GRANGER CAUSALITY AND REGIME INFERENCE IN MARKOV SWITCHING VAR MODELS WITH BAYESIAN METHODS

MATTHIEU DROUMAGUET,[a] ANDERS WARNE[b] AND TOMASZ WOŹNIAK[c]*

[a] *Department of Economics, European University Institute, Florence, Italy*
[b] *Directorate General Research, European Central Bank, Frankfurt, Germany*
[c] *Department of Economics, University of Melbourne, Victoria, Australia*

SUMMARY

In this paper, we derive restrictions for Granger noncausality in MS-VAR models and show under what conditions a variable does not affect the forecast of the hidden Markov process. To assess the noncausality hypotheses, we apply Bayesian inference. The computational tools include a novel block Metropolis–Hastings sampling algorithm for the estimation of the underlying models. We analyze a system of monthly US data on money and income. The results of testing in MS-VARs contradict those obtained with linear VARs: the money aggregate M1 helps in forecasting industrial production and in predicting the next period's state. Copyright © 2016 John Wiley & Sons, Ltd.

⌨ *Supporting information may be found in the online version of this article.*

1. INTRODUCTION

The concept of Granger causality was introduced by Granger (1969) and is based on the idea that variable x which causes another variable y should precede it. This idea has been formalized such that x is said *not* to Granger cause y if past and current information about x does not improve the forecast of y in a mean square error sense; see also Sims (1972). Granger causality has primarily been studied empirically in vector autoregressions (VARs) with a focus on one-step-ahead forecasts; see, for example, Lütkepohl (1993). In such a setting, x does not Granger cause y if the coefficients on lags of x in the equation for y are jointly zero. Knowledge of Granger causal relations may allow a researcher to formulate an appropriate model and obtain a better forecast of variables of interest. Note that this concept refers to the forecasting of variables, in contrast to, for example, the causality concept attributed to Rubin (1974), based on *ceteris paribus* effects (for a comparison of the two concepts, see, for example, Lechner, 2011). We also underline that, in general, Granger causality does not relate to any causal relation implied by structural economic theories either. Such correspondence has only been shown for linear models by Dufour and Tessier (1993).

Psaradakis *et al.* (2005) use a Markov switching (MS) VAR model to analyze *temporary Granger causality* within the money-income system, i.e. causality which holds in some periods but not in others. Technically, this means that they condition the causality analysis on realizations of the hidden Markov process and therefore focus only on linear relations between variables. That is, x does not Granger cause y temporarily in a certain state if the coefficients on lags of x in the equation for y are zero in that state. Since *all* parameters of an MS-VAR model may, in principle, vary with the hidden Markov process, the analysis in Psaradakis *et al.* (2005) neglects the possibility that x may be useful for predicting the states that affect the coefficients in the y equation.

---

* Correspondence to: Tomasz Woźniak, Department of Economics, University of Melbourne, 111 Barry Street, FBE Building, Level 4, 3053 Carlton, Victoria, Australia. E-mail: tomasz.wozniak@unimelb.edu.au

The approach to Granger causality that we consider in this paper takes into account the two sources of predictive relations between the variables of interest: first, the linear relations in the VAR model conditional on the states; and second, the fact that all of the variables are used to forecast the future probabilities of the states. The analysis of Granger causality is consequently not conditioned on the realizations of the hidden Markov process, but only on observed variables.

As a second contribution, we suggest an approach for performing Bayesian inference that allows us to test all of the restrictions of Granger noncausality jointly. The proposed framework consists of Bayesian estimation of the unrestricted model, allowing for Granger causality, and of the restricted models representing hypotheses of noncausality. For this purpose, we construct a novel block Metropolis–Hastings sampling algorithm that allows for the estimation of all these models. The hypotheses of Granger causality and noncausality can thereafter be evaluated with standard Bayesian methods using posterior odds ratios and Bayes factors.

An advantage of our approach is that we can test a hypothesis represented by several restricted models jointly. Another feature of the posterior odds analysis is that all the hypotheses are treated symmetrically. As a consequence, this method gives arguments *in favor of* a hypothesis. Finally, since a mixture model is a special case of an MS model, our analytical and empirical results apply also to such models.

The remainder of the paper is organized as follows. In Section 2 we present the model and some useful definitions, while Section 3 provides multiple restrictions for the considered relations between variables. Section 4 thereafter discusses Bayesian inference in the context of multiple hypotheses. The block Metropolis–Hastings algorithm used for estimating the models is described in Section 5. Section 6 gives an empirical application of the methodology, using the bivariate money-income system for monthly US data, and Section 7 concludes. The proofs and technical details are presented as supporting information in an online Appendix.[1]

2. THEORETICAL FRAMEWORK FOR GRANGER CAUSALITY AND REGIME INFERENCE

2.1. A Markov Switching VAR Model

Let $\mathbf{y}_T = (y_1, \ldots, y_T)$ denote a time series of T observations, where each y_t is an N-variate real-valued vector for $t \in \{1, \ldots, T\}$. Following Hamilton (1989), we consider a class of parametric Markov mixture distribution models in which the stochastic process y_t depends on the realizations of a hidden discrete stochastic process s_t with finite state space $\{1, \ldots, M\}$. Conditional on the regime, s_t, and the time series up to time $t - 1$, denoted by $\mathbf{y}_{t-1}$, y_t follows an independent identical normal distribution. Specifically, we let

$$y_t = \mu_{s_t} + \sum_{k=1}^{p} A_{s_t}^{(k)} y_{t-k} + \epsilon_t, \quad \epsilon_t | s_t \sim \text{ i.i. } \mathcal{N}(\mathbf{0}, \Sigma_{s_t}) \tag{1}$$

for $t \geq 1$. 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.

The variable s_t is assumed to be an irreducible aperiodic Markov chain with $\Pr(s_0 = i | \mathbf{P}) = \pi_i$, where $\pi = (\pi_1, \ldots, \pi_M)$ is the ergodic distribution of the MS process. It follows that $\pi_i > 0$ for all $i \in \{1, \ldots, M\}$ and $\pi' \iota_M = 1$, where ι_M is an M-dimensional vector with ones. The properties of the

---

[1] The online Appendix provides the general theoretical framework for studying Granger noncausality and regime inference in MS-VAR models with any number of variables and regimes, whereas the bivariate case with at most three regimes is discussed in the paper. The Appendix also gives the proofs of the propositions, technical details on the posterior samplers that are suggested in the paper and additional empirical results.

regime variable s_t are sufficiently described by the $(M \times M)$ transition probabilities matrix $\mathbf{P}$, with typical element p_{ij}, denoting the probability of a transition from state i to state j, $p_{ij} = \Pr(s_{t+1} = j | s_t = i)$. The elements of each row of $\mathbf{P}$ sum to unity, $\sum_{j=1}^{M} p_{ij} = 1$.

Furthermore, 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, $\theta_{s_t} \colon \mu_{s_t}, A_{s_t}^{(k)}, \Sigma_{s_t}$, for $k \in \{1, \ldots, p\}$ and $s_t \in \{1, \ldots, M\}$, and Θ is the parameter space of θ. To clarify the ideas presented in the paper, the analysis below is based on the simplifying assumptions that $N = 2$ and $M \leq 3$, while the general case is discussed in the online Appendix.

2.2. Defining Granger Causality and Noncausality

With $N = 2$ we let $y_t = [y_{1t}\ y_{2t}]'$ and define $\mathbf{y}_{it}$ such that it holds the observations of y_{it} up to period t for $i \in \{1, 2\}$. Let $u_{1.t+1}$ denote the one-step-ahead forecast error of $y_{1.t+1}$, conditional on $\mathbf{y}_t$ (and the parameters) when the predictor is given by the conditional expectation, i.e.

$$u_{1.t+1} = y_{1.t+1} - E[y_{1.t+1}|\mathbf{y}_t; \theta] \tag{2}$$

By construction, $u_{1.t+1}$ has conditional mean zero and non-negative variance. In what follows, we assume that the forecast error $u_{1.t+1}$ is mean zero stationary. Similarly, let $\tilde{u}_{1.t+1} = y_{1.t+1} - E[y_{1.t+1}|\mathbf{y}_{1t}, ; \theta]$ be the one-step-ahead forecast error of $y_{1.t+1}$ conditional on $\mathbf{y}_{1t}$ only, with comparable properties with respect to the smaller information set.

