Bayesian Testing of Granger Causality in Markov-Switching VARs ☆

Matthieu Droumaguet[a], Tomasz Woźniak[b,a]

[a] *Department of Economics, European University Institute, Via della Piazzuola 43, 50133 Florence, Italy*
[b] *Department of Economics, University of Melbourne, Arts West Building, Victoria 3010, Australia*

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 inference consists of a novel Gibbs 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.

Keywords: Granger Causality, Markov Switching Models, Hypothesis Testing, Posterior Odds Ratio, Gibbs Sampling
JEL classification: C11, C12, C32, C53, E32

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

☆This paper was presented during the poster session of the 22^{nd} EC^2 Conference: *Econometrics for Policy Analysis: after the Crisis and Beyond* in Florence in December 2011. The authors thank Anders Warne, Helmut Lütkepohl, Massimiliano Marcellino and Peter R. Hansen for their useful comments on the paper.

Email addresses: `matthieu.droumaguet@eui.eu` (Matthieu Droumaguet), `tomasz.wozniak@unimelb.edu.au` (Tomasz Woźniak)

URL: `http://bit.ly/twozniak` (Tomasz Woźniak)

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. 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, Dufour & Roy, 1994, and references therein), the Logistic Smooth Transition Vector Autoregressive (LST-VAR) model (Christopoulos & León-Ledesma, 2008), some models from the family of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models (Comte & Lieberman, 2000; Woźniak, 2011a). Finally, Warne (2000) derived conditions for different types of Granger noncausality for the Markov-switching VAR models on which we focus in this study. We present the model and its estimation in Section 2, while in Section 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 & Burda, 1997; Dufour, Pelletier & Renault, 2006, for h periods ahead inference in VAR models).

For some of these models, the derived restrictions on parameters are linear and therefore easily testable. For others, this is not the case. For Markov-switching VAR models, the restrictions for Granger noncausality and noncausality in variance are highly nonlinear functions of the original parameters of the model. Consequently, the asymptotic distributions of the Wald, Likelihood Ratio and Lagrange Multiplier tests are not known. Testing of these restrictions becomes cumbersome.

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 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 Gibbs sampling algorithm that allows for restricting the models. The algorithm is discussed in Section 4 and presented in Section 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 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 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 & 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 & 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, Ravn & Sola (2005) used the Markov-switching VAR models to analyze, the so called temporary Granger causality within the Money-Output system. They proposed a restricted MS-VAR specification that assumed four states of the economy: 1. both variables cause 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.

The restrictions that Psaradakis et al. (2005) imposed on the parameters of the model follow some logical inference. The model they proposed certainly investigates time-varying parameters, modeling linear interactions between current and lagged values of the variables. However, one may not interpret their restrictions in terms of Granger noncausality. In the model of Psaradakis et al. (2005), the parameters responsible for Granger noncausality in VAR models, i.e. off-diagonal elements of the matrices of auto-regressive terms, vary with states; restriction (A1)(vi), presented in Section 3, does not hold. In consequence, both the variables are used by the model to forecast the probability of the states one period ahead, $Pr(s_{t+1}|\theta, \mathbf{y})$. This is the channel through which all the considered variables Granger-cause each other, irrespective of whether the off-diagonal element of the autoregressive matrices is set to zero or not. In this study, we present a method of inference referring to precisely stated definitions of Granger causality for Markov-switching models. No confusion about the sources of Granger causality is admissible.

The remaining part of the paper is organized as follows. In Section 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, together with parameter restrictions representing them. Section 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 5. Section 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 7 concludes.

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 on the realizations, s_t, of a hidden discrete stochastic process S_t with finite state space $\{1, \ldots, M\}$. 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 $i = 1, \ldots, p$, and covariance matrices, Σ_{s_t}, depend on the state $s_t = 1, \ldots, M$.

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

$$\epsilon_t \sim i.i.\mathcal{N}(\mathbf{0}, \Sigma_{s_t}), \tag{2}$$

for $t = 1, \ldots, T$. We set the vector of initial values $\mathbf{y}_0 = (y_{p-1}, \ldots, y_0)'$ 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 $\pi = (\pi_1, \ldots, \pi_M)$, 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 $A^{(i)}_{s_t}$, 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.

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}: $\mu_{s_t}, A^{(i)}_{s_t}, \Sigma_{s_t}$, 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:

$$p(\mathbf{S}, \mathbf{y}|\theta) = p(\mathbf{y}|\mathbf{S}, \theta)p(\mathbf{S}|\theta). \tag{3}$$

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:

$$p(\mathbf{y}|\mathbf{S}, \theta) = \prod_{t=1}^{T} p(y_t|\mathbf{S}, \mathbf{y}^{t-1}, \theta) = \prod_{t=1}^{T} (2\pi)^{-K/2} |\Sigma_{s_t}|^{-1/2} \exp\left\{-\frac{1}{2}\epsilon_t' \Sigma_{s_t}^{-1} \epsilon_t\right\}. \tag{4}$$

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

$$p(\mathbf{S}|\theta) = p(s_0|\mathbf{P}) \prod_{i=1}^{M} \prod_{j=1}^{M} p_{ij}^{N_{ij}(\mathbf{S})}, \tag{5}$$

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) 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, $\mathbf{P}$:

$$p(\mathbf{y}, \mathbf{S}|\theta) = \prod_{i=1}^{M} \left(\prod_{t:s_t=i} p(y_t|\mathbf{y}^{t-1}, \theta_i) \right) \prod_{i=1}^{M} \prod_{j=1}^{M} p_{ij}^{N_{ij}(\mathbf{S})} p(s_0|\mathbf{P}). \tag{6}$$

Classical estimation of the model consists of the maximization of the likelihood function with e.g. the EM algorithm (see Krolzig, 1997; Kim & 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 (6) and the prior distribution:

$$p(\theta|\mathbf{y}, \mathbf{S}) \propto p(\mathbf{y}, \mathbf{S}|\theta)p(\theta). \tag{7}$$

Prior distribution. The convenient factorization of the likelihood function (6) is maintained by the choice of the prior distribution in the following form:

$$p(\theta) = \prod_{i=1}^{M} p(\theta_i)p(\mathbf{P}). \tag{8}$$

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 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 & 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 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, McCulloch & Meng (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 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 $s_t = 1, \dots, M$ and $j = 1, \dots, N$, follows a log-Normal distribution, with a mean parameter equal to 0 and the standard deviation parameter set to 2. Finally, we assume that the prior distributions of the correlation matrices $\mathbf{R}_{s_t}$ are proportional to a constant (the implications of such a prior specification are discussed in the original paper of Barnard et al., 2000).

