of nearly 6% on average, with a lower standard deviation than the income, below 6%.

Figure 3.1 plots the transformed series. Observation indicates that at least some heteroskedasticity is present, as can be seen with the money series, where a period

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

of higher volatility starts around 1980. Summary statistics and series observations all seem to indicate the possibility of different states in the series, in which case MS-VAR models can provide a useful framework for analysis. We, however, start our analysis with Granger causality testing in the context of linear VAR models.

Granger causal analysis with VAR model The reason why we begin by studying Granger causality with linear models is that we want to relate to the standard methodology, and to illustrate whether a non-linear approach brings added value to the analysis by comparing the results. Also, the Block Metropolis-Hastings sampler of Section 3.5 can easily be simplified to a Block Metropolis-Hastings sampler for VAR models. By doing so, estimating linear VAR models and comparing marginal densities, we will also compare whether or not these models are preferred by the data to more complex MS-VAR ones.

We estimate the data with the VAR models for different lag lengths, $p = 0, \ldots, 17$. Each of the Metropolis-Hastings algorithms is initiated by the OLS estimates of the VAR coefficients. Then follows a 10,000-iteration burn-in and, after convergence of the sampler, 5000 final draws are to constitute the posteriors. The prior distributions are as follow:

$$\begin{aligned} \beta_i &\sim \mathcal{N}(\mathbf{0}, 100 I_{N+pN^2}) \\ \sigma_{i.j} &\sim log\mathcal{N}(0, 2) \\ \mathbf{R}_{ij} &\sim \mathcal{U}(a, b) \end{aligned}$$

for $i = 1, \ldots, M$ and $j = 1, \ldots, N$.

Table 3.2 displays the marginal density of data for each model, computed with the modified harmonic mean obtained by applying formula (3.26) to the posteriors draws. As in Warne (2000), models with long lags are preferred. The VAR(12) model, i.e. with 12 lags for the autoregressive coefficients, yields the highest lnMHM and hence is the model we choose for the Granger causality analysis. Table C.1 in C.2 displays, for each parameter of the model, the mean, standard deviations, naive standard errors, autocorrelations of the Markov Chain at lag 1

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 3.2: Model selection for VAR(p) – determination of number of lags

Lags	0	1	2	3	4	5	6	7	8
lnMHM	-3149.63	-2991.7	-2983.4	-2966.49	-2970.25	-2954.49	-2948.57	-2944	-2939.52
Lags	9	10	11	12	13	14	15	16	17
lnMHM	-2936.67	-2941.2	-2917.97	**-2916.77**	-2917.87	-2926.21	-2923.23	-2930.82	-2936.96

and lag 10. Low autocorrelation at lag 10 indicates that the sampler has good properties.

The set of restrictions to impose on the parameters for vector autoregressive moving average models were covered in Sims (1972) and Boudjellaba et al. (1992). Translated into the VAR representation, and in the case of a bivariate VAR(p) model:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} A_{11}^{(i)} & A_{12}^{(i)} \\ A_{21}^{(i)} & A_{22}^{(i)} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix},$$

for $t = 1, \dots, T$, the restrictions for money, $y_{2,t}$, being Granger noncausal on income, $y_{1,t}$, read:

$$A_{12}^{(i)} = 0 \text{ for } i = 1, \dots, p.$$

Note that these restrictions, with assumed normal residual terms, are simultaneously encompassing Granger noncausality in mean, variance, and distribution.

The estimation of the restricted VAR(12) model, with its upper-right autoregressive coefficients $A_{12}^{(i)}$ set to 0 for all lags returns posteriors that yield a lnMHM of -2901.63. Expressed in logarithms, the posterior odds ratio of the null hypothesis of Granger causality from money to income is equal to 15.13. Table 3.3 summarizes the results for VAR models. This is a very strong acceptance of the restricted model $\mathcal{M}_1$ over the nonrestricted one $\mathcal{M}_0$, hence Bayesian testing provides evidence in favor of Granger noncausality from money to income, within the VAR framework. This result is in line with Christiano and Ljungqvist (1988), where Granger noncausality from money to output is established for the VAR model with log-differences with US data. The authors contest this result and argue for a specification error for models with first differences. We continue our analysis with nonlinear models that allow switches within their parameters.

Table 3.3: Noncausality and conditional regime independence in a VAR(12) model. Numerical efficiency results for these models are presented in table C.3 of C.3.

$\mathcal{M}_j$	Hypothesis	Restrictions	# restrictions	$\ln p(\mathbf{y}\vert\mathcal{M}_j)$	$\ln \mathcal{B}_{j0}$
	$\mathcal{H}_0$: *Unrestricted model*				
$\mathcal{M}_0$	VAR(12)	-	0	-2,916.77	0
	$\mathcal{H}_1$: *Granger noncausality from money to income*				
$\mathcal{M}_1$	(A1)	$A_{12}^{(i)} = 0$	p	-2,901.63	15.13

for $i = 1, \dots, p$.

Granger causal analysis with MS-VARs MS-VAR models capture the nonlinearities of the data, such as heteroskedasticity. Endogeneity in the regime estimation gives lots of latitude for the capture of a variety of nonlinear features of the data, hence in a way reducing the risk of model misspecification. The legitimacy of these models against VARs can easily be tested through the computation of the marginal distribution of data for the respective models.