The definition of *Granger causality*, originally given by Granger (1969), states simply that y_2 is not causal for y_1 when past and current information about y_2 cannot improve the mean square forecast error of $y_{1.t+1}$.

Definition 1. y_2 does not Granger cause y_1 if and only if

$$E\left[u_{1.t+1}^2; \theta\right] = E\left[\tilde{u}_{1.t+1}^2; \theta\right], \quad t \geq 0 \tag{3}$$

It is well known that the variance of $u_{1.t+1}$ is smaller than or equal to the variance of $\tilde{u}_{1.t+1}$. Specifically, it holds that $E\left[\tilde{u}_{1.t+1}^2; \theta\right] = E\left[u_{1.t+1}^2; \theta\right] + E\left[(E[y_{1.t+1}|\mathbf{y}_t; \theta] - E[y_{1.t+1}|\mathbf{y}_{1t}; \theta])^2; \theta\right]$. From this relationship we see that the variances of the two forecast errors are equal when the second term on the right-hand side is zero, i.e. when the two forecasts and, thus, the two forecast errors are equal for all time periods.

It is important to note that the definition involves conditioning on the parameters. Since equation (3) will generally not hold for just any parameter value, it implies that the parameters must satisfy some restrictions. With this in mind, let the system in equation (1) be expressed as

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1.s_t} \\ \mu_{2.s_t} \end{bmatrix} + \sum_{k=1}^{p} \begin{bmatrix} A_{11.s_t}^{(k)} & A_{12.s_t}^{(k)} \\ A_{21.s_t}^{(k)} & A_{22.s_t}^{(k)} \end{bmatrix} \begin{bmatrix} y_{1.t-k} \\ y_{2.t-k} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \tag{4}$$

The 2×2 covariance matrix of the residuals conditional on the regime can be likewise decomposed with typical element $\Sigma_{ij.s_t}$ for $i, j \in \{1, 2\}$.

For expositional purposes, let us first assume that all regimes are known. The prediction of $y_{1.t+1}$ conditional on s_{t+1} and $\mathbf{y}_t$ is

$$E[y_{1.t+1}|s_{t+1}, \mathbf{y}_t; \theta] = y_{1.t+1} - \epsilon_{1.t+1} \tag{5}$$

Accordingly, the forecast error is given by $\epsilon_{1.t+1}$ and the conditional forecast error variance by $\Sigma_{11.s_{t+1}}$. The necessary and sufficient condition for y_2 not to Granger cause y_1 is that $A_{12.s_t}^{(k)}$ in equation (4) is equal to zero, for all lags and regimes.

Let us now drop the assumption that the regimes are known. While the regime variable s_{t+1} conditional on s_t is independent of $\mathbf{y}_t$, it can be predicted using only $\mathbf{y}_t$. Let $\Pr[s_{t+1}|\mathbf{y}_t, \theta]$ denote the probability of a particular state occurring at $t+1$ conditional on the information available at t. The prediction of $y_{1.t+1}$ is then given by

$$E\left[y_{1.t+1}|\mathbf{y}_t; \theta\right] = \sum_{s_{t+1}=1}^{M} E\left[y_{1.t+1}|s_{t+1}, \mathbf{y}_t; \theta\right] \Pr\left[s_{t+1}|\mathbf{y}_t; \theta\right] \tag{6}$$

The role for y_2 is different in equation (6) relative to equation (5) in that the history of y_2 can also predict y_1 by containing information which helps predict the next period's state.

Since s_{t+1} conditional on s_t is independent of $\mathbf{y}_t$, it follows that

$$\Pr\left[s_{t+1}|\mathbf{y}_t; \theta\right] = \sum_{s_t=1}^{M} \Pr\left[s_{t+1}|s_t; \theta\right] \Pr\left[s_t|\mathbf{y}_t; \theta\right] \tag{7}$$

From this relationship we may conjecture that there are only two instances when there is no additional information in the history of y_2 for predicting the next period's state. The first is when $\Pr[s_{t+1}|s_t; \theta] = \Pr[s_{t+1}; \theta]$, i.e. the Markov process is serially uncorrelated. The second case occurs when $\Pr[s_t|\mathbf{y}_t, \theta] = \Pr[s_t|\mathbf{y}_{1t}; \theta]$.

With these observations in mind, it is useful to introduce the concept of predictive redundancy for MS-VAR models:

Definition 2. The time series $\mathbf{y}_{2t}$ is said to be *predictively redundant* for $s_{t+\tau}$ when

$$\Pr\left[s_{t+\tau} = j|\mathbf{y}_t, \theta\right] = \Pr\left[s_{t+\tau} = j|\mathbf{y}_{1t}, \theta\right] \tag{8}$$

for $j \in \{1, \ldots, M\}$, $t \geq 0$, and where τ is an integer.

It is noteworthy that both cases discussed below equation (7) imply that $\mathbf{y}_{2t}$ is predictively redundant for s_{t+1}. The first case obtains this property at the cost of having a serially uncorrelated Markov process, i.e. a mixture model. The second case allows for a serially correlated Markov process at the cost of the condition that $\mathbf{y}_{2t}$ is also predictively redundant for s_t.

2.3. Comparison with the Approach by Psaradakis *et al.* (2005)

A feature that distinguishes the forecast errors in equations (2) and (5) is the conditioning of the forecast on the realization of the state, s_{t+1}, in the latter equation. This conditioning also differentiates our approach from the analysis of so-called *temporary Granger causality* by Psaradakis *et al.* (2005). For the sake of comparison, we propose a formal definition of their concept maintaining our notation:[2]

Definition 3. y_2 does not *temporarily* Granger cause y_1 in state j if and only if

$$E\left[\epsilon_{1.t+1}^2 | s_{t+1} = j, \mathbf{y}_t; \theta\right] = E\left[\tilde{\epsilon}_{1.t+1}^2 | s_{t+1} = j, \mathbf{y}_{1t}; \theta\right], \quad t \geq 0 \tag{9}$$

where $\tilde{\epsilon}_{1.t+1} = y_{1.t+1} - E\left[y_{1.t+1}|s_{t+1} = j, \mathbf{y}_{1t}; \theta\right]$.

---

[2] Psaradakis *et al.* (2005) do not provide a formal definition of the concept.

A comparison of Definitions 1 and 3, as well as equations (3) and (9), emphasizes some of the differences. Psaradakis *et al.* not only introduce conditioning on the realization of the unobserved next period state, but also they do not require that condition (9) holds for all time periods. Rather, this condition only needs to hold for the time periods when $s_{t+1} = j$. We can therefore deduce that the necessary and sufficient condition for y_2 not to *temporarily* Granger cause y_1 in state j is that $A_{12.j}^{(k)}$ is equal to zero for $k \in \{1, \dots, p\}$.

3. REGIME INFERENCE AND GRANGER CAUSALITY ANALYSIS

3.1. Regime Inference

The first result in this paper concerns the restrictions that the MS-VAR system needs to satisfy to guarantee that we can make optimal inference about the regime process s_{t+1} using only $\mathbf{y}_{1.t}$ and the parameters. Proofs of the propositions are given in the online Appendix.

Proposition 1. Given the MS-VAR system for y_t in equation (1) with $N = 2$, s_t satisfying the assumptions in Section 2.1, and $M \in \{2, 3\}$, $\mathbf{y}_{2t}$ is predictively redundant for s_{t+1} if and only if either:

(A1): $\mu_{2.s_t} = \mu_2$, $A_{21.s_t}^{(k)} = A_{21}^{(k)}$, $A_{22.s_t}^{(k)} = A_{22}^{(k)}$, $\Sigma_{22.s_t} = \Sigma_{22}$, $\Sigma_{12.s_t} = 0$, and $A_{12.s_t}^{(k)} = 0$ for $k \in \{1, \dots, p\}$ and $s_t \in \{1, \dots, M\}$; or

(A2): $\mathbf{P} = \iota_M \pi'$.