To summarize, the prior specification (8) now takes the detailed form of:

$$p(\theta) = \prod_{i=1}^{M} p(\mathbf{P}_i)p(\beta_i)p(\mathbf{R}_i)\left(\prod_{j=1}^{N} p(\sigma_{i.j})\right), \tag{9}$$

where each of the prior distributions is as assumed:

$$\begin{aligned} \mathbf{P}_{i\cdot} &\sim \mathcal{D}_M(\imath_M' + 9I_{M.i\cdot}) \\ \beta_i &\sim \mathcal{N}(\mathbf{0}, 100I_{N+pN^2}) \\ \sigma_{i.j} &\sim log\mathcal{N}(0, 2) \\ \mathbf{R}_i &\propto 1 \end{aligned}$$

for $i = 1, \dots, M$ and $j = 1, \dots, N$, 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 (6) and the prior distribution (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 (7), resulting from the assumed specification, is as follows:

$$p(\theta|\mathbf{y}, \mathbf{S}) \propto \prod_{i=1}^{M} p(\theta_i|\mathbf{y}, \mathbf{S})p(\mathbf{P}|\mathbf{y}, \mathbf{S}). \tag{10}$$

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

$$p(\mathbf{P}|\mathbf{S}) \propto p(s_0|\mathbf{P}) \prod_{i=1}^{M} \prod_{j=1}^{M} p_{ij}^{N_{ij}(\mathbf{S})} p(\mathbf{P}), \tag{11}$$

and the posterior density of the state-dependent parameters:

$$p(\theta_i|\mathbf{y}, \mathbf{S}) \propto \prod_{t:S_t=i} p(y_t|\theta_i, \mathbf{y}_{t-1},)p(\theta_i). \tag{12}$$

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 & 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 & Chib (1993) and McCulloch & Tsay (1994).

In the Gibbs sampling algorithm, parameters of the model are split into sub-vectors, the full conditional densities of which are of convenient form. Firstly, however, we draw a vector of the states of the economy, $\mathbf{S}$. We initialize the algorithm, conditioning on the starting values for the parameters, $\theta^{(0)}$. Then, using the BLHK filter and smoother (see Frühwirth-Schnatter, 2006, Chapter 11 and references therein), we 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.

Secondly, we draw from the posterior distribution of the transition probabilities matrix (11), conditioning on the states drawn in the previous step of the current iteration, $\mathbf{P}^{(l)} \sim p(\mathbf{P}|\mathbf{S}^{(l)})$. Assuming the Dirichlet prior distribution and that the hidden Markov process starts from its ergodic distribution, π, makes the posterior distribution not of standard form. In this step of the Gibbs sampler, we use the Metropolis-Hastings algorithm as described in (Frühwirth-Schnatter, 2006, Section 11.5.5).

Thirdly, we draw the state-dependent parameters of the VAR process collected in one vector, $\beta = (\beta_1', \ldots, \beta_M')'$. Due to the form of the likelihood function and Normal prior distribution, the full conditional distribution is also normal $f(\beta|\mathbf{y}, \mathbf{S}^{(l)}, \mathbf{P}^{(l)}, \sigma^{(l-1)}, \mathbf{R}^{(l-1)}) = \mathcal{N}\left(\bar{\beta}^*, \bar{V}_{\beta^*}\right)$, from which we draw $\beta^{(l)}$. $\bar{\beta}^*$ and $\bar{V}_{\beta^*}$ are the parameters of the full conditional distribution specified in Section 5 (see also Frühwirth-Schnatter, 2006, Section 8.4.3).

Finally, we collect all the standard deviations in one vector, $\sigma = (\sigma_1', \ldots, \sigma_M')'$, and all the unknown correlation coefficients into a vector, $\mathbf{R} = (\text{vecl}(\mathbf{R}_1)', \ldots, \text{vecl}(\mathbf{R}_M)')'$, where function, vecl, stacks all the lower-diagonal elements of the correlation matrix into a vector. In order to draw from the full conditional densities of these two vectors, $f(\sigma|\mathbf{y}, \mathbf{S}^{(l)}, \mathbf{P}^{(l)}, \beta^{(l)}, \mathbf{R}^{(l-1)})$ and $f(\mathbf{R}|\mathbf{y}, \mathbf{S}^{(l)}, \mathbf{P}^{(l)}, \beta^{(l)}, \sigma^{(l)})$, we employ the Griddy-Gibbs sampling algorithm of Ritter & Tanner (1992), as described by Barnard et al. (2000).

The algorithm for the restricted models is presented in detail in Section 5.

3. Granger Causality

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

$$y_t = (y_{1t}', y_{2t}', y_{3t}', y_{4t}')', \tag{13}$$

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 contained 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.:

$$u_{t+1} = y_{t+1} - E[y_{t+1}|\mathbf{y}_t]. \tag{14}$$

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

$$E\left[u_{t+1}^2\right] = E\left[\tilde{u}_{t+1}^2\right] < \infty \quad \forall t = 1, \ldots, T. \tag{15}$$

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, Dufour & Roy, 1992). It is argued, however, that this definition cannot give a full insight into relations between variables under changing economic circumstances (see Woźniak, 2011a). 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, Granger & Engle (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 (15).

Definition 2. y_4 does not Granger-cause in variance y_1, denoted by $y_4 \overset{V}{\nrightarrow} y_1$, if and only if:

$$E\left[u_{t+1}^2 | \mathbf{y}_t\right] = E\left[\tilde{u}_{t+1}^2 | \mathbf{v}_{1t}, \mathbf{y}_{3t}\right] < \infty \quad \forall t. \tag{16}$$

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

Definition 3. y_4 does not Granger-cause in distribution y_1, denoted by $y_4 \overset{D}{\nrightarrow} y_1$, if and only if:

$$g_{t+1}\left(u_{t+1}^2 | \mathbf{y}_t, \theta\right) = h_{t+1}\left(\tilde{u}_{t+1}^2 | \mathbf{v}_{1t}, \mathbf{y}_{3t}, \theta\right) \quad \forall t, \tag{17}$$

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

Granger noncausality in distribution is the strictest concept of the three defined. It implies Granger noncausality in variance (16) and Granger noncausality (15), which gives the order of inference that (17) implies (16) that implies (15). All three definitions are identical in linear Gaussian models. Therefore, in MS-VAR models, which are not linear, stating (17) implies (15) and (16), but stating (15) does not determine either (16) or (17). All the definitions are given in the form following Warne (2000).

Comte & Lieberman (2000) introduce a new definition of *second-order Granger noncausality* and distinguish it from *Granger noncausality in variance* of Robins et al. (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 noncausality 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 (2011a).

Putting these definitions in the context of economic (macroeconomic or financial) time series gives a new perspective. From the practitioner's point of view, when two variables appear in a time series model, one usually expects them to be somehow related. If one cannot reject a null hypothesis of noncausality in distribution, then the result should be interpreted as very informative. Such a finding, however, will be met extremely rarely, and in fact testing the noncausality in distribution will give little information. Moreover, an even less strict hypothesis of noncausality in variance is also found to be problematic. Woźniak (2011a) shows using the example of trivariate daily series of exchange rates, that testing this hypothesis is not informative due to the very high frequency of rejections of the null hypothesis. Only when the problem is reformulated the Granger causality and second-order noncausality are tested separately, is the analysis informative.

