Granger-causality in Markov switching models

Abstract

In this paper we propose a new approach for characterizing and testing Granger-causality, which is well-equipped to handle models where the change in regime evolves according to multiple Markov chains. Differently from the existing literature, we propose a method for analyzing causal links that specifically takes into account Markov chains. Tests for independence are also provided. We illustrate the methodology with an empirical application, and in particular we investigate the causality and interdependence between financial and economic cycles in U.S. using the bivariate Markov switching model proposed by Hamilton and Lin (1996). We find that financial variables are useful in forecasting the aggregate economic activity, and vice versa.

Keywords: Granger-Causality, Markov-Switching models, Markov chain.

J.E.L. classification: C-12, C-32

1 Introduction

The most popular causality concept in econometrics literature has been introduced by Granger (1969). He defines a variable X to be causal for a time series variable Y if the former helps to improve the forecasts of the latter. In the great majority of practical applications, Granger-causality has been analyzed in the context of linear Vector Autoregressive (VAR) models (see for example Hamilton, 1994; Sims *et al.*, 1990; Toda and Phillips, 1993). For such models, the necessary and sufficient condition for X to be non-causal for Y is that all coefficients of lags of X are zero in the equation that describes Y.

Recently, some attempts have been made to extend Granger-causality to Markov switching (MS) models[1].

For example, Warne (2000) and Psaradakis *et al.* (2005) propose different definitions of causality based on Granger's ideas and provide a set of (economically and statistically meaningful) parametric Granger non-causality restrictions in the context of MS-VAR models. The common feature of these works is that they suppose independent Markov chains. Clearly, in some empirical applications this hypothesis could be too restrictive and the Markov chains can be correlated.

Accordingly, in this paper we propose a method for analyzing causal links by working directly with Markov chains[2]. Our method allows us to consider different links between the Markov chains that are interesting in empirical applications and that are ruled-out by using the independence hypothesis.

[1] See Hamilton (1989, 1994), Kim and Nelson (1999) and McCulloch and Tsay (1994) for a extensive discussion of the MS models.

[2] Our work is closely related to Mosconi and Seri (2006) results but proposes a different and simpler parametrisation useful for Markov chains.

More precisely, our strategy is based on a particular decomposition of the transition probabilities that allows us to test directly if a Markov chain causes another one in the Granger sense, that is, if one Markov chain helps to predict another one.

Causality analysis based on Markov chains is attractive for a number of reasons. First, in a multi-country/multi-sector framework, this methodology allows the explanation of the interactions among macro-areas. In fact, this approach is very useful to study the relationships between phases in different countries or sectors, and allows us to determine the causality of these relationships improving our comprehension and description of how these phases evolve. Moreover, it is certainly useful in order to describe the relationships between leading and lagging countries or to describe the relationships between business surveys and macroeconomic variables.

We illustrate our methodology with an empirical application. We investigate the causality and interdependence between financial and economic cycles using the MS model proposed by Hamilton and Lin (1996, hereafter HL). HL find that economic variables may be useful in forecasting stock returns volatility but no rigorous test based on Granger's ideas is provided. With our methodology it is possible to determine, in a formal way, whether business cycles have predicting power for financial variables and/or vice versa.

The paper is organized as follows. In section 2 we introduce the Multiple chain Markov switching VAR, i.e. a general MS-VAR model with correlated Markov chains. In section 3, we discuss the basic assumptions and the definitions of Granger-causality we adopt in this paper and we define the parametric conditions for Granger-causality. An application of these tests is presented in section 4.

Section 5 concludes.

2 The Multiple chain Markov switching VAR Model

Let $Y_t = (Y_{1t}, ..., Y_{nt})'$ denote a time series n-variate vector for $t \in \{1, ..., T\}$.

We consider the following model that we call Multi-chain Markov switching (MCMS) VAR model:

$$\begin{aligned} Y_t &= \mu(S_t) + \sum_{k=1}^{p} \Phi_k(S_t) Y_{t-k} + \varepsilon_t \quad (1) \\ and\ \varepsilon_t &\sim i.i.d.N(0, \Omega(S_t)) \end{aligned}$$

where the mean $\mu(S_t)$, the lag polynomial matrices $\Phi_k(S_t)$, $k = 1, ..., p$, and the variances and covariances of the error terms ε_t, all depend on the unobserved latent random vector $S_t = (S_{1t}, ..., S_{nt})$ which denotes the unobserved state (or regime) of the system at time t[3]. S_{jt} represents the state variable associated with Y_{jt} and can assume M possible values. The regimes of the joint Markov chain S_t are all the possible combinations of the regimes of the specific Markov chains S_{jt} $\{j = 1, ..., n\}$. Therefore S_t can assume M^n values.