Moreover, the Markov-switching models, framework provides a more detailed analysis of causality, as MS-VAR models produce different sets of restrictions for different types of noncausality, i.e. noncausality in mean, variance, or distribution. Therefore, we distinguish between more and less strict hypotheses, and make inferences that are more informative by investigating causality in moments of different order.

We estimate the data MSIAH(m)-VAR(p) models for different number of regimes $m = 2, 3, 4$ and different lag lengths, $p = 0, \dots, 6$. Each of the Gibbs algorithm is initiated by the estimates from the EM algorithm of the corresponding model. Then follows a 10,000-iteration burn-in and, after convergence of the sampler, we

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 3.4: Model selection for MSIAH(2)-VAR(p) – determination of the lag order

Lags	0	1	2	3	4	5	6
lnMHM	-3,002.64	-2,926.42	-2,903.89	-2,898.21	**-2,895.22**	-2,914.87	-2,913.49

sample 5000 final draws from the posteriors. The prior distributions are as defined in Section 3.2.

Table 3.4 reports the lnMHMs for the estimated models with 2 regimes. Though we also estimated models with 3 or 4 regimes, estimation encountered difficulties of low occurrences of regimes. These phenomena indicate that the data does not support MS-VAR models with 3 or more regimes, and explains why we only present results with 2 regimes. The number of estimated lags for the autoregressive coefficients is limited to 6 lags – less than the 12 lags for VAR models – also due to insufficient state occurrences when the number of AR parameters increases. The model preferred by the data is the MSIAH(2)-VAR(4), i.e. with 2 regimes and VAR process of order 4. Table C.2 in C.2 displays, for each parameter of the model, the mean, standard deviations, naive standard errors, and autocorrelations of the Markov chains at lag 1 and lag 10. Decaying autocorrelation between draws indicates that the sampler has desirable properties.

Figure 3.2 plots the regime probabilities from the selected model. In comparison with the second regime, the first regime matches times of higher variance for both variables. As well the constant for income growth, $\mu_{1,1}$, is negative during the occurrences of the first regime. Hence, the first regime can be interpreted as the bad state of the economy.

Note that comparing the best unrestricted MS-VAR model from Table 3.4 to the best VAR model of Table 3.3 (that is to the restricted model) yields a logarithm of the posterior odds ratio of 6.41 in favor of the MS-VAR model.

Similarly to Warne (2000), we proceed with the analysis of Granger noncausality for the selected MSIAH(2)-VAR(4) model. The Bayesian testing strategy we employ renders the process straightforward: each type of causality implies differ-

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

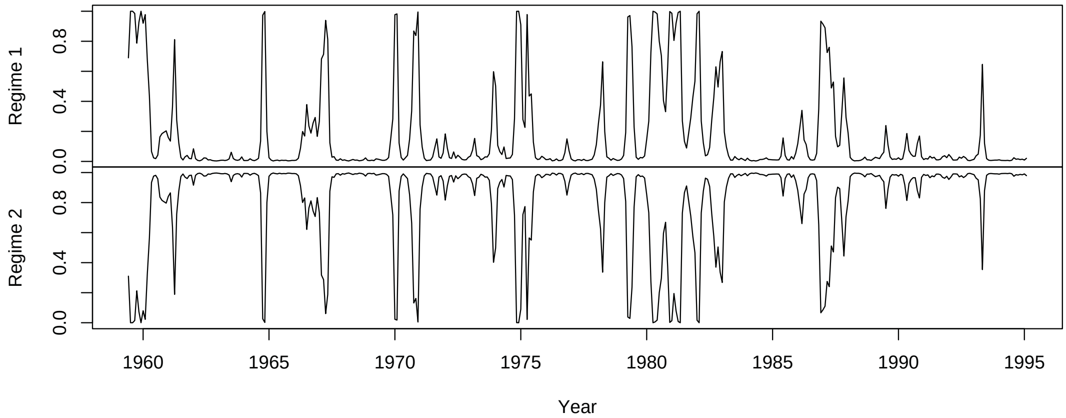


Figure 3.2: Estimated probabilities of regimes for a MSIAH(2)-VAR(4) model

ent restrictions on the model parameters; we impose them, estimate the models and compute all marginal densities of data. Table 3.5 summarizes all the sets of restrictions to impose when testing the noncausality from money to income, and also logarithms of the marginal densities of data given the model, $\ln p(\mathbf{y}|\mathcal{M}_j)$, and logarithms of the Bayes factors, $\ln \mathcal{B}_{j0}$ for $j = 1, \dots, 7$. A positive logarithm of the Bayes factor is to be interpreted as evidence in favour of the restricted model. In a symmetric way, negative logarithm of the Bayes factor indicates that the non-restricted model is preferred by the data.

Analysis of Table 3.5 shows that only model $\mathcal{M}_5$ is more probable *a posteriori* than the unrestricted model $\mathcal{M}_0$. This model represents one of the sets of restrictions for Granger noncausality in mean. All other models, however, are less probable than the unrestricted model, which is represented with the negative values of the logarithms of the Bayes factors.

Table 3.6 presents a summary of the assessment of the considered hypotheses. We found strong support for Granger noncausality in mean. This hypothesis has much bigger posterior probability compared to all other hypotheses, including the unrestricted model. Warne (2000) found a similar result, but holding only at the 10% level of significance. However, Bayesian testing establishes this strong result,

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 3.5: Noncausality and conditional regime independence in a MSIAH(2)-VAR(4) model. Numerical efficiency results for these models are presented in table C.3 of C.3.