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 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 (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}:

$$\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1.s_{1t}} \\ \mu_{2.s_{2t}} \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} A^{(i)}_{11.s_{1t}} & A^{(i)}_{12.s_{1t}} \\ A^{(i)}_{21.s_{2t}} & A^{(i)}_{22.s_{2t}} \end{bmatrix} \begin{bmatrix} v_{1t-i} \\ v_{2t-i} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}. \tag{18}$$

The residual term in (18) has zero conditional mean and conditional covariance matrix decomposed into sub-matrices as on the left-hand side of (19):

$$\operatorname{Var}\left(\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{11.s_{1t}} & \Sigma'_{21.s_t} \\ \Sigma_{21.s_t} & \Sigma_{22.s_{2t}} \end{bmatrix}, \qquad \operatorname{Var}\left(\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{bmatrix}\right) = \begin{bmatrix} \Omega_{11.s_{1t}} & \Omega'_{21.s_t} & \Omega'_{31.s_t} & \Omega'_{41.s_t} \\ \Omega_{21.s_t} & \Omega_{22.s_{1t}} & \Omega'_{32.s_t} & \Omega'_{42.s_t} \\ \Omega_{31.s_t} & \Omega_{32.s_t} & \Omega_{33.s_{2t}} & \Omega'_{43.s_t} \\ \Omega_{41.s_t} & \Omega_{42.s_t} & \Omega_{43.s_t} & \Omega_{44.s_{2t}} \end{bmatrix}, \tag{19}$$

where covariance matrices may be decomposed respectively into:

$$\Sigma_{ij.s_t} = \operatorname{diag}(\sigma_{i.s_t})\mathbf{R}_{ij.s_t}\operatorname{diag}(\sigma_{j.s_t}), \qquad \Omega_{ij.s_t} = \operatorname{diag}(\omega_{i.s_t})R_{ij.s_t}\operatorname{diag}(\omega_{j.s_t}). \tag{20}$$

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 (19). The decomposition of the Markov process is maintained, as in (18):

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

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.:

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

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

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

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

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 for one period ahead forecast:

$$\bar{m}_{1t} \equiv E\left[m_{1.s_{t+1}} | \mathbf{y}_t, \theta\right], \tag{22a}$$

$$\bar{a}^{(k)}_{1rt} \equiv E\left[a^{(k)}_{1r.s_{t+1}} | \mathbf{y}_t, \theta\right], \tag{22b}$$

for all $r = 1, \dots, N$ and $k = 1, \dots, p$. 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 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) $\sum_{j=1}^{M} \left[(a^{(k)}_{1r.j} - \bar{a}^{(k)}_{1r}) \otimes (a^{(l)}_{1s.j} - \bar{a}^{(l)}_{1s}) \right] p_{ij} = \varsigma^{(k.l)}_{r.s}$,

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

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

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

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

is satisfied.

In condition (A4), ς_m, $\varsigma^{(k.l)}_{r.s}$, $\varsigma^{(k)}_{\mu.r}$ 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) $a^{(k)}_{1r.j} = a^{(k)}_{1r.j_1}$,

(iv) $a^{(k)}_{14.j} = 0$, 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.

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 $\text{rank}(\mathbf{P}^{(2)}) = M_2$. We present these in Appendix A.

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 problem with classical tests comes from the fact that conditions (A3) and (A4) are nonlinear functions of the original parameters of the model. The proposed solution consists of a new Gibbs 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. We start the discussion of the available tests with the problem of testing the restrictions, that state the causality relations for a hidden Markov process, (A1) and (A2), and the restrictions for noncausality in distribution, (A5). All these conditions are linear in parameters and can be easily tested with the Wald, Lagrange Multiplier (LM) or the Likelihood Ratio (LR) test. Given the asymptotic normality of the maximum likelihood estimator (the result established for Markov-switching models by Lindgren, 1978), all the test statistics are χ^2-distributed.

However, testing of the conditions for Granger noncausality, (A3), and for noncausality in variance, (A4), becomes cumbersome in classical inference. It is worth emphasizing that these two restrictions are of special economic interest, as they deliver easily interpretable information about the dependencies between variables in first two conditional moments. Moreover, they are less strict than conditions (A1), (A2) and (A6), and therefore testing them gives more information, due to a potentially lower rate of rejections of the null hypothesis.

The problem, as already pointed out, is in the nonlinearity of these restrictions. Consequently, in the general case, for any allowed value of M_1, M_2, p, and for any value of the rank of matrix $\mathbf{P}$, the matrices of first partial derivatives of the restrictions (A3) or (A4) with respect to θ are not of full rank. In effect, the asymptotic distribution of the test statistics of the Wald, LM and LR tests is unknown. For further explanations and examples, the reader is referred to Section 4.2 of Warne (2000).

This problem is well known in the studies on testing 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 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 & 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.

In case of the Logistic Smooth Transition Vector Autoregressive (LST-VAR) model, the restrictions derived in Christopoulos & León-Ledesma (2008) are linear and can be tested.

Multivariate models for second conditional moments have also been considered for testing second-order noncausality and noncausality in variance. First, Comte & Lieberman (2000) derive the conditions for second-order noncausality between two vectors of variables for the family of BEKK-GARCH and vec-GARCH models. Nevertheless, they do not propose any method of testing them, due to problems with the asymptotic distribution of the test statistic, as described in this section. Hafner & Herwartz (2008) derive a new set of linear restrictions that are easily testable, but which are only a sufficient condition for second-order noncausality.

It is necessary to mention at this stage the approach to testing noncausality in variance proposed by Cheung & Ng (1996). They propose a two-stage procedure. In the first, for each of the variables a univariate ARMA-GARCH model is fitted and estimated. In the second, a test based on the cross-correlation function between squared residuals from the first stage for all the considered variables is performed. Other papers, such as Hong (2001), Pantelidis & Pittis (2004) and van Dijk, Osborn & Sensier (2005), work on the same testing procedure, improving its power and size properties for different features of data. This approach allows for detection of the Granger causal relations, however, it does not model the spillovers and is instead proposed as a pre-estimation method of constructing multivariate models.