In addition, $\mathbf{y}_{2.t}$ is predictively redundant for s_t if and only if (A1) holds.

First, note that conditions (A1) and (A2) give linear restrictions on the parameters of the model. Condition (A2) states that the hidden Markov processes is serially uncorrelated and the distribution of ϵ_t is therefore mixed normal. As a consequence, $\mathbf{y}_{2.t}$ is (trivially) predictively redundant for s_{t+1} under (A2).

The intuition behind condition (A1) is in fact also straightforward. Suppose $p = 1$, $M = 2$, while ϵ_{2t} is i.i.d. and independent of ϵ_{1t} for all t. Now consider the experiment of drawing two y_{2t}'s, one for each regime, when $y_{1.t-1}$ and $y_{2.t-1}$ are fixed. The difference between these two draws is: $y_{2t|s_t=2} - y_{2t|s_t=1} = (\mu_{2,2} - \mu_{2,1}) + (A_{21,2} - A_{21,1}) y_{1.t-1} + (A_{22,2} - A_{22,1}) y_{2.t-1}$. The right-hand side of this equation is zero for all vectors $(y_{1.t-1}, y_{2.t-1})$ when the coefficients in the y_2 equation are constant across states. Accordingly, if these restrictions are satisfied, then $\Pr[s_t | \mathbf{y}_{1t}, \mathbf{y}_{2.t}] = \Pr[s_t | \mathbf{y}_{1t}, \mathbf{y}_{2.t-1}]$ and all information about s_t is found in the equation for y_1. If the coefficient on $y_{2.t-1}$ in that equation is zero for both regimes and, hence, that y_2 is temporarily Granger noncausal for y_1 in both regimes, then $\mathbf{y}_{2.t-1}$ plays no role for predicting s_t either.

Since $\mathbf{y}_{2.t}$ is predictively redundant for s_t, it follows from the Markov property of the regime process that $\mathbf{y}_{2.t}$ is also predictively redundant for s_{t+1}. In addition, the expected value of $y_{1.t+1}$ conditional on s_{t+1} and $\mathbf{y}_t$ on the right-hand side of equation (6) does not depend on $\mathbf{y}_{2.t}$, and it therefore follows that y_2 does not Granger cause y_1 in this situation. Hence (A1) is not only a sufficient condition for predictive redundancy but also for Granger noncausality.

3.2. Granger Noncausality Analysis

The analysis in the previous section has established that there is an interesting connection between the restrictions for predictive redundancy in (A1) and Granger noncausality. These restrictions are, however, only sufficient and in this section we shall focus on both necessary and sufficient conditions for this type of noncausal relations.

To examine Granger causal relations for MS-VARs in more detail, the one-step-ahead prediction of $y_{1.t+1}$ is given by

$$E\left[y_{1.t+1}\mid \mathbf{y}_t;\theta\right] = \bar{\mu}_{1.t} + \sum_{k=1}^{p}\left(\bar{A}_{11.t}^{(k)}y_{1.t+1-k} + \bar{A}_{12.t}^{(k)}y_{2.t+1-k}\right) \qquad (10)$$

where

$$\bar{\mu}_{1.t} = \sum_{i=1}^{M}\sum_{j=1}^{M}\mu_{1.j}\,p_{ij}\,\Pr\left[s_t = i\mid \mathbf{y}_t;\theta\right], \quad \bar{A}_{1r.t}^{(k)} = \sum_{i=1}^{M}\sum_{j=1}^{M}A_{1r.j}^{(k)}\,p_{ij}\,\Pr\left[s_t = i\mid \mathbf{y}_t;\theta\right] \qquad (11)$$

for $r \in \{1,2\}$, are the expected values of the random coefficients in the equation for y_1. From these equations it can be seen that $\bar{\mu}_{1.t}$ and $\bar{A}_{1r.t}^{(k)}$ depend on $\mathbf{y}_{2t}$ only via the filter probabilities $\Pr[s_t\mid \mathbf{y}_t;\theta]$. Furthermore, $\bar{A}_{12.t}^{(k)}$ must be zero to directly exclude $\mathbf{y}_{2.t}$ from the one-step-ahead forecasts of $y_{1.t+1}$.

This brings us to the main result about Granger noncausality for bivariate MS-VARs with at most three regimes.

Proposition 2. y_2 does not Granger cause y_1 when $M \in \{2,3\}$ if and only if either: (A1) or

(A3): (i) $\sum_{j=1}^{M}\mu_{1.j}\,p_{ij} = \bar{\mu}_1$, (ii) $\sum_{j=1}^{M}A_{1r.j}^{(k)}\,p_{ij} = \bar{A}_{1r}^{(k)}$, (iii) $\bar{A}_{12}^{(k)} = 0$, for all $i \in \{1,\dots,M\}$, $r \in \{1,2\}$, and $k \in \{1,\dots,p\}$,

is satisfied.

Since (A1) need not hold under (A3), it follows that the filter probabilities $\Pr[s_t = i\mid \mathbf{y}_t;\theta]$ are allowed to depend on $\mathbf{y}_{2.t}$. Consequently, to ensure that these filter probabilities can simply be integrated out of the one-step-ahead forecast of $y_{1.t+1}$, the nonlinear restrictions in conditions (A3)(i) and (A3)(ii) state that the expected value of each random coefficient in the equation for $y_{1.t+1}$ is constant. Condition (A3)(iii) thereafter sets each expected value of the coefficients on lags of y_2 to zero.

Note that the restrictions in (A3) do not take the properties of the transition matrix $\mathbf{P}$ into account. Hence they are very general and it is not possible to determine the number of restrictions or their actual form without specifically referring to the rank of the transition matrix.

To address this issue, condition (A2) in Proposition 1 sets the rank of the Markov transition matrix to unity. Such a mixture VAR model provides us with the following interesting result.

Corollary 1. Suppose that condition (A2) is satisfied, then condition (A3) is equivalent to

(A4): $\bar{A}_{12}^{(k)} = \sum_{j=1}^{M}A_{12.j}^{(k)}\pi_j = 0$, for all $k \in \{1,\dots,p\}$.

Corollary 1 provides the necessary and sufficient conditions for y_2 not to Granger cause y_1 in a mixture VAR model. The only requirement that such a model needs to satisfy is that the expected value of the random coefficients on lags of y_2 in the equation for y_1 are all zero.

Note also that all $A_{12.j}$ coefficients can be nonzero under (A4). This means that condition (A4) does not imply temporary Granger noncausality in at least one state. Likewise, temporary noncausality in one state does not imply that (A4) holds. Hence, while both types of noncausality may be of interest to examine in practice, there is not a logical link between them in Markov switching or mixture VAR models.

Corollary 1 covers some but not all of the cases when condition (A3) in Proposition 2 holds. With at most three regimes, the remaining two cases are: (i) $\mathbf{P}$ has full rank M; and (ii) rank$[\mathbf{P}] = 2$ when $M = 3$.

Let us first assume that $\mathbf{P}$ has full rank M. Turning first to (A3)(i) we see that it can be rewritten as $\sum_{j=1}^{M}\left(\mu_{1.j} - \bar{\mu}_1\right)p_{ij} = 0$, for $i \in \{1,\dots,M\}$. These M orthogonality conditions can only hold

for a full rank $(M \times M)$ matrix $\mathbf{P}$ when $\mu_{1.j} = \bar{\mu}_1 = \mu_1$ for all j, i.e. when the constant term in the y_1 equation does not vary with the regime. The same principle applies to (A3)(ii) such that these coefficients are also constant across regimes. Consequently, the $A_{12.j}^{(k)}$ coefficients are regime invariant, and (A3)(iii) can therefore be expressed as $\bar{A}_{12}^{(k)} = \sum_{j=1}^{M} A_{12}^{(k)} p_{ij} = A_{12}^{(k)} = 0$, for $k \in \{1, \ldots, p\}$. In other words, if the Markov transition matrix has full rank M, then the Granger noncausality restrictions in (A3) are all linear, stating that all coefficients in the equation for y_1 are constant across regimes, and that the coefficients on lags of y_2 are all zero.