$\mathcal{M}_j$	Hypothesis	Restrictions	# restrictions	$\ln p(\mathbf{y}\|\mathcal{M}_j)$	$\ln \mathcal{B}_{j0}$
	$\mathcal{H}_0$: *Unrestricted model*				
$\mathcal{M}_0$	MS(2)-VAR(4)	-	0	-2895.22	0
	$\mathcal{H}_1$: *History of money does not impact on the regime forecast of income*				
$\mathcal{M}_1$	(A1) $M_1 = 1, M_2 = 2$	$\mu_{1,s_t} = \mu_1, A^{(i)}_{11,s_t} = A^{(i)}_{11}, A^{(i)}_{12,s_t} = 0$ $\Sigma_{11,s_t} = \Sigma_{11}, \Sigma_{12,s_t} = 0$	$3p$+4	-2964.72	-69.50
$\mathcal{M}_2$	(A1) $M_1 = 2, M_2 = 1$	$\mu_{2,s_t} = \mu_2, A^{(i)}_{21,s_t} = A^{(i)}_{21}, A^{(i)}_{22,s_t} = A^{(k)}_{22},$ $\Sigma_{22,s_t} = \Sigma_{22}, \Sigma_{12,s_t} = 0, A^{(i)}_{12,s_t} = 0$	$4p$+4	-2921.54	-26.32
	(A2) $M_1 = 1, M_2 = 2$	Always holds, no restrictions	-	-	-
$\mathcal{M}_3$	(A2) $M_1 = 2, M_2 = 1$	$p_{11} = p_{21}$	1	-2907.39	-12.17
	$\mathcal{H}_2$: *Granger noncausality in mean*				
	(A1) or	-	-	-	-
$\mathcal{M}_4$	(A6) $M_1 = 1, M_2 = 2$	$\mu_{1,s_t} = \mu_1, A^{(i)}_{11,s_t} = A^{(i)}_{11}, A^{(i)}_{12,s_t} = 0$	$3p$+1	-2880.63	14.59
$\mathcal{M}_5$	(A6) $M_1 = 2, M_2 = 1$	$p_{11} = p_{21}, \sum_{j=1}^{2} A^{(i)}_{12,j}\pi_j = 0$	p+1	-2897.24	-2.02
	$\mathcal{H}_3$: *Granger noncausality in variance*				
	(A1) or	-	-	-	-
$\mathcal{M}_6$	(A7) $M_1 = 1, M_2 = 2$	$\mu_{1,s_t} = \mu_1, A^{(i)}_{11,s_t} = A^{(i)}_{11}, A^{(i)}_{12,s_t} = 0$ $\Sigma_{11,s_t} = \Sigma_{11}$	$3p$+2	-2953.15	-57.93
$\mathcal{M}_7$	(A7) $M_1 = 2, M_2 = 1$	$p_{11} = p_{21}, A^{(i)}_{12,s_t} = 0$	$2p$+1	-2900.58	-5.36
	$\mathcal{H}_4$: *Granger noncausality in distribution*				
	(A1) or	-	-	-	-
	(A7)	-	-	-	-

for $i = 1, \ldots, p$.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table 3.6: Summary of the hypotheses testing

$\mathcal{H}_i$	Hypothesis	Represented by models	$\ln \frac{Pr(\mathcal{H}_i\|\mathbf{y})}{Pr(\mathcal{H}_0\|\mathbf{y})}$
$\mathcal{H}_0$	Unrestricted model	$\mathcal{M}_0$	0
$\mathcal{H}_1$	History of money does not impact on the regime forecast of income	$\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_4$	-12.17
$\mathcal{H}_2$	Granger noncausality in mean	$\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_4$, $\mathcal{M}_5$	14.59
$\mathcal{H}_3$	Granger noncausality in variance	$\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_6$, $\mathcal{M}_7$	-5.36
$\mathcal{H}_4$	Granger noncausality in distribution	$\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_6$, $\mathcal{M}_7$	-5.36

and the conditional mean of income is invariant to the history of money. Table 3.6 provides strong evidence for Granger causal relations in variance and, in effect, in distribution, as these two hypotheses for the considered model are represented by the same set of models.

Summary The results of Bayesian testing for Granger causality from money to input on the US monthly series covering the period 1959–1995 are in line with the narration of Psaradakis et al. (2005), in the sense that the strongly established noncausality in mean within VAR models (which is equivalent to the noncausality in variance and in distribution) does not hold with MS-VAR models. Allowing nonlinearity in the models' coefficients, here by a Markov chain permitting switches between regimes of the economy, and testing for causality from money to income yields a different result and the strong noncausal evidence is decomposed. We found that the history of money helps to predict the regimes of income. We also found that money causes income both in variance and in distribution. However, we did find evidence for Granger noncausality in mean from money to income, as did Warne (2000). Bayesian model estimation associated with Bayesian testing provided tools with which to select the correct model specification, and also with which to compare it to the VAR specifications, and the posterior odds ratio tests allowed us to test for the three types of Granger noncausality.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

These findings have particular consequences for the forecasting of the income. Despite the fact that past information about money does not change the forecast of the conditional mean of income, it is still crucial for its modeling. Past observations of money improves the forecast of the state of the economy when modeled with a Markov-switching process. Therefore, if one is interested in forecasting regime switches in the income equation, then one should add the money variable into the considered system. The same conclusion applies to the forecasting of future variability of income and, in particular, for its density forecast. The last finding is especially relevant for the Bayesian Markov-switching vector autoregressions. We justify this statement with two features of such a model. First, Markov-switching vector autoregressions are designed to model and forecast a complicated distribution of the residuals with heteroskedastic variances and non-normal distribution. Second, the Bayesian inference is particularly suitable for the density forecast with MS-VARs, due to the fact that the predictive density is constructed by integrating out the parameters of the models. In consequence, the forecast incorporates the uncertainty with respect to the parameter values. Moreover, the integration required in order to construct the forecasts conditioned only on past observations of the variables, and not conditioned on the unobserved states, as in classical forecasting (see Hamilton, 1994), is straightforward.