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 & Burda (1997) seems applicable to the problem considered in this work. 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, following two papers of Woźniak (2011a, 2011b). Both of the papers work on the Extended CCC-GARCH model of Jeantheau (1998). Woźniak

(2011b) derives the restrictions for the second-order noncausality between two vectors of variables. In order to compare the unrestricted model, denoted by $\mathcal{M}_i$, and the restricted model, $\mathcal{M}_j$ and $= j \neq i$, he uses 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:

$$\text{POR} = \frac{Pr(\mathcal{M}_i|\mathbf{y})}{Pr(\mathcal{M}_j|\mathbf{y})} = \frac{p(\mathbf{y}|\mathcal{M}_i)}{p(\mathbf{y}|\mathcal{M}_j)} \frac{Pr(\mathcal{M}_i)}{Pr(\mathcal{M}_j)}, \tag{23}$$

where $p(\mathbf{y}|\mathcal{M})$ is the marginal density of data and $Pr(\mathcal{M})$ is the prior probability of a model. Note that if one chooses not to discriminate any of the hypotheses *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:

$$\mathcal{B}_{ij} = \frac{Pr(\mathbf{y}|\mathcal{M}_i)}{Pr(\mathbf{y}|\mathcal{M}_j)}. \tag{24}$$

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 & Raftery, 1995).

In the other paper, Woźniak (2011a) analyses a family of VARMA-GARCH models in terms of Granger noncausality, second-order noncausality and noncausality in variance. For joint testing of the multiple restrictions, he proposes a function summarizing all the restrictions:

$$\kappa(R(\theta)) = \left[R(\theta) - E[R(\theta)|\mathbf{y}]\right]' V[R(\theta)|\mathbf{y}]^{-1} \left[R(\theta) - E[R(\theta)|\mathbf{y}]\right], \tag{25}$$

where $R(\theta)$ is a vector collecting all the restrictions representing a hypothesis of interest, and $E[R(\theta)|\mathbf{y}]$ and $V[R(\theta)|\mathbf{y}]$ are posterior expected value and posterior covariance matrix of $R(\theta)$ respectively. In case of testing the hypothesis of noncausality, the null hypothesis restrictions are compared to a vector of zeros, $R(\theta) = \mathbf{0}$. In order to decide whether the null hypothesis is credible, the value of the function in (25) for the vector representing the null hypothesis, $\kappa(\mathbf{0})$, is placed in the posterior density for the unrestricted model, $p(\kappa(R(\theta))|\mathbf{y})$. The relation of $\kappa(\mathbf{0})$ to $p(\kappa(R(\theta))|\mathbf{y})$ may be summarized by a percentile. It is up to the researcher to evaluate whether $\kappa(\mathbf{0})$ lays in a probability region of $p(\kappa(R(\theta))|\mathbf{y})$, such that it should be rejected, or not. Such a method of evaluation of the credibility of hypotheses was used e.g. by Geweke (2010) and Hoogerheide, van Dijk & van Oest (2009). This method of testing does not suffer from the drawbacks of the Wald test. Moreover, the analysis is based on the posterior distribution, which is the exact finite sample distribution, and there is no need to refer to the asymptotic theory if the assumed model is correct.

Testing the noncausality restrictions in MS-VARs. Both of the presented approaches are applicable to the problem of testing noncausality in Markov-switching VAR models. However, taking into account the complicated structure of the restrictions, i.e. the fact that the form of restrictions depends on the rank of matrix $\mathbf{P}$ and whether restriction (A2) holds (see Appendix A), we focus on the first of the presented methods, namely the Posterior Odds Ratio (23). 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, $p(\theta|\mathbf{y})$, the marginal density of data is computed using:

$$p(\mathbf{y}) = \left(S^{-1} \sum_{i=1}^{S} \frac{h(\theta^{(i)})}{L(\mathbf{y};\theta^{(i)})p(\theta^{(i)})} \right)^{-1}, \tag{26}$$

where $L(\mathbf{y};\theta^{(i)})$ is a likelihood function. $h(\theta^{(i)})$, 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, Waggoner & Zha (2008), Chib (1995) and Chib & Jeliazkov (2001). In fact, the results of several methods should be reported in order to present the robustness of the estimators. 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, 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 Gibbs sampling algorithm specially constructed for the purpose of testing noncausality hypotheses in the MS-VAR models in Section 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 & 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, **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, 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 & 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 & Spagnolo, 2003; Psaradakis & Sola, 1998), the Bayesian model selection proposed in Step 1 is a proper method free of 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.

5. The Gibbs sampler for restricted MS-VAR models

This section scrutinizes the Gibbs 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, and conditional on the parameter draws from the four other blocks, we describe how we sample from the posterior distribution.

The vector of states of the economy is drawn from a discrete conditional distribution, for which the weights are given by the filter of the multi-move Gibbs sampler. Transition probabilities are sampled from Dirichlet distributions in a Metropolis-Hastings sampler. For the covariance matrices, two steps are considered. In the first, posteriors for the standard deviations are sampled thanks to a griddy-Gibbs sampler. In the second, another griddy-Gibbs setup allows us to draw from the full conditional distribution of correlations bounded by the necessity of positive definiteness of the resulting covariance matrix. Finally, the last block of parameters, the regime-dependent autoregressive parameters, are simultaneously drawn from a multivariate normal distribution. In the upcoming notation, the symbols l and $l-1$ often appear. They refer to the iteration of the Gibbs sampler, i.e. iteration l will often make use of draws from iteration $l-1$. For the first iteration of a Gibbs run, $l = 1$, initial parameter values either come from the EM parameter estimates in the case of a burn-in run of the Gibbs, or they are the last posterior draws from a burn-in run. The rest of this section describes all the constituting blocks that form the Gibbs sampler.

5.1. Sampling of the vector of the states of the economy

The first drawn parameter is the vector representing the states of the economy, **S**. Being a latent variable, there are no priors nor restrictions on **S**. We first use a filter (see **?**, [Section 11.2)nd references therein]Fruhwirth-Schnatter2006 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.

Algorithm 1. *Multi-move sampling of the states.*

1. *BLHK* filter: Inherited from classical inference, and following its description from Krolzig (1997), it performs the filtering and smoothing operations on the regime probabilities ξ_t. ξ_t denotes the probabilities for the unobserved state of the system.

$$\xi_t = \begin{bmatrix} Pr(s_t = 1) \\ \vdots \\ Pr(s_t = M) \end{bmatrix}$$

The filter, introduced by Hamilton (1989), is an iterative algorithm calculating the optimal forecast of the value of ξ_{t+1} on the basis of the information set in t consisting of the observed values of y_t, namely $\mathbf{y}_t = (y_t', y_{t-1}', \ldots, y_{1-p}')'$. The initial state $\hat{\xi}_{1|0}$ is initialized with the vector of ergodic regime probabilities $\bar{\xi} = \pi$, where π satisfies the equation $\mathbf{P}\pi = \pi$. This step is a forward recursion, i.e. for $t = 1, \ldots, T$, written as:

$$\hat{\xi}_{t+1|t} = \frac{\mathbf{P}'\left(\eta_t \odot \hat{\xi}_{t|t-1}\right)}{1_M'\left(\eta_t \odot \mathbf{P}'\hat{\xi}_{t-1|t-1}\right)},$$

where $\odot$ denotes the element-wise matrix multiplication and η_t is the collection of M densities, defined as:

$$\eta_t = \begin{bmatrix} Pr\left(y_t | s_t = 1, \mathbf{y}_{t-1}, \theta^{(l-1)}\right) \\ \vdots \\ Pr\left(y_t | s_t = M, \mathbf{y}_{t-1}, \theta^{(l-1)}\right) \end{bmatrix}.$$