Second, suppose that the number of regimes is three while the Markov transition matrix has rank two. Letting p_i. denote the ith row of $\mathbf{P}$, it can now be shown that

$$\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ c & 1-c \end{bmatrix} \begin{bmatrix} p_{1.} \\ p_{2.} \end{bmatrix} \tag{12}$$

where rank$[\mathbf{P}] = 2$ is equivalent to stating that $p_{1.} \neq p_{2.}$. Note here that the rank of two implies one restriction on $\mathbf{P}$.

Letting $m_1 = [\mu_{1.1} \ \mu_{1.2} \ \mu_{1.3}]$, the restrictions (A3)(i) can be rewritten as

$$m_1 \mathbf{P}' = \begin{bmatrix} m_1 p'_{1.} & m_1 p'_{2.} & m_1 \left(c p'_{1.} + (1-c) p'_{2.} \right) \end{bmatrix} = \bar{\mu}_1 \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \tag{13}$$

These restrictions hold simultaneously when $m_1 p'_{1.} = m_1 p'_{2.}$, i.e. when the restriction $\sum_{j=1}^{3} \mu_{1.j} \left(p_{1j} - p_{2j} \right) = 0$ is satisfied. Since $p_{1j} = p_{2j}$ for all j is ruled out by the rank assumption, this relationship yields one orthogonality restriction that the μ_{1j} parameters satisfy. Similarly, the restrictions (A3)(ii) are satisfied when $\sum_{j=1}^{3} A_{1r.j}^{(k)} (p_{1j} - p_{2j}) = 0$ for $r \in \{1, 2\}$ and $k \in \{1, \ldots, p\}$. Finally, (A3)(iii) holds when $\sum_{j=1}^{3} A_{12.j}^{(k)} p_{1j} = 0$. The Granger noncausality and the predictive redundancy restrictions are, for convenience, summarized in Table I for the case of $M = 3$.

One important implication of these theoretical Granger noncausality results for MS-VAR models is that the overall hypothesis is consistent with several individual sub-hypotheses. A standard requirement in classical analysis of a multiple hypotheses is that each individual sub-hypothesis is *minimal*

Table I. Summary of regime inference and Granger noncausality restrictions on the parameters of MS-VAR models with $M = 3$

	No. of restrictions
Condition (A1)	
$\mathcal{M}_1 :$ $\mu_{2.s_t} = \mu_2, A_{21.s_t}^{(k)} = A_{21}^{(k)}, A_{22.s_t}^{(k)} = A_{22}^{(k)}, A_{12.s_t}^{(k)} = 0$, for $k \in \{1, \ldots, p\}$ and $\Sigma_{22.s_t} = \Sigma_{22}, \Sigma_{12.s_t} = 0$	$7p + 7$
Condition (A2)	
$\mathcal{M}_2 :$ $\mathbf{P} = \iota_3 \pi'$	4
Condition (A3)	
$\mathcal{M}_3 :$ $\mathbf{P} = \iota_3 \pi'$, and $\sum_{j=1}^{3} A_{12.j}^{(k)} \pi_j = 0$, for $k \in \{1, \ldots, p\}$	$p + 4$
$\mathcal{M}_4 :$ $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ c & 1-c \end{bmatrix} \begin{bmatrix} p_{1.} \\ p_{2.} \end{bmatrix}, \sum_{j=1}^{3} \mu_{1.j} (p_{1j} - p_{2j}) = 0, \sum_{j=1}^{3} A_{11.j}^{(k)} (p_{1j} - p_{2j}) = 0$ and $\sum_{j=1}^{3} A_{12.j}^{(k)} p_{ij} = 0$ for $i \in \{1, 2\}$ and $k \in \{1, \ldots, p\}$	$3p + 2$
$\mathcal{M}_5 :$ $\mu_{1.s_t} = \mu_1, A_{11.s_t}^{(k)} = A_{11}^{(k)}$, and $A_{12.s_t}^{(k)} = 0$, for $k \in \{1, \ldots, p\}$	$5p + 2$

Note: ι_n is an n-dimensional vector of ones, p_i. is the ith row of $\mathbf{P}$, π is a vector of ergodic state probabilities, and c is a parameter estimated by maximizing the value of the full conditional posterior density of $\mathbf{P}$.

in the sense of Gabriel (1969). That is, a hypothesis is said to be minimal if it does not imply the truth of any other hypothesis in a multiple hypotheses set. Concerning the Granger noncausality restrictions that we have obtained, the case of two regimes gives us three sub-hypotheses, and the case of three regimes gives us four sub-hypotheses. It is straightforward to show that all these sets of restrictions are minimal in the sense of Gabriel (1969).

4. BAYESIAN TESTING

Restrictions (A1)–(A4) can be tested using either classical or Bayesian inference, and in this study we make use of the latter. The approach we suggest can deal with both multiple hypotheses and nonlinearities, issues that are known to often make classical inference cumbersome; see Droumaguet *et al.* (2015) for a discussion on these issues in MS-VARs, as well as Woźniak (2015, 2016), where Granger noncausality is studied within the extended CCC-GARCH model of Jeantheau (1998).

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, $\mathcal{H}_i = \{ j : \mathcal{M}_j$ represents the ith hypothesis$\}$. The models that are included in this set are assumed to be minimal. Furthermore, suppose that we are interested in comparing the posterior probability of this hypothesis to hypothesis $\mathcal{H}_0$, represented by the unrestricted model $\mathcal{M}_0$. The credibility of hypothesis $\mathcal{H}_i$ compared to hypothesis $\mathcal{H}_0$ may then be assessed with the posterior odds ratio given by

$$\text{POR} = \frac{\Pr(\mathcal{H}_i \,|\, \mathbf{y}_T)}{\Pr(\mathcal{H}_0 \,|\, \mathbf{y}_T)} = \frac{\sum_{j \in \mathcal{H}_i} \Pr(\mathcal{M}_j \,|\, \mathbf{y}_T)}{\Pr(\mathcal{M}_0 \,|\, \mathbf{y}_T)} = \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_0)} \cdot \frac{p(\mathbf{y}_T \,|\, \mathcal{H}_i)}{p(\mathbf{y}_T \,|\, \mathcal{H}_0)} \tag{14}$$

If we set equal prior probabilities for all the hypotheses, then the posterior odds ratio is equal to the Bayes factor and is given by the ratio of marginal data densities (MDDs) when conditioning on $\mathcal{H}_i$ and $\mathcal{H}_0$, respectively.

The MDD is typically calculated for a model $\mathcal{M}_j$ rather than for a hypothesis $\mathcal{H}_i$. To determine the MDD for the hypothesis $\mathcal{H}_i$ using the MDDs for the models that are consistent with it, we can utilize the following: $p(\mathbf{y}_T \,|\, \mathcal{H}_i) = \sum_{j \in \mathcal{H}_i} p(\mathbf{y}_T \,|\, \mathcal{M}_j) \Pr(\mathcal{M}_j \,|\, \mathcal{H}_i)$. If we assume that all models $\mathcal{M}_j$ are equally likely a priori given that the hypothesis $\mathcal{H}_i$ is true, then the MDD given the hypothesis is equal to the average of the MDDs for the models.