A note Using Bayes factors for the comparison of the models is not uncontroversial. It appears that Bayes factors are sensitive to the specification of the prior distributions for the parameters being tested. The more diffuse a prior distribution the more informative it is about the the parameter tested with a Bayes factor. This phenomenon is called Bartlett's paradox (see Bartlett, 1957) and is a version of the Lindley's paradox. Moreover, Strachan and van Dijk (2011) show that assuming a diffuse prior distribution for the parameters of the model, results in wrongly defined Bayes factors. As a solution to this problem Strachan and van Dijk recommend using a prior distribution belonging to a class of shrinkage distributions.

In this study, normal prior densities with mean zero and variance equal to

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

100 are assumed. This prior distribution for the VAR parameters belongs to a class of diffuse prior distributions. Therefore, the critique of Bayes factors applies. The problem is recognized and solved by employment of the shrinkage prior distribution for these parameters in the newest version of this work (see Droumaguet et al., 2012). However, I do not include this results in this work.

3.7 Conclusions

We proposed a method of testing the nonlinear restrictions for the hypotheses of Granger noncausality in mean, in variance and in distribution for Markov-switching Vector Autoregressions. The employed Bayes factors and Posterior Odds Ratios overcome the limitations of the classical approach. It requires, however, an algorithm of estimation of the unrestricted model and of the restricted models, representing hypotheses of interest. The algorithm we proposed, allows for the restriction of all the groups of parameters of the model in an appropriate way. It combines several existing algorithms and improves them in order to maintain the desired properties of the model and the efficiency of estimation. The estimation method allows us to use many of the existing methods of computing of the marginal density of data that are required for both Bayes factors and Posterior Odds Ratios.

The Bayesian approach to testing has also consequences for the way in which the competing hypotheses are treated. Contrary to classical tests, the hypotheses of Granger causality or noncausality of different types are, in our approach, treated symmetrically. We obtain this effect by comparing the posterior probabilities of the hypotheses (or models). In consequence, the output of our inference, in the form of choosing the hypothesis of the highest posterior probability, reflects the choice of the hypothesis supported in the biggest rate by the data. This applies, of course, to cases in which the chosen prior probabilities and densities do not discriminate *a priori* some of the hypotheses.

In the empirical illustration of the methodology, we have found that in the USA money does not cause income in mean. We have, however, found that the money impacts on the forecast of the future state of the economy, as well as on the forecast of the variability of the income and on its density forecast. If the empirical analysis is to be something more than just an illustration of the methodology, and in effect be conclusive, robustness checks are required. In particular, considering more relevant variables in the system could impact on the conclusions of the analysis of the Granger causality between money and income.

As the main limitation of the whole analysis of Granger causality for MS-VAR models, we find that only *one period ahead* Granger causality is considered in this study. The conditions for *h periods ahead* noncausality should be further explored. We only mention that potentially establishing that one variable does not improve the forecast of the hidden Markov process, taking into account the Markov property, may imply the same for all periods in the future. Still, establishing conditions for the noncausality h periods ahead for the autoregressive parameters, including covariances, would potentially require tedious algebra. This statement is motivated by the complexity of formulating forecasts with MS-VAR models.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Appendix C

C.1 Alternative restrictions for Granger noncausality

The following restrictions were set by Warne (2000), and are all derived under the condition (A2) and $\text{rank}(\mathbf{P}^{(2)}) = M_2$.

Restriction 5. Suppose that $\mathbf{P} = (\imath_{M_1}\pi^{(1)\prime} \otimes \mathbf{P}^{(2)})$ with $\text{rank}(\mathbf{P}^{(2)}) = M_2$, then condition (A3) is equivalent to:

(A6): (i) $\sum_{j_1=1}^{M_1} m_{1.(j_1,j_2)}\pi_{j_1}^{(1)} = \bar{m}_1$,

(ii) $\sum_{j_1=1}^{M_1} a_{1r.(j_1,j_2)}^{(k)}\pi_{j_1}^{(1)} = \bar{a}_{1r}^{(k)}$, and

(iii) $\bar{a}_{14}^{(k)} = 0$

for all $j_2 \in \{1, \dots, M_2\}$, $r \in \{1, 2, 3\}$, and $k \in \{1, \dots, p\}$,

Restriction 6. Suppose that $\mathbf{P} = (\imath_{M_1}\pi^{(1)\prime} \otimes \mathbf{P}^{(2)})$ with $\text{rank}(\mathbf{P}^{(2)}) = M_2$, then condition (A4) is equivalent to:

(A7): (i) (A3),

(ii) $\sum_{j_1=1}^{M_1} \left[(m_{1.(j_1,j_2)} - \bar{m}_1) \otimes (m_{1.(j_1,j_2)} - \bar{m}_1) \right] \pi_{j_1}^{(1)} = \varsigma_\mu$,