To compute the smoothed probabilities, full-sample information is used to make an inference about the unobserved regimes by incorporating the previously neglected sample information $\mathbf{y}_{t+1 \cdots T} = (y_{t+1}', \cdots, y_T')'$ into the inference about ξ_t. This step is a backward recursion, for $j = 1, \ldots, T-1$. The iteration consists of the following equation:

$$\hat{\xi}_{T-j|T} = \left[\mathbf{P}\left(\hat{\xi}_{T-j+1|T} \oslash \hat{\xi}_{T-j+1|T-j}\right)\right] \odot \hat{\xi}_{T-j|T-j},$$

where $\oslash$ denotes the element-wise matrix division.

2. Using the smoothed probability $\hat{\xi}_{T|T}$ as the conditional distribution for $s_T | \mathbf{y}, \theta^{(l-1)}$, we sample s_T.
3. The conditional distributions for $s_t | s_{t+1}, \mathbf{y}, \theta^{(l-1)}$ with $t = T-1, T-2, \ldots, 0$ are given by the smoother:

$$\hat{\xi}_{t|t+1} | s_{t+1}, \mathbf{y}, \theta^{(l-1)} = \left[\mathbf{P}\left(\hat{\xi}_{t+1} \oslash \hat{\xi}_{t+1|t-j}\right)\right] \odot \hat{\xi}_{t|t}.$$

$\mathbf{s}$ is thereby sampled for all periods, $t = 1, \ldots, T$.

5.2. Sampling the transition probabilities

In this step of the Gibbs 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 **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 **M**, the transition matrix is decomposed as:

$$p = \mathbf{M}\mathbf{w}, \tag{27}$$

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

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

where each M_{ij} satisfies the following conditions:

1. For each (i, j), all elements of M_{ij} are non-negative.
2. $\iota_M' M_{ij} = \Lambda_{ij} \iota_{d_j}'$, where Λ_{ij} is the sum of the elements in any column of M_{ij}.
3. Each row of **M** has, at most, one non-zero element.
4. M is such that **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 **P** is irreducible, forbidding the presence of an absorbing state that would render the Markov chain **S** non-stationary. The non-independence of the rows of **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 **P**, direct Gibbs 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, $w_j \sim \mathcal{D}_{d_j}(\beta_{1,j}, \ldots, \beta_{d_j,j})$. We then transform the column vector **w** into our candidate matrix of transitions probabilities using equation (27). Finally, we compute the acceptance rate before retaining or discarding the draw.

Algorithm 2. *Metropolis-Hastings for the restricted transition matrix.*

1. $s_0 \sim \pi$. The initial state is drawn from the ergodic distribution of **P**.
2. $w_j \sim \mathcal{D}_{d_j}(n_{1,j} + \beta_{1,j}, \ldots, n_{d_j,j} + \beta_{d_j,j})$ for $j = 1, \ldots, Q$. $n_{i,j}$ corresponds to the number of transitions from state i to state j, counted from **S**. The candidate transition probabilities matrix – in the transformed notation – are sampled from a Dirichlet distribution.
3. $\mathbf{P}^{new} = \mathbf{M}\mathbf{w}$. The proposal for the transitions probabilities matrix is reconstructed.
4. Accept $\mathbf{P}^{new}$ if $u \leq \frac{\pi_{s_0}^{new}}{\pi_{s_0}^{l-1}}$, where $u \sim \mathcal{U}[0, 1]$.

5.3. *Sampling the covariance matrices*

In the standard non-restricted case, the conditional posterior of Σ_{s_t} can be simulated from inverse-Wishart distributions. However, in the present case we aim to estimate covariance matrices upon which we will impose restrictions such as state-invariant variance or correlation, and zero correlation for any parameter. In that case, the simultaneous sampling of all the covariance parameters from an inverted-Wishart becomes impossible. Adapting the approach proposed by Barnard et al. (2000) to Markov-switching models, we are able to 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 matrix, $\mathbf{R}_{s_t}$, thanks to 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 & 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 σ_{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 3. *Griddy-Gibbs for the standard deviations.* The algorithm iterates on all the standard deviation parameters σ_{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 $(\hat{\sigma}_{i,s_t} - 2\sigma_{\hat{\sigma}_{i,s_t}}, \hat{\sigma}_{i,s_t} + 2\sigma_{\hat{\sigma}_{i,s_t}})$ into G grid points.

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

Algorithm 4. *Griddy-Gibbs for the correlations* The algorithm iterates on all the correlation parameters $\mathbf{R}_{i,s_t}$ for $i = 1, \ldots, \frac{(N-1)N}{2}$ and $s_t = 1, \ldots, M$. Similarly to Barnard et al. (2000), we assume uniform distribution on the feasible set of correlations, $\mathbf{R}_{i,s} \sim \mathcal{U}(a, b)$, 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 $s_t = 1, \ldots, M$, 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: $\mathbf{R}_{i,s_t}|\mathbf{y}, \mathbf{S}, \mathbf{P}, \beta, \sigma, \mathbf{R}_{-\mathbf{R}_{i,s_t}} \propto p(\mathbf{y}|\mathbf{S}, \theta) \cdot p(\mathbf{R}_{i,s_t})$.

5.4. Sampling the vector autoregressive parameters

Finally, we draw the state-dependent autoregressive parameters, β_{s_t} for $s_t = 1, \ldots, M$. 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 $\beta = (\beta_0, \beta_1, \ldots, \beta_M)$, 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. The regression model becomes:

$$y_t = Z_t\beta_0 + Z_t D_{i1}\beta_1 + \cdots + Z_t D_{iM}\beta_M + \epsilon_t, \tag{28}$$

$$\epsilon_t \sim i.i.\mathcal{N}(\mathbf{0}, \Sigma_{s_t}). \tag{29}$$

We have here introduced the D_{is}, 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 & 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:

$$Z_t = \mathrm{diag}(z_{1,0,t}, \ldots, z_{N,0,t}, z_{1,1,t}, \ldots, z_{N,1,t}, \ldots, z_{1,M,t}, \ldots, z_{N,M,t}).$$

This notation enables the restriction of each parameter, by simply setting $z_{n,s_t,t}$ to 0 where desired.

Algorithm 5. *The Gibbs sampler for the autoregressive parameters.* We assume normal prior for β, i.e. $\beta \sim \mathcal{N}(\underline{\beta}, \underline{V_\beta})$.