For example, as discussed in Section 3.2, condition (A3) depends on the rank of $\mathbf{P}$ and this value is not of interest from the point of view of the Granger causality testing, but nevertheless affects the restrictions. In this case, we can integrate out the rank of $\mathbf{P}$ from the testing problem by computing

$$p(\mathbf{y}_T \,|\, (\text{A3})) = \sum_{i=1}^{M} p(\mathbf{y}_T \,|\, (\text{A3}) \wedge \text{rank}(\mathbf{P}) = i) \Pr(\text{rank}(\mathbf{P}) = i \,|\, (\text{A3})) \tag{15}$$

where $p(\mathbf{y}_T \,|\, (\text{A3}) \wedge \text{rank}(\mathbf{P}) = i)$ are estimated by a MDD estimator, whereas $\Pr(\text{rank}(\mathbf{P}) = i \,|\, (\text{A3})) = 1/M$ is a possible choice for the conditional prior probabilities. Furthermore, the Granger noncausality hypothesis is valid when either (A1) or (A3) is true. Since the former sub-hypothesis only involves one model and the latter sub-hypothesis M models, one option is to set the prior weight of (A1) to $1/(M+1)$, while the weight of (A3) is $M/(M+1)$.

5. BLOCK METROPOLIS–HASTINGS SAMPLER FOR MS-VAR MODELS

This section describes the likelihood function, prior distributions and algorithm for the estimation of the unrestricted and restricted models; the details of the algorithm are given in the online Appendix. The parametrization allows for the estimation of all the restricted models, where the restrictions on the parameters are given by the regime inference restrictions in Proposition 1 or the Granger noncausality restrictions in Proposition 2.

The complete-data likelihood function is equal to the joint sampling distribution $p(\mathbf{S}_T, \mathbf{y}_T | \theta)$ for the complete data $(\mathbf{S}_T, \mathbf{y}_T)$ given the parameters, where $\mathbf{S}_T = (s_0, s_1, \ldots, s_T)$; see, for example, Frühwirth-Schnatter (2006). This likelihood is further decomposed into $p(\mathbf{S}_T, \mathbf{y}_T | \theta) = p(\mathbf{y}_T | \mathbf{S}_T, \theta) \Pr(\mathbf{S}_T | \mathbf{P})$, where the parameters $\theta = (\theta_1, \ldots, \theta_M, \mathbf{P})$. The two components on the right-hand side are the same as the standard MS-VAR model of Frühwirth-Schnatter (2006, Section 11.3.1).

5.1. Prior Analysis

We assume that the prior distribution of the state-specific parameters for each state and the transition probabilities matrix are independent: $p(\theta) = \prod_{i=1}^{M} p(\theta_i) p(\mathbf{P})$. For the unrestricted MS-VAR 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 p_{ii}, for $i \in \{1, \ldots, M\}$, for which it is set to the hyperparameter $\lambda_{\mathbf{P}}$, i.e. $\mathbf{P}_{i\cdot} \sim \mathcal{D}_M (\iota'_M + (\lambda_p - 1) I_{M.i\cdot})$. If $\lambda_{\mathbf{P}} > 1$, the regimes are persistent over time.

Furthermore, the state-dependent parameters of the VAR process are collected in vectors $\beta_{s_t} = \left(\mu'_{s_t}, \mathrm{vec}\left(A_{s_t}^{(1)}\right)', \ldots, \mathrm{vec}\left(A_{s_t}^{(p)}\right)' \right)'$, for $s_t \in \{1, \ldots, M\}$. These parameters follow a priori a $(N + pN^2)$-variate normal distribution, with mean equal to zero and a diagonal covariance matrix, $V_\beta: \beta_i \sim \mathcal{N}(\mathbf{0}, V_\beta)$.

The prior distribution nests many popular prior specifications in empirical macroeconomics, including shrinkage priors, and can be extended to hierarchical priors. Furthermore, the mean vector of the normal prior distribution of parameters β_{s_t} is set to zero since in the empirical analysis we use first differences of the log-levels of the observed variables (see Section 6).[3]

Let σ_{s_t} be a two-dimensional vector of standard deviations and $\mathbf{R}_{s_t}$ a 2×2 correlation matrix. The covariance matrix of $\epsilon_t | s_t$ can now be decomposed as $\Sigma_{s_t} = \mathrm{diag}(\sigma_{s_t}) \mathbf{R}_{s_t} \mathrm{diag}(\sigma_{s_t})$, Modeling covariance matrices using a decomposition into standard deviations and a correlation matrix was proposed for Bayesian inference by Barnard *et al.* (2000). This approach is adapted to MS-VAR models in this paper since it allows us to impose restrictions on the covariance matrix in a convenient manner. The unrestricted model is treated analogously, thereby allowing for direct comparisons of the prior distributions for the unrestricted and the restricted models. Thus each standard deviation $\sigma_{j.s_t}$ for $s_t \in \{1, \ldots, M\}$ and $j \in \{1, 2\}$ follows a log-normal distribution, with a mean parameter equal to 0 and the standard deviation parameter set to $\lambda_\sigma > 0$, such that $\sigma_{j.i} \sim \log \mathcal{N}(0, \lambda_\sigma)$.

Finally, we assume that the marginal prior distribution for the below-diagonal element of the (symmetric) correlation matrix $\mathbf{R}_{s_t}$ is a uniform distribution on the interval $(-1, 1)$, and is denoted by $\mathbf{R}_{21.s_t} \sim \mathcal{U}(-1, 1)$.[4] We collect all the standard deviations in one vector, $\sigma = (\sigma'_1, \ldots, \sigma'_M)'$, and all the unknown correlation coefficients into a vector, $\mathbf{R} = (\mathrm{vecl}(\mathbf{R}_1)', \ldots, \mathrm{vecl}(\mathbf{R}_M)')'$, where the vecl operator stacks all the lower-diagonal elements of the correlation matrix into a vector.

5.2. Posterior Simulations

In this section we discuss the estimation algorithm in broad terms, while the technical details are given in the online Appendix. The algorithm is designed to facilitate the estimation of unrestricted

---

[3] If the observed variables are instead measured in log-levels, then the prior mean of β_{s_t} can be specified such that it contains, for example, unit values for the diagonal elements of the matrix $A_{s_t}^{(1)}$ and zeros elsewhere for each regime (see Doan *et al.*, 1983).

[4] Barnard *et al.* (2000) discuss the implications of such a prior specification. However, histograms of the marginal posterior distributions for the correlation coefficients simulated in our empirical example have regular shapes. Therefore, no significant unfavorable effects of such prior assumptions are noticed. Barnard *et al.* (2000) mentions alternative prior distributions that could be used as well.

and restricted MS-VAR models. In particular, it allows some of the covariances of the error term to be constrained by restricting corresponding correlation coefficients, to set some of the autoregressive parameters to zero, to estimate a reduced rank transition probabilities matrix, to make some groups of parameters state-invariant, as well as to impose nonlinear restrictions on the autoregressive parameters. Combinations of such restrictions are needed to estimate the models that are consistent with predictive redundancy or Granger noncausality.

In the block Metropolis–Hastings algorithm, the parameters are split into sub-vectors for which the full conditional densities are of a convenient form. The integer l denotes the iteration of the Markov chain Monte Carlo (MCMC) sampler, such as $\theta^{(l)}$ for the θ vector. The initial value of θ is computed with the EM algorithm and is denoted by $\theta^{(0)}$.

1. Draw the realizations of s_t, for $t = 1, \ldots, T$, using the probabilities $\Pr(s_t = i | \mathbf{y}_T, \theta^{(l-1)})$, for $i = 1, \ldots, M$, obtained with the regime smoother, and create $\mathbf{S}_T^{(l)}$, for the lth iteration of the algorithm.

2. Draw from the posterior distribution of the transition probabilities matrix conditioning on the states drawn in the previous step of the current iteration, $\mathbf{P}^{(l)} \sim p(\mathbf{P}|\mathbf{S}_T^{(l)})$. Given the assumptions of a Dirichlet prior distribution of $\mathbf{P}$ and that the Markov process starts from its ergodic distribution, $\Pr[s_0] = \pi$, the posterior distribution is not of a standard form. In this step of the MCMC sampler, we use the Metropolis–Hastings algorithm as described in the online Appendix.