(iii) $\sum_{j_1=1}^{M_1} \left[(a_{1r.(j_1,j_2)}^{(k)} - \bar{a}_{1r}^{(k)}) \otimes (a_{1s.(j_1,j_2)}^{(l)} - \bar{a}_{1s}^{(l)}) \right] \pi_{j_1}^{(1)} = \varsigma_{r.s}^{(k.l)}$,

(iv) $\sum_{j_1=1}^{M_1} \left[(m_{1.(j_1,j_2)} - \bar{m}_1) \otimes (a_{1r.(j_1,j_2)}^{(k)} - \bar{a}_{1r}^{(k)}) \right] \pi_{j_1}^{(1)} = \varsigma_{\mu.r}^{(k)}$,

(v) $\sum_{j_1=1}^{M_1} \sigma_{1.(j_1,j_2)}\pi_{j_1}^{(1)} = \varsigma_\omega$, and

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

(vi) $a_{14.j}^{(k)} = 0$

for all $j \in \{1, \ldots, M\}$, $j_2 \in \{1, \ldots, M_2\}$, $r, s \in \{1, 2, 3\}$, and $k, l \in \{1, \ldots, p\}$

is satisfied.

Restriction 7. Suppose rank($\mathbf{P}$) $\in \{1, M\}$, then y_4 does not Granger-cause in distribution y_1 if and only if it does not Granger-cause y_1 in variance.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

C.2 Summary of the posterior densities simulations

Table C.1: VAR(12): posterior properties

	Mean	Std. dev.	Naive Std. error	Autocorr. lag 1	Autocorr. lag 10
Standard deviations					
σ_1	9.192	0.137	0.002	0.028	0.006
σ_2	4.912	0.095	0.001	0.046	0.002
Correlations					
ρ_1	-0.025	0.058	0.001	0.060	-0.014
Intercepts					
μ_1	-0.004	0.300	0.004	0.001	-0.009
μ_2	0.582	0.266	0.004	-0.011	0.006
Autoregressive coefficients					
$A_{11}^{(1)}$	0.284	0.049	0.001	-0.007	0.005
$A_{12}^{(1)}$	0.138	0.088	0.001	-0.006	-0.028
$A_{21}^{(1)}$	0.027	0.027	0.000	-0.024	-0.016
$A_{22}^{(1)}$	0.361	0.049	0.001	0.020	0.027
$A_{11}^{(2)}$	0.076	0.049	0.001	-0.009	0.014
$A_{12}^{(2)}$	0.108	0.094	0.001	-0.034	-0.014
$A_{21}^{(2)}$	-0.044	0.026	0.000	-0.001	0.012
$A_{22}^{(2)}$	-0.005	0.052	0.001	0.007	-0.001
$A_{11}^{(3)}$	0.068	0.049	0.001	0.002	0.011
$A_{12}^{(3)}$	0.133	0.093	0.001	-0.035	0.009
$A_{21}^{(3)}$	-0.054	0.026	0.000	-0.014	-0.009
$A_{22}^{(3)}$	0.199	0.052	0.001	0.001	-0.001
$A_{11}^{(4)}$	0.085	0.049	0.001	0.004	0.009
$A_{12}^{(4)}$	-0.053	0.092	0.001	-0.014	-0.008
$A_{21}^{(4)}$	-0.024	0.027	0.000	0.012	-0.011
$A_{22}^{(4)}$	-0.106	0.051	0.001	-0.026	0.002
$A_{11}^{(5)}$	-0.054	0.049	0.001	-0.003	-0.010

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

	Mean	Std. dev.	Naive Std. error	Autocorr. lag 1	Autocorr. lag 10
$A_{12}^{(5)}$	0.032	0.094	0.001	-0.019	-0.010
$A_{21}^{(5)}$	0.007	0.026	0.000	0.008	-0.005
$A_{22}^{(5)}$	0.228	0.051	0.001	0.004	0.008
$A_{11}^{(6)}$	0.004	0.047	0.001	0.000	0.009
$A_{12}^{(6)}$	0.106	0.095	0.001	0.009	0.019
$A_{21}^{(6)}$	0.000	0.026	0.000	0.004	0.011
$A_{22}^{(6)}$	0.067	0.052	0.001	0.008	-0.010
$A_{11}^{(7)}$	0.035	0.048	0.001	-0.002	-0.007
$A_{12}^{(7)}$	-0.100	0.095	0.001	-0.008	0.003
$A_{21}^{(7)}$	0.001	0.025	0.000	0.017	-0.002
$A_{22}^{(7)}$	-0.012	0.053	0.001	-0.025	-0.008
$A_{11}^{(8)}$	0.031	0.048	0.001	0.035	-0.017
$A_{12}^{(8)}$	0.056	0.094	0.001	0.005	-0.005
$A_{21}^{(8)}$	0.052	0.025	0.000	-0.015	0.005
$A_{22}^{(8)}$	0.104	0.051	0.001	0.011	0.010
$A_{11}^{(9)}$	0.015	0.048	0.001	-0.016	0.019
$A_{12}^{(9)}$	-0.054	0.093	0.001	0.006	0.004
$A_{21}^{(9)}$	-0.043	0.025	0.000	0.016	-0.004
$A_{22}^{(9)}$	0.181	0.052	0.001	0.023	-0.012
$A_{11}^{(10)}$	0.020	0.047	0.001	0.023	0.020
$A_{12}^{(10)}$	0.008	0.090	0.001	0.007	-0.022
$A_{21}^{(10)}$	-0.008	0.026	0.000	-0.010	-0.005
$A_{22}^{(10)}$	-0.077	0.052	0.001	0.018	-0.012
$A_{11}^{(11)}$	0.008	0.048	0.001	-0.017	0.021
$A_{12}^{(11)}$	-0.064	0.093	0.001	-0.014	0.001
$A_{21}^{(11)}$	-0.036	0.026	0.000	0.007	-0.006
$A_{22}^{(11)}$	-0.023	0.052	0.001	-0.022	0.001
$A_{11}^{(12)}$	-0.069	0.044	0.001	0.008	0.003
$A_{12}^{(12)}$	-0.042	0.087	0.001	-0.031	0.006
$A_{21}^{(12)}$	0.061	0.024	0.000	0.010	-0.013
$A_{22}^{(12)}$	-0.029	0.049	0.001	-0.004	-0.002