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:

$$\overline{V_\beta} = \left(\underline{V_\beta^{-1}} + \sum_{t=1}^{T} Z_t' \Sigma_{s_t}^{-1} Z_t \right)^{-1}$$

and

$$\overline{\beta} = \overline{V_\beta} \left(\underline{V_\beta^{-1}} \underline{\beta} + \sum_{t=1}^{T} Z_t' \Sigma_{s_t}^{-1} y_t \right).$$

5.5. Simulating restrictions in the form of functions of the parameters.

Some of the restrictions for Granger noncausality presented in Section 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, the full conditional distributions for β or σ are no longer 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 6. *Metropolis-Hastings for the restricted vector autoregressive parameters β.*

1. Form a candidate draw, β^{new}, using Algorithm 7.
2. Compute the probability of acceptance of a draw:

$$\alpha(\beta^{l-1}, \beta^{new}) = \min \left[\frac{p(\mathbf{y}|\mathbf{S}, \mathbf{P}, \beta^{new}, \sigma, \mathbf{R}) p(\beta^{new})}{p(\mathbf{y}|\mathbf{S}, \mathbf{P}, \beta^{l-1}, \sigma, \mathbf{R}) p(\beta^{l-1})}, 1 \right]. \tag{30}$$

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 & Chib (1995). The formula for computing the acceptance probability from equation (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 5, whereas β_i is determined by a deterministic function.

Algorithm 7. *Generating a candidate draw* β*.*

1. Restrict parameter β_i to zero. Draw all the parameters $(\beta_1, \ldots, \beta_{i-1}, \mathbf{0}, \beta_{i+1}, \ldots, \beta_k)'$ according to the algorithms described in Section 5.4.
2. Compute $\beta_i = g(\beta_{-i}, \sigma, \mathbf{R}, \mathbf{P})$.
3. Return the vector $(\beta_1, \ldots, \beta_{i-1}, \mathbf{g}(\beta_{-\mathbf{i}}, \sigma, \mathbf{R}, \mathbf{P}), \beta_{i+1}, \ldots, \beta_k)'$

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),[1] 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 & 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 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.

[1]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.

However, the model of Psaradakis et al. (2005) is ad-hoc, in that the number of states is imposed by the necessity of the analysis, while not necessarily supported by the data. Moreover, despite the fact that their model resembles the setting of restriction (A1), i.e. the most strict restriction, implying noncausality in mean, variance and distribution, it cannot be interpreted as testing the Granger noncausality hypothesis. The reason for this is that the off-diagonal elements of autoregressive polynomial matrices are not set to zero in all states of the model (e.g. the elements $A_{12}^{(i)}$ using our notation are not set to zero in states one and three). This is a violation of restriction (A1)(vi). Therefore, the model of Psaradakis et al. (2005) is a suitable tool for modeling parameters that change over time and are responsible for Granger noncausality in VAR models, however it cannot be used as a tool for investigating Granger noncausality in MS-VARs.

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.

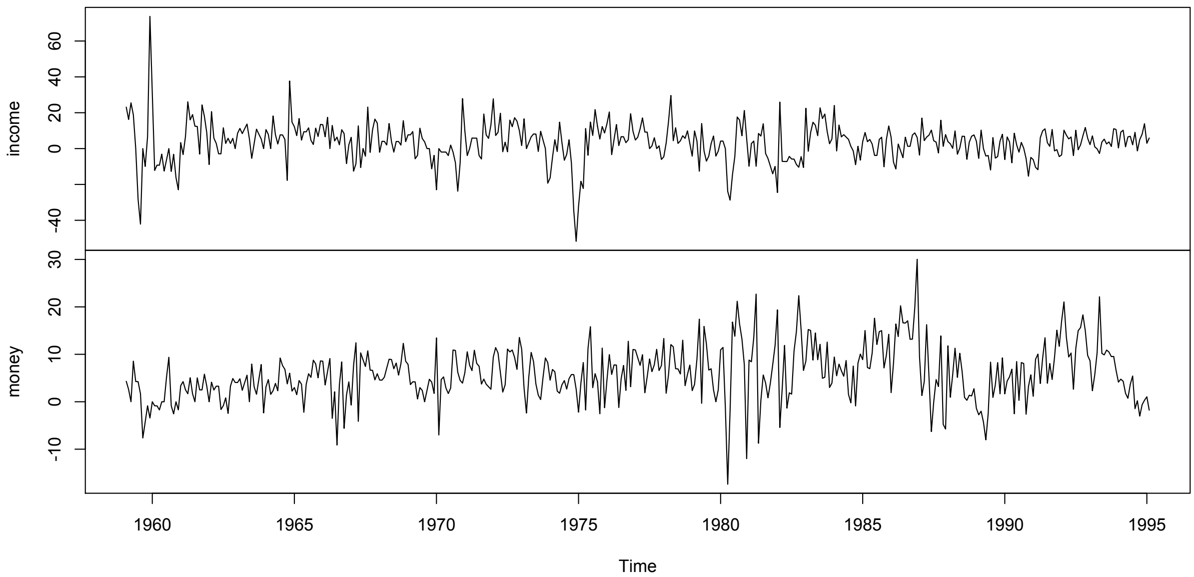

Figure 1: Log-differentiated series of money and income.

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

Table 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.

1200. Warne (2000) performed Johansen tests for cointegration, and – unlike for level series – trace statistics 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 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 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 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 Gibbs sampler of Section 5 can easily be simplified to a Gibbs 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, $p = 0, \dots, 17$. Each of the Gibbs 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:

$$\beta_i \sim \mathcal{N}(\mathbf{0}, 100 I_{N+pN^2})$$
$$\sigma_{i.j} \sim log\mathcal{N}(0, 2)$$
$$\mathbf{R}_i \propto 1$$

for $i = 1, \dots, M$ and $j = 1, \dots, N$.

Table 2 displays the marginal density of data for each model, computed with the modified harmonic mean obtained by applying formula (26) to the posteriors draws. As

Table 2: Model selection for VAR(p) – determination of number of lags

Lags	0	1	2	3	4	5	6	7	8
MHM	-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
MHM	-2936.67	-2941.2	-2917.97	**-2916.77**	-2917.87	-2926.21	-2923.23	-2930.82	-2936.96

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 MHM and hence is the model we choose for the Granger causality analysis. Table B.8 in Appendix B displays, for each parameter of the model, the mean, standard deviations, naive standard errors, autocorrelations of the Markov Chain at lag 1 and lag 10. Low autocorrelation at lag 10 indicates that the Gibbs 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:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} A_{11}^{(i)} & A_{12}^{(i)} \\ A_{21}^{(i)} & A_{22}^{(i)} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix},$$

the restrictions for money, $y_{2,t}$, being Granger noncausal on income, $y_{1,t}$, read:

$$A_{12}^{(i)} = 0 \text{ for } i = 1, \ldots, p.$$

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 MHM 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 summarizes the results for VAR models.