3. Draw from the full conditional distributions of the σ and the $\mathbf{R}$ parameters, denoted by $p\left(\sigma | \mathbf{y}_T, \mathbf{S}_T^{(l)}, \mathbf{P}^{(l)}, \beta^{(l-1)}, \mathbf{R}^{(l-1)}\right)$ and $p\left(\mathbf{R} | \mathbf{y}_T, \mathbf{S}_T^{(l)}, \mathbf{P}^{(l)}, \beta^{(l-1)}, \sigma^{(l)}\right)$, respectively, with the Griddy–Gibbs sampling algorithm of Ritter and Tanner (1992) and described by Barnard *et al.* (2000).

4. 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 for unrestricted models (or for models with zero restrictions for some of the elements of β) is also normal, $f\left(\beta | \mathbf{y}_T, \mathbf{S}_T^{(l)}, \mathbf{P}^{(l)}, \sigma^{(l)}, \mathbf{R}^{(l)}\right) = \mathcal{N}\left(\bar{\beta}^*, \bar{V}_{\beta^*}\right)$. $\bar{\beta}^*$ and $\bar{V}_{\beta^*}$ are the parameters of the full conditional distribution specified in the online Appendix.[5]

6. GRANGER NONCAUSALITY ANALYSIS OF US MONEY AND INCOME

In many empirical studies of causal relations between money and income the focus is on US data. At the heart of this issue is the empirical analysis conducted in Friedman and Schwartz (1971) asserting that money changes lead 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 debate, since economists have not reached a consensus on this topic. This historical debate is well narrated by Psaradakis *et al.* (2005), and the interested reader is advised to consult that paper for a depiction of events.

From a more recent perspective, the financial crisis of 2007–8 and the Great Recession that followed also provides a background for looking into the issue of Granger causality from money to income. With conventional monetary policy quickly becoming constrained by the zero lower bound

---

[5] The Metropolis–Hastings step for the case when some of the elements of β are restricted to be nonlinear functions of other parameters is also described in the online Appendix.

on nominal interest rates, unconventional monetary policy measures were instead implemented by major central banks in an attempt to stimulate the economy by lowering longer-term interest rates. The quantitative easing (QE) programs of the Federal Reserve System that were pursued from 2008 and onwards in various stages also led to a substantial increase in the growth of M1, for example. According to Joyce *et al.* (2012), the empirical evidence supports the view that such unconventional monetary policy has had an effect on economic activity, although the size and duration of this effect are uncertain.[6] In other words, the QE programs may be connected with, first, an increase of M1 growth and, thereafter, a rise in output growth. From a Granger causality perspective, this suggests that the history of money may contain useful information for improving the forecasts of income in a bivariate framework, although the forecast horizon for this improvement need not be the one-step-ahead horizon. In addition, the change from conventional to unconventional monetary policy is likely to be better reflected by a nonlinear regime switching model than a linear model.[7]

As pointed out by Psaradakis *et al.* (2005) in the motivation of their approach, there is a problem in the instability of the empirical results found for the causality between money and output. Depending on the samples considered, 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 restricted Markov switching VAR model that assumes four states of the economy: (1) both variables may cause each other; (2) money does not cause output; (3) output does not cause money; (4) none of the variables causes the other.

Warne (2000) and Psaradakis *et al.* (2005) use bivariate MS-VAR models to study Granger causality for US data on money and income.[8] As outlined in the Introduction, with the approach of Warne (2000) which we follow, the benchmark MS-VAR models are unrestricted and we perform Bayesian model selection through the comparison of their MDDs to determine the number of states as well as the number of autoregressive lags. Moreover, we make use of the original definition of Granger noncausality (see Definition 1), applied to MS-VAR models. Below, we use the Bayesian testing apparatus to investigate the money–income relationship once again.

6.1. Data

The data are similar to those used by Christiano and Ljungqvist (1988) and Warne (2000), but the sample is longer and spans a period of 53 years. Two monthly US time series are included—the M1 money stock and the industrial production index—a standard *proxy* for monthly income in Granger causality studies. Both variables have 646 observations covering the period from 1959:1 to 2012:11 and are taken from the Citibase database. The data are seasonally adjusted, transformed using first differences of the natural logarithm, and multiplied by 1200.

There are strong signs of time variation in the volatility of the series. In the money series, for example, a period of higher volatility starts around 1980 and ends before 1985. The period of the global financial crisis is also characterized by increased volatility in both series, especially after August 2008.[9]

---

[6] The interested reader may also refer to other articles in the same issue of the *Economic Journal*, as well as references therein.

[7] Sargent and Surico (2011) investigate two quantity theory of money propositions highlighted by Lucas (1980): a given change in the quantity of money gives rise to an equal change in the inflation rate; and the nominal interest rates. The authors document periods over which these two propositions are supported by US data and when they have broken down. They relate their findings to changes in the monetary policy rule of a dynamic stochastic general equilibrium model, especially with respect to the responses to inflationary pressures. Their set-up also includes observed data on output growth and thereby allows for an investigation of the implications of changes in the monetary policy rules to the relationship between money growth and output growth.

[8] Warne (2000) uses monthly data whereas Psaradakis *et al.* (2005) use quarterly data.

[9] The time series observations of money and income are plotted in Figure 1 of the online Appendix.

Table II. Model selection for VAR(p) and MS(M)-VAR(p) models over the sample period 1959:1–2012:11

VAR models									
Lags (p)	0	1	2	3	4	5	6	7	8
ln p_{MHM} ($\mathbf{y}_T \mid p$)	−4739.81	−4654.11	−4642.09	−4613.66	−4616.61	−4609.95	−4592.88	−4585.46	−4582.27
Lags (p)	9	10	11	12	13	14	15	16	17
ln p_{MHM} ($\mathbf{y}_T \mid p$)	−4581.74	−4584.36	−4556.16	−4553.18	−4551.15	**−4544.68**	−4546.68	−4549.14	−4546.53

MS-VAR models with number of states $M = 2$								
Lags (p)	0	1	2	3	4	5	6	7
ln p_{MHM} ($\mathbf{y}_T \mid p, M$)	−4578.72	−4456.06	−4440.07	−4412.75	−4418.02	−4420.24	−4411.57	−4411.62

MS-VAR models with number of states $M = 3$								
Lags (p)	0	1	2	3	4	5	6	7
ln p_{MHM} ($\mathbf{y}_T \mid p, M$)	−4567.22	−4415.06	−4402.98	**−4384.28**	−4390.89	−4392.09	−4387.6	-4390.79

Note: ln p_{MHM} ($\mathbf{y}_T \mid p$) denotes the marginal data density using the modified harmonic mean estimator suggested by Geweke (1999, 2005) and computed for VAR models with lag order p, as well as for MS-VAR models with number of states M and lag order, p.

6.2. Granger Noncausality Analysis with VAR Models

To study whether a nonlinear approach brings added value to the Granger causality analysis, we begin by examining a standard VAR model, i.e. the case of $M = 1$. The block Metropolis–Hastings sampler of Section 5 can be simplified to the single-regime case and thus be used for standard Bayesian VAR models. We can thereafter proceed with comparing the VAR models with the more complex MS-VARs with MDDs and also examine whether the Granger causal analysis suggests similar conclusions in VARs and MS-VARs.

The prior distributions are as follows:

$$\beta \sim \mathcal{N}(\mathbf{0}, \lambda_\beta I_{N + pN^2}), \quad \sigma_j \sim \log \mathcal{N}(0, \lambda_\sigma), \quad \mathbf{R}_{21} \sim \mathcal{U}(-1, 1) \tag{16}$$

for $j \in \{1, 2\}$ (and $N = 2$). The hyperparameter $\lambda_\sigma = 2$, while $\lambda_\beta = 0.3$.[10]

To estimate VAR models for different lag lengths, $p \in \{0, \dots, 17\}$, using the Metropolis–Hastings algorithm the parameters are initialized by the OLS estimates of the VAR coefficients. Then follows a 10,000-iteration burn-in and, after convergence of the sampler, 50,000 final draws from the posterior.