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Table C.2: MSIAH(2)-VAR(4): posterior properties

	Mean	Std. dev.	Naive Std. error	Autocorr. lag 1	Autocorr. lag 10
Transition probabilities					
$p_{1,1}$	0.734	0.066	0.001	0.557	-0.005
$p_{2,1}$	0.059	0.018	0.000	0.624	0.088
Standard deviations					
$\sigma_{1,1}$	17.129	1.207	0.017	0.625	0.150
$\sigma_{2,1}$	8.746	0.646	0.009	0.559	0.111
$\sigma_{1,2}$	6.983	0.276	0.004	0.669	0.173
$\sigma_{2,2}$	4.011	0.179	0.003	0.666	0.105
Correlations					
$\rho_{1,1}$	-0.173	0.127	0.002	0.203	0.008
$\rho_{1,2}$	0.078	0.070	0.001	0.284	0.018
Intercepts regime 1					
$\mu_{1,1}$	-0.213	0.949	0.013	0.014	0.032
$\mu_{2,1}$	1.107	0.885	0.013	0.101	0.011
Autoregressive coefficients regime 1					
$A^{(1)}_{11,1}$	0.497	0.147	0.002	0.128	0.016
$A^{(1)}_{12,1}$	0.209	0.287	0.004	0.142	-0.018
$A^{(1)}_{21,1}$	0.069	0.075	0.001	0.156	0.027
$A^{(1)}_{22,1}$	0.419	0.156	0.002	0.222	-0.002
$A^{(2)}_{11,1}$	-0.253	0.191	0.003	0.238	0.020
$A^{(2)}_{12,1}$	-0.134	0.361	0.005	0.191	-0.005
$A^{(2)}_{21,1}$	-0.018	0.094	0.001	0.131	0.025
$A^{(2)}_{22,21}$	-0.092	0.202	0.003	0.237	0.002
$A^{(3)}_{11,1}$	0.172	0.218	0.003	0.173	0.001
$A^{(3)}_{12,1}$	-0.176	0.376	0.005	0.105	0.008
$A^{(3)}_{21,1}$	-0.126	0.122	0.002	0.265	0.006
$A^{(3)}_{22,1}$	0.112	0.217	0.003	0.191	0.004

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

	Mean	Std. dev.	Naive Std. error	Autocorr. lag 1	Autocorr. lag 10
$A^{(4)}_{11,1}$	-0.490	0.217	0.003	0.325	0.078
$A^{(4)}_{12,1}$	0.409	0.343	0.005	0.164	0.019
$A^{(4)}_{21,1}$	0.088	0.106	0.001	0.252	0.029
$A^{(4)}_{22,1}$	0.098	0.205	0.003	0.281	0.031
Intercepts regime 2					
$\mu_{1,2}$	0.295	0.634	0.009	0.163	-0.005
$\mu_{2,2}$	2.058	0.420	0.006	0.210	-0.012
Autoregressive coefficients regime 2					
$A^{(1)}_{11,2}$	0.237	0.059	0.001	0.391	0.041
$A^{(1)}_{12,2}$	0.028	0.099	0.001	0.333	-0.002
$A^{(1)}_{21,2}$	-0.026	0.031	0.000	0.259	0.025
$A^{(1)}_{22,2}$	0.398	0.058	0.001	0.297	-0.024
$A^{(2)}_{11,2}$	0.130	0.048	0.001	0.210	0.014
$A^{(2)}_{12,2}$	0.165	0.088	0.001	0.195	0.013
$A^{(2)}_{21,2}$	-0.032	0.028	0.000	0.194	0.005
$A^{(2)}_{22,2}$	0.092	0.057	0.001	0.321	0.038
$A^{(3)}_{11,2}$	0.099	0.053	0.001	0.377	0.057
$A^{(3)}_{12,2}$	0.214	0.086	0.001	0.195	0.006
$A^{(3)}_{21,2}$	-0.014	0.026	0.000	0.176	0.023
$A^{(3)}_{22,2}$	0.285	0.053	0.001	0.284	0.007
$A^{(4)}_{11,2}$	0.106	0.052	0.001	0.394	0.039
$A^{(4)}_{12,2}$	-0.174	0.092	0.001	0.272	0.014
$A^{(4)}_{21,2}$	-0.019	0.025	0.000	0.200	0.009
$A^{(4)}_{22,2}$	-0.066	0.055	0.001	0.323	0.031