In Table 4, we reproduce the interpretation scale for the Bayes factor (equivalent to the posterior odds ratio when not discriminating between the hypotheses a priori) of Kass & Raftery (1995). 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 favour of Granger noncausality from money to income, within the VAR framework. This result is in line with Christiano & 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: Noncausality and conditional regime independence in a VAR(12) model. Numerical efficiency results for these models are presented in table C.10 of Appendix C.

$\mathcal{M}_j$	Hypothesis	Restrictions	# restrictions	$\ln p(\mathbf{y}\vert\mathcal{M}_j)$	$\ln \mathcal{B}_{j1}$
	$\mathcal{H}_0$: *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 $i = 1, \ldots, p$.

Table 4: Interpretation of the Bayes factor in weights against the null hypothesis

$\ln\left(\frac{Pr(\mathcal{M}\vert\mathbf{y})}{Pr(\mathcal{M}_0\vert\mathbf{y})}\right)$	$\frac{Pr(\mathcal{M}\vert\mathbf{y})}{Pr(\mathcal{M}_0\vert\mathbf{y})}$	Evidence against H_0
0 to 1	1 to 3	Not worth more than a bare mention
1 to 3	3 to 20	Positive
3 to 5	20 to 150	Strong
> 10	> 150	Very strong

Source: Kass & Raftery (1995).

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, $p = 0, \ldots, 6$. 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 sample 5000 final draws from the posteriors. The prior distributions are as defined in Section 2.

Table 5 reports the MHMs 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

Table 5: Model selection for MSIAH(2)-VAR(p) – determination of the lag order

Lags	0	1	2	3	4	5	6
MHM	-3,002.64	-2,926.42	-2,903.89	-2,898.21	**-2,895.22**	-2,914.87	-2,913.49

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 B.9 in Appendix B 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 Gibbs draws indicates that the Gibbs sampler has desirable properties.

Figure 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.

Comparing the best MS-VAR model to the best VAR model yields a posterior odds ratio of 6.41 in favour of the MS-VAR model, which in Table 4 enters the category of strong evidence for MS-VARs against VARs.

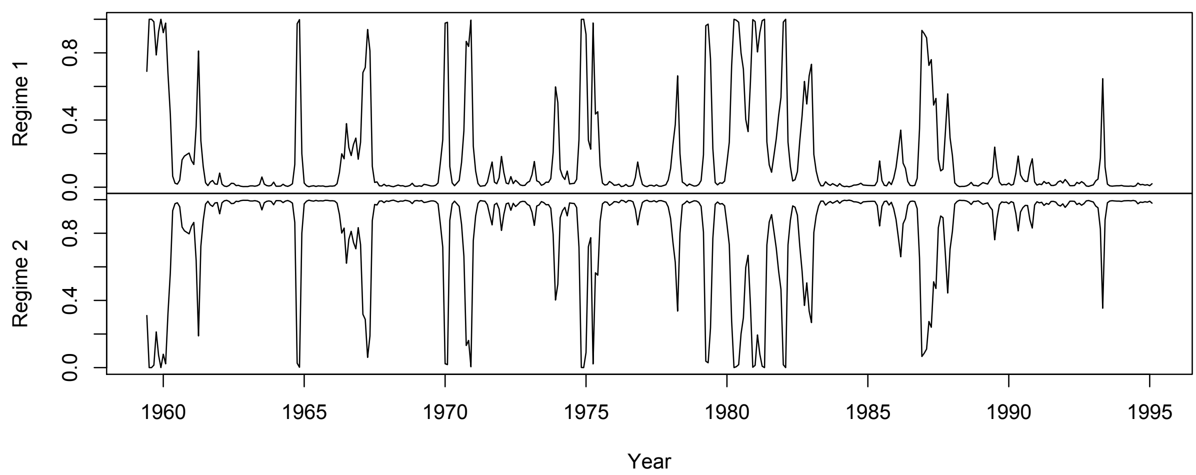

Figure 2: Estimated probabilities of regimes for a MSIAH(2)-VAR(4) model

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 different restrictions on the model parameters; we impose them, estimate the models and compute all marginal densities of data. The final test statistic to consider is the posterior odds ratio (which, with our prior probabilities of models, is equal to the Bayes factor), where the non-restricted MSIAH(2)-VAR(4) model constitutes the null hypothesis. Table 6 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, $\ln \mathcal{B}_{j1}$ for $j = 1, \dots, 7$. A positive logarithm of the Bayes factor is to be interpreted as evidence in favour of the restricted model. The scale of Table 4 is informative about the strength of the evidence. In a symmetric way, negative logarithm of the Bayes factor indicates that the non-restricted model is preferred by the data.

Analysis of Table 6 shows that only model $\mathcal{M}_5$ is more probable *a posteriori* than the unrestricted model $\mathcal{M}_1$. 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.

In order to asses the credibility of the hypotheses, each of which is represented by several sets of restrictions, and thus models, we compute Posterior Odds Ratios for the hypotheses in Table 7. 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:

$$\text{POR} = \frac{Pr(\mathcal{H}_i|\mathbf{y})}{Pr(\mathcal{H}_0|\mathbf{y})} = \frac{\sum_{j\in\mathcal{H}_i} Pr(\mathcal{M}_j|\mathbf{y})Pr(\mathcal{M}_j)}{Pr(\mathcal{M}_0|\mathbf{y})Pr(\mathcal{M}_0)}. \tag{31}$$

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

Table 7 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, which cast doubt on his conclusion. However, Bayesian testing establishes this strong result, and the conditional mean of income is invariant to the history of money. Table 7 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

Table 6: Noncausality and conditional regime independence in a MSIAH(2)-VAR(4) model. Numerical efficiency results for these models are presented in table C.10 of Appendix C.

$\mathcal{M}_j$	Hypothesis	Restrictions	# restrictions	$\ln p(\mathbf{y}\vert\mathcal{M}_j)$	$\ln \mathcal{B}_{j1}$
	$\mathcal{H}_0$: *Unrestricted model*				
$\mathcal{M}_0$	MS(2)-VAR(4)	-	0	-2895.22	0
	$\mathcal{H}_1$: *History of money down 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	-	-	-	-
	Restriction 7 ≡ (A7)	-	-	-	-

for $i = 1, \ldots, p$.

mean within VAR models (which is equivalent to the noncausality in variance and in distribution) does not hold with MS-VAR models. Allowing non-linearity 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

Table 7: Summary of the hypotheses testing

$\mathcal{H}_i$	Hypothesis	Represented by models	$\ln \frac{Pr(\mathcal{H}_i\vert\mathbf{y})}{Pr(\mathcal{H}_0\vert\mathbf{y})}$
$\mathcal{H}_0$	Unrestricted model	$\mathcal{M}_0$	0
$\mathcal{H}_1$	History of money down 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

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.