Table II displays the MDDs for each model, computed with the modified harmonic mean (MHM) estimator by Geweke (1999, 2005). As in Christiano and Ljungqvist (1988) and Warne (2000), models with long lags are preferred, and the VAR(14) model yields the highest MDD, denoted by ln $p_{\mathrm{MHM}}(\mathbf{y}_T \mid p)$, and equal to −4544.68, and is therefore the model we choose for the Granger causality analysis.

Estimation of the restricted VAR model, where the coefficients on lags of money in the income equation are equal to zero, yields an MDD of −4518.43. Expressed in base-10 logarithms rather than natural logarithms, the posterior odds ratio of the hypothesis of Granger noncausality from money to income is therefore equal to 11.4. Hence Bayesian testing provides strong evidence in favor of the hypothesis that money does not Granger cause income within the VAR framework for log-differences.[11]

---

[10] The value of the hyperparameter $\lambda_\beta = 0.3$ has been established by maximizing the marginal data density for VAR models over a grid of admissible values. For an alternative approach to optimal prior selection in VAR models, see Giannone *et al.* (2015).

[11] This result is in line with Christiano and Ljungqvist (1988), where Granger noncausality from money to output is established for the VAR model with log-differences with US data. Christiano and Ljungqvist (1988), however, contest this result and argue for a specification error for models with first differences. Warne (2000) also finds that money does not Granger cause income in the bivariate VAR model for the log-differences, but that money does Granger cause income in the log-levels specification.

6.3. Granger Noncausality Analysis with MS-VAR Models

To estimate the number of regimes, M, and the number of lags, p, we consider MS-VAR models with a maximum of four regimes and seven lags. The prior distributions are as defined in Section 5 with a diagonal prior covariance matrix of β_i given by $V_\beta = \lambda_\beta I_{N+pN^2}$, with $\lambda_\beta = 0.3$ as in the VAR model analyzed before, and $\lambda_\sigma = 2$. The value of the hyperparameter $\lambda_\mathbf{P} = 10$ implies that the states are persistent over time. The expected duration of the states implied by this prior depends on the number of states, M. For instance, for the models with two regimes, the prior distribution implies an expected duration of the states of around 11 periods, whereas for the model with three regimes the expected duration is around six periods.[12] The block Metropolis–Hastings algorithm has 10,000 burn-in draws before 100,000 final posterior draws are sampled.[13]

Table II reports the estimated MDDs for the MS-VAR models with two and three regimes. The case of $M = 4$ is not provided since the computations suggest that MS-VAR models with more than three regimes are not supported by the data.[14] The number of lags for the autoregressive coefficients is limited to seven—fewer than the 17 lags for VAR models—also owing to insufficient state occurrences when the number of lags increases. The model preferred by the data is the MS-VAR with three regimes and a lag order equal to three. Billio *et al.* (2016), for example, finds similar evidence for the US cycle.

Figure 2 in the online Appendix plots the regime probabilities from the selected model. State 1 has the highest probabilities of occurrence in the period before 1978, and is characterized by moderate average growth (represented by the μ parameters) and volatility (represented by σ parameters) of the time series; see the online Appendix for the posterior estimates of the unrestricted MS-VAR model with three states and three lags. The second state has the highest probabilities of occurrence in the period starting from 1984, and this state is the one with the highest average income growth and the lowest standard deviation. The second state is also a state of the highest average growth rate of money with a moderate level of volatility. The third state has probabilities close to one for the whole year, starting in August 2008, and is also highly likely after July 2011, as well as in the early 1980s and in 1959. This state has the largest standard deviations: 2.5 times higher than in any other state for income, and three times higher than in any other state for money. Moreover, this is the only state in which the average growth of income is negative, as measured by the posterior mean of the intercept term of the VAR equation.

A comparison of the best unrestricted MS-VAR model to the best VAR model (i.e. to the restricted model) yields a posterior odds ratio of over 69 (in terms of the base-10 logarithm) in favor of the MS-VAR model when the models are given equal prior probability. Hence the MS-VAR model with three regimes and three lags is strongly supported by the data over the VAR model with 14 lags.

We proceed with the analysis of Granger noncausality for the selected MS(3)–VAR(3) model. The Bayesian testing strategy we employ renders the process straightforward: each model $\mathcal{M}_j$ implies different restrictions on θ; we impose them, estimate the models and compute the MDDs. Table I gives the restrictions in (A1)–(A3) for MS-VAR models with three regimes and provides an accounting of the number of restrictions imposed on the parameters.

---

[12] The parameters of each of the prior Dirichlet distributions, $e_i = \iota_M + (\lambda_p - 1)I_{M.i}$, depend on the number of regimes. Through these parameters, M affects prior transition probabilities, equal to $e_i / \left(\sum_{j=1}^M e_{i.j}\right)$, with $e_{i.j}$ being the jth element of e_i, and the prior expected state durations (see Hamilton, 1989).

[13] The number of iterations is increased for models that require simulation of the latent Markov process owing to a slightly lower efficiency of simulations for these models.

[14] The computations encountered difficulties for MS-VAR models with four regimes that are due to insufficient occurrence of one regime. We assume that the hidden Markov process is stationary, which implies nonzero ergodic state probabilities. A situation in which at some Gibbs iteration one of the states has zero occurrences violates this assumption and is not allowed in our algorithms. This restriction made sampling from the posterior distribution of parameters of many of the considered models with four states practically impossible. This indicates that the data do not support MS-VAR models with four or more regimes.

Table III. Noncausality and regime inference testing in an MS-VAR models for US monthly data on money and income, 1959:1–2012:11

$\mathcal{M}_j$	Restrictions	$\ln p_{\mathrm{MHM}}\left(\mathbf{y_T}|\mathcal{M}_j\right)$	$\log_{10}\mathcal{B}_{j0}$
$\mathcal{H}_0$	Unrestricted model	**−4384.28**	**0**
$\mathcal{M}_0$	MS(3)−VAR(3)	−4384.28	0
$\mathcal{H}_1$	History of money has no effect on the regime forecast	**−4422.76**	**−16.71**
$\mathcal{M}_1$	(A1)	−4556.22	−74.68
$\mathcal{M}_2$	(A2)	−4422.07	−16.41
$\mathcal{H}_2$	Granger noncausality	**−4392.96**	**−3.77**
$\mathcal{M}_1$	(A1)	−4556.22	−74.68
$\mathcal{M}_3$	(A3) and rank $(\mathbf{P}) = 1$	−4430.35	−20.01
$\mathcal{M}_4$	(A3) and rank $(\mathbf{P}) = 2$	−4488.47	−45.25
$\mathcal{M}_5$	(A3) and rank $(\mathbf{P}) = 3$	−4391.57	−3.17
	(A3)	−4392.67	−3.64

Note: For the definition of $\ln p_{\mathrm{MHM}}\left(\mathbf{y_T}|\mathcal{M}_j\right)$ and $\log_{10}\mathcal{B}_{j0}$ for models $\mathcal{M}_j$ (see note to Table II). Section 4 includes explanations on its computations for hypotheses $\mathcal{H}_i$. For the exact restrictions on parameters for the restricted models see Table I. The Bayes factors for models are computed relative to model $\mathcal{M}_0$, whereas in the case of hypotheses they are computed relative to hypothesis $\mathcal{H}_0$.

It can be seen from Table I that condition (A1) imposes the largest number of restrictions, and condition (A2) the smallest. The restrictions in (A1) allow the regime process to be serially correlated, while condition (A2) implies that it is not. As can be seen from Table I, these two models are minimal and from Proposition 1 it follows that if one of them is true then there is not any information in the history of money for improving the predictions of next period's regime.