C.3 Characterization of estimation efficiency

Table C.3: Characterization of the efficiency in the models' estimations

	RNE			*Autocorr. lag 1*			*Autocorr. lag 10*			*Geweke z-score*		
$\mathcal{M}_j$	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max
Vector autoregressive models												
$\mathcal{M}_0$	1.00	0.85	1.19	0.00	-0.03	0.06	0.00	-0.03	0.03	-0.10	-2.37	2.38
$\mathcal{M}_1$	1.00	0.76	1.08	0.01	-0.03	0.07	0.00	-0.04	0.02	0.07	-2.57	2.43
Markov switching vector autoregressive models												
$\mathcal{M}_0$	0.48	0.10	1.00	0.24	0.01	0.67	0.02	-0.02	0.17	-0.56	-2.14	3.27
$\mathcal{M}_1$	0.47	0.06	1.00	0.17	-0.02	0.78	0.01	-0.03	0.29	0.22	-1.98	2.58
$\mathcal{M}_2$	0.71	0.13	1.12	0.14	-0.02	0.71	0.01	-0.03	0.08	0.13	-2.10	1.59
$\mathcal{M}_3$	0.30	0.02	0.94	0.27	0.03	0.89	0.04	-0.01	0.56	-0.32	-2.43	1.94
$\mathcal{M}_4$	0.46	0.08	0.83	0.25	0.07	0.78	0.01	-0.03	0.23	-0.20	-1.57	1.56
$\mathcal{M}_5$	0.22	0.02	0.43	0.44	0.12	0.85	0.07	-0.01	0.50	-0.10	-2.39	2.16
$\mathcal{M}_6$	0.24	0.02	0.92	0.26	0.04	0.90	0.04	-0.02	0.58	-0.08	-1.43	1.91
$\mathcal{M}_7$	0.33	0.05	0.83	0.31	0.03	0.84	0.04	-0.01	0.39	-0.16	-2.34	1.67

Table C.3 reports statistics for assessing the efficiency of each estimated model. Three types of statistics are presented: the relative numerical efficiency of Geweke (1989), autocorrelations at different lags, and the convergence diagnostic of Geweke (1992). Statistics should be presented separately for each parameter of each model, but to save space, we summarize each model with a median, minimum, and maximum.

The relative numerical efficiency represents the ratio of the variance of a hypothetical draw from the posterior density over the variance of the Gibbs sampler. Thus, it can be interpreted as a measure of the computational efficiency of the algorithm. The columns of Table C.3, unsurprisingly, tell us that the algorithm for VAR models is more efficient than that for MS-VAR. The same observation can be made when comparing unrestricted models with restricted ones. What is interesting for us is the magnitude of the RNE statistics between unrestricted and restricted models. Those are comparable, which is

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

a good sign that the algorithm for constrained models are, computationally, reasonable efficient.

The columns displaying the autocorrelations at lag 1 and lag 10 are here to ensure that there is a decay over time. This is the case here, and the Gibbs samplers explore the entire posterior distribution.

Geweke (1992) introduces the z scores test which tests the stationarity of the draws from the posterior distribution simulation comparing the mean of the first 30% of the draws with the last 40% of the draws. We compare the two means with a z-test. Typically, values outside $(-2, 2)$ indicate that the mean of the series is still drifting, and this occurs for some parameters in each models, except $\mathcal{M}_4$ and $\mathcal{M}_6$ for MS-VARs. Increasing the burn in period might improve the scores and stationarity of the MCMC chain.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Bibliography

Albert, J. H. and S. Chib (1993). Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts. *Journal of Business & Economic Statistics 11*(1), 1–15.

Barnard, J., R. McCulloch, and X.-l. Meng (2000). Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Applicationto Shrinkage. *Statistica Sinica 10*, 1281–1311.

Bartlett, M. S. (1957). A Comment on D. V. Lindley's Statistical Paradox. *Biometrika 44*(3/4), 533–534.

Boudjellaba, H., J.-M. Dufour, and R. Roy (1992). Testing Causality Between Two Vectors in Multivariate Autoregressive Moving Average Models. *Journal of the American Statistical Association 87*(420), 1082– 1090.

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

Casella, G. and E. I. George (1992). Explaining the Gibbs Sampler. *The American Statistician 46*(3), 167–174.

Chib, S. and I. Jeliazkov (2001). Marginal Likelihood from the Metropolis-Hastings Output. *Journal of the American Statistical Association 96*(453), 270– 281.

Christiano, L. J. and L. Ljungqvist (1988). Money does granger-cause output in the bivariate money-output relation. *Journal of Monetary Economics 22*(2), 217 – 235.

Christopoulos, D. K. and M. A. León-Ledesma (2008). Testing for Granger (Non-) causality in a Time-Varying Coeffi cient. *Journal of Forecasting* (27), 293–303.

Comte, F. and O. Lieberman (2000). Second-Order Noncausality in Multivariate GARCH Processes. *Journal of Time Series Analysis 21*(5), 535–557.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Diebold, F. X. and K. Yilmaz (2009, January). Measuring Financial Asset Return and Volatility Spillovers, with Application to Global Equity Markets. *The Economic Journal 119*(534), 158–171.

Droumaguet, M., A. Warne, and T. Woźniak (2012). Granger causality and regime inference in bayesian markov-switching vars. Unpublished manuscript, European University Institute, Florence, Italy.