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 income, it is still crucial for its modeling. Past observations of money improves the forecast of the state of the economy when modelled 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.

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 all the existing methods of computating 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 (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.

Appendix A. Alternative restrictions for Granger noncausality

The following restrictions were set by Warne (2000), and are all derived under the condition (A2) and rank($\mathbf{P}^{(2)}$) = M_2.

Restriction 5. Suppose that $\mathbf{P} = (\imath_{M_1}\pi^{(1)\prime} \otimes \mathbf{P}^{(2)})$ with 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 $j_2 \in \{1,\ldots,M_2\}$, $r \in \{1,2,3\}$, and $k \in \{1,\ldots,p\}$,

Restriction 6. Suppose that $\mathbf{P} = (\imath_{M_1}\pi^{(1)\prime} \otimes \mathbf{P}^{(2)})$ with 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

(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.

Appendix B. Summary of the posterior densities simulations

Table B.8: 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
$A_{12}^{(5)}$	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

	Mean	Std. dev.	Naive Std. error	Autocorr. lag 1	Autocorr. lag 10
$A_{22}^{(7)}$	-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

Table B.9: 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
$A^{(4)}_{11,1}$	-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					

	Mean	Std. dev.	Naive Std. error	Autocorr. lag 1	Autocorr. lag 10
$A^{(1)}_{11,2}$	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

Appendix C. Characterization of estimation efficiency

Table C.10: 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.10 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.10, 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 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.

References

Albert, J. H., & Chib, S. (1993). Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts. *Journal of Business & Economic Statistics*, *11*, 1–15.

Barnard, J., McCulloch, R., & Meng, X.-l. (2000). Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Applicationto Shrinkage. *Statistica Sinica*, *10*, 1281–1311.

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

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

Casella, G., & George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, *46*, 167–174.

Cheung, Y. W., & Ng, L. K. (1996). A causality-in-variance test and its application to financial market prices. *Journal of Econometrics*, *72*, 33–48.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association*, *90*, 1313–1321.

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

Christiano, L. J., & Ljungqvist, L. (1988). Money does granger-cause output in the bivariate money-output relation. *Journal of Monetary Economics*, *22*, 217 – 235.

Christopoulos, D. K., & León-Ledesma, M. A. (2008). Testing for Granger (Non-) causality in a Time-Varying Coeffi cient. *Journal of Forecasting*, (pp. 293–303).

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

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

van Dijk, D., Osborn, D. R., & Sensier, M. (2005). Testing for causality in variance in the presence of breaks. *Economics Letters*, *89*, 193–199.

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

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

Friedman, M., & Schwartz, A. (1971). *A monetary history of the United States, 1867-1960* volume 12. Princeton Univ Pr.

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *Econometrics Journal*, *7*, 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*, pp. 1317–1339.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. Bernardo, J. Berger, A. Dawid, & A. Smith (Eds.), *Bayesian Statistics 4* (pp. 169–194). Clarendon Press, Oxford, UK.

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*, 1–73.

Geweke, J. (2010). *Complete and Incomplete Econometric Models*. Princeton University Press.

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

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

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

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, *57*, 357–384.
Hamilton, J. D. (1994). State-Space Models. In R. F. Engle, & D. L. McFadden (Eds.), *Handbook of Econometrics* chapter 50. (pp. 3039–3080). Elsevier. (Volume iv ed.).
Hong, Y. (2001). A test for volatility spillover with application to exchange rates. *Journal of Econometrics*, *103*, 183–224.
Hoogerheide, L. F., van Dijk, H. K., & van Oest, R. (2009). Simulation Based Bayesian Econometric Inference: Principles and Some Recent Computational Advances. Handbook of Computational Econometrics chapter 7. (pp. 215–280). Wiley.
Jeantheau, T. (1998). Strong consistency of estimators for multivariate arch models. *Econometric Theory*, *14*, 70–86.
Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.
Kim, C.-J., & Nelson, C. R. (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*, 608–616.
Kim, C.-J., & Nelson, C. R. (1999b). *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*. MIT press.
Koop, G., & Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, *3*, 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*, 109–127.
Lindgren, G. (1978). Markov Regime Models for Mixed Distributions and Switching Regressions. *Scandinavian Journal of Statistics*, *5*, 81–91.
Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*. Springer-Verlag.
Lütkepohl, H., & Burda, M. M. (1997). Modified Wald tests under nonregular conditions. *Journal of Econometrics*, *78*, 315–332.
McCulloch, R. E., & Tsay, R. S. (1994). Statistical analysis of economic time series via markov switching models. *Journal of Time Series Analysis*, *15*, 523–539.
Pantelidis, T., & Pittis, N. (2004). Testing for Granger causality in variance in the presence of causality in mean. *Economics Letters*, *85*.
Psaradakis, Z., Ravn, M. O., & Sola, M. (2005). Markov switching causality and the money-output relationship. *Journal of Applied Econometrics*, *20*, 665–683.
Psaradakis, Z., & Sola, M. (1998). Finite-sample properties of the maximum likelihood estimator in autoregressive models with Markov switching. *Journal of Econometrics*, *86*, 369–386.
Psaradakis, Z., & Spagnolo, N. (2003). On the determination ot the number of regimes in markov-switching autoregressive models. *Journal of Time Series Analysis*, *24*, 237–252.
Ritter, C., & Tanner, M. A. (1992). Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of the American Statistical Association*, *87*, 861– 868.
Robins, R. P., Granger, C. W. J., & Engle, R. F. (1986). Wholesale and Retail Prices: Bivariate Time-Series Modeling with forecastable Error Variances. In *Model Reliability* (pp. 1–17). The MIT Press.
Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688 – 701.
Sims, C. A. (1972). Money, Income, and Causality. *The American Economic Review*, *62*, 540 – 552.
Sims, C. A., Waggoner, D. F., & Zha, T. (2008). Methods for inference in large multiple-equation markov-switching models. *Journal of Econometrics*, *146*, 255–274.
Taylor, J. B., & Williams, J. C. (2009). A Black Swan in the Money Market. *American Economic Journal: Macroeconomics*, *1*, 58–83.
Tobin, J. (1970). Money and Income: Post Hoc Ergo Propter Hoc? *The Quarterly Journal of Economics*, *84*, 301–317. ArticleType: research-article / Full publication date: May, 1970 / Copyright © 1970 Oxford University Press.

Warne, A. (2000). *Causality and Regime Inference in a Markov Switching VAR*. Working Paper Series 118 Sveriges Riksbank (Central Bank of Sweden).
Woźniak, T. (2011a). Granger causal analysis of varma-garch models.
Woźniak, T. (2011b). Testing causality between two vectors in multivariate garch models.