Models $\mathcal{M}_3$–$\mathcal{M}_5$ imply that condition (A3) holds and are based on the different values for the rank of the matrix with Markov transition probabilities. The first two of these (A3) models have nonlinear restrictions, while the last model has only linear restrictions. Note that the number of restrictions for these models is typically increasing with the rank of the $\mathbf{P}$ matrix.[15]

Table III reports the natural logarithms of the MDD for each model and the base-10 logarithm of the Bayes factors with respect to model $\mathcal{M}_0$. A positive value of the Bayes factor is to be interpreted as evidence in favor of the restricted model. In a symmetric way, negative values of the Bayes factor indicates that the unrestricted model is preferred by the data. The results in Table III show that model $\mathcal{M}_5$ has the highest MDD among the five restricted models and is comparable to the MDD of the unrestricted model, $\mathcal{M}_0$. The other models ($\mathcal{M}_1$–$\mathcal{M}_4$), however, are much less probable than the unrestricted model, as represented by the large negative values of the logarithms of the Bayes factors. Moreover, the MDDs and Bayes factors of conditions (A1)–(A3) are reported. Owing to the inclusion of model $\mathcal{M}_5$, only condition (A3) is given some posterior support.

Table III also presents a summary of the assessment of the two multiple hypotheses, denoted by $\mathcal{H}_1$ and $\mathcal{H}_2$. The first hypothesis states that the history of money is predictively redundant for next time period regime (see Proposition 1) is covered by the two minimal individual hypotheses represented by models $\mathcal{M}_1$ and $\mathcal{M}_2$. The base-10 logarithm of the Bayes factor is here close to −17 compared with the unrestricted MS-VAR model and, hence, the empirical evidence is strongly in favor of the model where the history of money is useful to improve the predictions of the next period regime.

Second, the MDD of the Granger noncausality hypothesis is lower than the MDD of the unrestricted model by roughly 8.7 natural logarithm units. Translated to logarithms of base-10 scale, this yields a Bayes factor of roughly −3.8, suggesting that the empirical evidence of Granger noncausality from money to income on monthly US data is, at best, very weak when we condition on MS-VAR models.

---

[15] The only exception is when $p = 1$ for $\mathcal{M}_3$ and $\mathcal{M}_4$, when the number of restrictions of these two models are equal.

A byproduct of the analysis for the Markov switching model is Granger noncausality for mixture VARs. As already discussed in Section 3, the mixture-VARs are nested within the MS-VARs by setting the rank of the transition probabilities matrix, $\mathbf{P}$, to unity. The hypothesis of causality is represented by model $\mathcal{M}_2$, while the hypothesis of noncausality is given by $\mathcal{M}_3$. The empirical results are qualitatively similar to those for MS-VARs, with a Bayes factor equal to roughly $10^{3.6}$ in favor of Granger causality over noncausality.[16]

7. CONCLUSIONS

In this paper we derive sets of restrictions on the parameters of MS-VAR models that can be used to test for Granger noncausality and for examining which observed variables have information relevant for improving the predictions of the underlying and unobserved Markov process that determines the regimes.

It is shown that both the Granger noncausality and the regime inference hypotheses imply multiple sets of restrictions on the parameters of the MS-VAR. For bivariate MS-VARs, the number of such sets depends on the number of regimes, while also the number of observed variables (N) matter for N-variate MS-VARs. Granger noncausality results in some of the sets containing nonlinear restrictions, with the nonlinearity being dependent on the rank of the matrix with Markov transition probabilities. Moreover, the number of restrictions actually being tested depends on the rank of this matrix.

In our application we found that Granger noncausality from monthly US money growth to income growth established for linear (single-regime) VARs is contradicted by the evidence from nonlinear models. The causality analysis of MS-VARs suggests that money is essential for the forecasting of the probabilities of the states which influence the behavior of income. Since mixture VAR models are nested in MS-VAR models, our analytical results on Granger noncausality can also be applied to such models. In our application, we also find strong support in favor of the hypothesis that money Granger causes income in a mixture VAR. Moreover, we find that MS-VARs dominate mixture VARs, while mixture VARs strongly dominate the linear VAR specification. Our methodology, however, is more general and could be applied to the analysis of factors driving the Great Moderation, the Great Recession, business cycles or speculative bubbles formalized within the MS-VARs framework.

The Granger noncausality analysis that we have presented in this paper focuses on the properties of the mean squared errors of the forecasts. It is possible that a variable y_2 does not Granger cause another variable y_1 from this perspective, but may nevertheless incorporate important information which is valuable for determining higher moments than the mean of the predictive distribution of y_1. Warne (2000) provides additional noncausality concepts, namely the second-order Granger causality and the Granger causality in distribution, which provide an interesting avenue for future empirical analysis on causal relations in MS-VARs.

---

[16] Note that three regimes and three lags may not be the preferred choice of these parameters if we were to allow only for mixture VARs when estimating them.

REFERENCES

Barnard J, McCulloch R, Xl Meng. 2000. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* **10**: 1281–1311.

Billio M, Casarin R, Ravazzolo F, van Dijk HK. 2016. Interconnections between eurozone and US booms and busts using a Bayesian panel Markov-switching VAR model. *Journal of Applied Econometrics*. DOI:10.1002/jae.2501.

Christiano LJ, Ljungqvist L. 1988. Money does Granger-cause output in the bivariate money–output relation. *Journal of Monetary Economics* **22**: 217–235.

Doan T, Litterman RB, Sims CA. 1983. Forecasting and conditional projection using realistic prior distributions. *NBER Working Paper*: 1202.

Droumaguet M, Warne A, Woźniak T. 2015. Granger causality and regime inference in Bayesian Markov switching VARs. *Working Paper 1794*, European Central Bank.

Dufour JM, Tessier D. 1993. On the relationship between impulse response analysis, innovation accounting and Granger causality. *Economics Letters* **42**: 327–333.

Friedman M, Schwartz A. 1971. *A Monetary History of the United States, 1867–1960*, Vol. 12. Princeton University Press: Princeton, NJ.

Frühwirth-Schnatter S. 2006. *Finite Mixture and Markov Switching Models*. Springer: Berlin.

Gabriel KR. 1969. Simultaneous test procedures: some theory of multiple comparisons. *Annals of Mathematical Statistics* **40**: 224–250.

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

Geweke J. 2005. *Contemporary Bayesian Econometrics and Statistics*. Wiley: Hoboken, NJ.

Giannone D, Lenza M, Primiceri GE. 2015. Prior selection for vector autoregressions. *Review of Economics and Statistics* **97**: 436–451.

Granger CWJ. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**: 424–438.

Hamilton JD. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**: 357–384.

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

Joyce M, Miles D, Scott A, Vayanos D. 2012. Quantitative easing and unconventional monetary policy: an introduction. *Economic Journal* **122**: F271—F288.

Lechner M. 2011. The relation of different concepts of causality used in time series and microeconometrics. *Econometric Reviews* **30**: 109–127.

Lucas RE Jr. 1980. Two illustrations of the quantity theory of money. *American Economic Review* **70**: 1005–1014.

Lütkepohl H. 1993. *Introduction to Multiple Time Series Analysis*. Springer: Berlin.

Psaradakis Z, Ravn MO, Sola M. 2005. Markov switching causality and the money-output relationship. *Journal of Applied Econometrics* **20**: 665–683.

Ritter C, Tanner MA. 1992. Facilitating the Gibbs sampler: the Gibbs stopper and the Griddy–Gibbs sampler. *Journal of the American Statistical Association* **87**: 861–868.

Rubin DB. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**: 688–701.

Sargent T, Surico P. 2011. Two illustrations of the quantity theory of money: breakdowns and revivals. *American Economic Review* **101**: 109–128.

Sims CA. 1972. Money, income, and causality. *American Economic Review* **62**: 540–552.

Tobin J. 1970. Money and income: post hoc ergo propter hoc *Quarterly Journal of Economics* **84**: 301–317.

Warne A. 2000. Causality and regime inference in a Markov switching VAR. *Working Paper Series 118*, Sveriges Riksbank (Central Bank of Sweden).

Woźniak T. 2015. Testing causality between two vectors in multivariate GARCH models. *International Journal of Forecasting* **31**: 876–894.

Woźniak T. 2016. Granger-causal analysis of GARCH models: a Bayesian approach. *Econometric Reviews*. DOI:10.1080/07474938.2015.1092839.