Droumaguet, M. and T. Woźniak (2012). Bayesian Testing of Granger Causality in Markov-Switching VARs. Working paper series, European University Institute, Florence, Italy. Download at: http://cadmus.eui.eu/bitstream/handle/1814/20815/ECO_2012_06.pdf?sequence=1.

Dufour, J.-M. (1989). Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions. *Econometrica 54*(2), 335–355.

Dufour, J.-M., D. Pelletier, and E. Renault (2006). Short run and long run causality in time series: inference. *Journal of Econometrics 132*(2), 337–362.

Friedman, M. and A. Schwartz (1971). *A monetary history of the United States, 1867-1960*, Volume 12. Princeton Univ Pr.

Frühwirth-Schnatter, S. (2004, June). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *Econometrics Journal 7*(1), 143–167.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.

Geweke, J. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica 57*(6), pp. 1317–1339.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 4*, pp. 169–194. Clarendon Press, Oxford, UK.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Geweke, J. (1995). Bayesian Comparison of Econometric Models.

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

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica 37*(3), 424–438.

Greenberg, E. and S. Chib (1995, November). Understanding the Metropolis-Hastings Algorithm. *The American Statistician 49*(4), 327–335.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica 57*(2), 357–384.

Hamilton, J. D. (1994). State-Space Models. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics* (Volume IV ed.)., Chapter 50, pp. 3039–3080. Elsevier.

Hoogerheide, L. F., H. K. van Dijk, and R. van Oest (2009). *Simulation Based Bayesian Econometric Inference: Principles and Some Recent Computational Advances*, Chapter 7, pp. 215–280. Handbook of Computational Econometrics. Wiley.

Jarociński, M. and B. Maćkowak (2011). Choice of Variables in Vector Autoregressions.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate arch models. *Econometric Theory 14*(01), 70–86.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kim, C.-J. and C. R. Nelson (1999a). Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle. *Review of Economics and Statistics 81*(4), 608–616.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Kim, C.-J. and C. R. Nelson (1999b). *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*. MIT press.

Koop, G. and D. Korobilis (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics 3*(4), 267–358.

Krolzig, H. (1997). *Markov-switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Springer Verlag.

Lechner, M. (2011). The Relation of Different Concepts of Causality Used in Time Series and Microeconometrics. *Econometric Reviews 30*(1), 109–127.

Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*. Springer-Verlag.

Lütkepohl, H. and M. M. Burda (1997). Modified Wald tests under nonregular conditions. *Journal of Econometrics 78*(1), 315–332.

McCulloch, R. E. and R. S. Tsay (1994). Statistical analysis of economic time series via markov switching models. *Journal of Time Series Analysis 15*(5), 523–539.

Pajor, A. (2011). A Bayesian Analysis of Exogeneity in Models with Latent Variables. *Central European Journal of Economic Modelling and Econometrics 3*(2), 49–73.

Psaradakis, Z., M. O. Ravn, and M. Sola (2005). Markov switching causality and the money-output relationship. *Journal of Applied Econometrics 20*(5), 665–683.

Psaradakis, Z. and M. Sola (1998, October). Finite-sample properties of the maximum likelihood estimator in autoregressive models with Markov switching. *Journal of Econometrics 86*(2), 369–386.

Psaradakis, Z. and N. Spagnolo (2003, March). On the determination ot the number of regimes in markov-switching autoregressive models. *Journal of Time Series Analysis 24*, 237–252.

Ritter, C. and M. A. Tanner (1992). Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of the American Statistical Association 87*(419), 861– 868.

Wozniak, Tomasz (2012), Granger-Causal analysis of conditional mean and volatility models
European University Institute DOI: 10.2870/63858

Robins, R. P., C. W. J. Granger, and R. F. Engle (1986). *Wholesale and Retail Prices: Bivariate Time-Series Modeling with forecastable Error Variances*, pp. 1–17. The MIT Press.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5), 688 – 701.

Sims, C. A. (1972). Money, Income, and Causality. *The American Economic Review 62*(4), 540 – 552.

Sims, C. A., D. F. Waggoner, and T. Zha (2008). Methods for inference in large multiple-equation markov-switching models. *Journal of Econometrics 146*(2), 255–274.

Strachan, R. W. and H. K. van Dijk (2011). Divergent Priors and Well Behaved Bayes Factors.

Taylor, J. B. and J. C. Williams (2009, January). A Black Swan in the Money Market. *American Economic Journal: Macroeconomics 1*(1), 58–83.

Tobin, J. (1970, May). Money and Income: Post Hoc Ergo Propter Hoc? *The Quarterly Journal of Economics 84*(2), 301–317. ArticleType: research-article / Full publication date: May, 1970 / Copyright © 1970 Oxford University Press.

Warne, A. (2000, December). Causality and regime inference in a markov switching var. Working Paper Series 118, Sveriges Riksbank (Central Bank of Sweden).

Woźniak, T. (2012). Granger causal analysis of varma-garch models. EUI Working Papers ECO 2012/19, European University Institute, Florence, Italy. Download at: http://cadmus.eui.eu/bitstream/handle/1814/23336/ECO_2012_19.pdf.

Woźniak, T. (2012). Testing Causality Between Two Vectors in Multivariate GARCH Models. EUI Working Papers ECO 2012/20, European University Institute, Florence, Italy. Download at: http://cadmus.eui.eu/bitstream/handle/1814/23337/ECO_2012_20.pdf.