The random sequence $\{S_t\}$ is assumed to be a time-homogenous first order Markov process with one step ahead transition probabilities given by:

[3] Through the paper the following notation will be taken: capital letters denote the random variable, while small letters denote a particular realization. Moreover, $\{Q_t\}$ denotes a stochastic process, while Q_t represents the value of the process at time t.

$$P_{ij} = \Pr(S_t = j | S_{t-1} = i), \quad i, j \in S = \{1, ..., M^n\}. \tag{2}$$

where all relevant information in the transition probabilities in t reduces to the knowledge of the state of the process in $(t-1)$. These transition probabilities are collected in the transition matrix $\Pi = \{P_{ij}\}_{i,j \in S}$ which contains $M^n \times M^n$ elements. If we suppose that the Markov chains S_{jt} $\{j = 1, ..., n\}$ are independent, then the transition matrix is given by the Kronecker product of the transition matrix of each specific chain: $\Pi = \Pi_1 \otimes \Pi_2 \otimes ...\Pi_n$. Note that the MCMS-VAR model is more general than the MS-VAR model proposed by Krolzig (1997) where all components in the process Y_t depend on a single Markov chain and thus share the same dynamics. In the MCMS-VAR model, the regime shift of each of the n endogenous variables is driven by a specific Markov chain S_{jt} with M regimes and these Markov chains are potentially related, thus this model represents an ideal framework to study the causal relationship between the Markov chains.

In the rest of the paper, to make the analysis simpler, we consider Granger-causality between the components of the bivariate time series $Y_t = (Y_{1t}, Y_{2t})'$ with $M = 2$ and we denote these states by 0 and 1. In this case the Markov chain $S_t = (S_{1t}, S_{2t})$ can assume four different values $\{(0,0), (1,0), (0,1), (1,1)\}$ for any $t \in \{1, ..., T\}$, and the matrix P contains 12 elements, being the sum of each column equal to one.

Accordingly, the basic set up of the analysis is the following bivariate MCMS-VAR model:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1(S_t) \\ \mu_2(S_t) \end{bmatrix} + \sum_{k=1}^{p} \begin{bmatrix} \phi_{11}^k(S_t) & \phi_{12}^k(S_t) \\ \phi_{21}^k(S_t) & \phi_{22}^k(S_t) \end{bmatrix} \begin{bmatrix} Y_{1t-k} \\ Y_{2t-k} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}. \tag{3}$$

3 Granger-causality

3.1 Basic Definitions

The aim of this section is to provide a mathematically rigorous definition of non-causality based on predictability. Accordingly, we first specify the stochastic process we want to predict, the available information set, and the reduced information set. In the relevant theoretical literature, several generalization of non-causality are available. In this paper we follow Mosconi and Seri (2006) and adopt the concept of discrete-time one-step ahead strong non-causality proposed by Florens and Fougère (1996). Following their definition, one-step ahead is referred to the prediction horizon while strong, as opposed to weak, means that we focus on predicting the whole distribution, rather than only the mean.

Let $\{S_t = (S_{1t}, S_{2t}),\ t \in I \subseteq \mathbb{N} = \{1, 2, ...\}\}$, or $\{S_t\}$ for short, be a discrete-time stochastic process on the probability space (Ω, Υ, P). The usual statistical problem of non-causality is to test if P satisfies the non-causality conditions. The filtration $\{F_t, t \in I\} = \{F_t\}$ provides the information available at time t. To make the analysis simpler, we assume $\{F_t\}$ to be the canonical filtration associated with the stochastic process $\{(S_t)\} = \{(S_{1t}, S_{2t})\}$,[4] where $\{S_{1t}\}$ and $\{S_{2t}\}$ may be either scalar or vector processes. Finally, we introduce the reduced information set, which is represented by the canonical filtrations $\{G_t^1\} = \{\sigma\{(S_{1\tau}), 1 \leq \tau \leq t\}\}$ and $\{G_t^2\} = \{\sigma\{(S_{2\tau}), 1 \leq \tau \leq t\}\}$. Note that $G_t^i \subseteq F_t$, $\forall t \in I$ and $i = 1, 2$.

Through the paper the following set of definitions, which are fixed in terms of

[4] Recall that a canonical filtration associated with a general process $\{Q_t\}$ defined on (Ω, Υ, P) is a family $\{F_t\}$ of $sub-\sigma$ fields of Υ, where $F_t = \sigma\{Q_i, 1 \leq i \leq t\}$. In a more intuitive way, F_t represents the knowledge of the history of $\{Q_t\}$ up to time t.

conditional independence of sub-σ-fields of Υ, will be adopted (see Florens and Mouchart, 1982, for further details):

Definition 1. *Strong one-step ahead Granger non-causality*: $\{S_{2t}\}$ does not strongly cause $\{S_{1t}\}$ one-step ahead given $\left\{G_{t-1}^{1}\right\}$, briefly $S_{1t} \nleftarrow S_{2t}$, if:

$$G_t^1 \perp\!\!\!\perp G_{t-1}^2 | G_{t-1}^1 \qquad \forall t \in I. \tag{4}$$

Similarly,

$\{S_{1t}\}$ does not strongly cause $\{S_{2t}\}$ one-step ahead given $\left\{G_{t-1}^{2}\right\}$, briefly $S_{1t} \nrightarrow S_{2t}$, if:

$$G_t^2 \perp\!\!\!\perp G_{t-1}^1 | G_{t-1}^2 \qquad \forall t \in I. \tag{5}$$

Definition 2. *Strong simultaneous independence*: $\{S_{1t}\}$ and $\{S_{2t}\}$ are strongly simultaneous independent given $\{F_{t-1}\}$, briefly $S_{1t} \nLeftrightarrow S_{2t}$, if:

$$G_t^1 \perp\!\!\!\perp G_t^2 | F_{t-1} \qquad \forall t \in I. \tag{6}$$

We next show how we can apply definitions (1) and (2) to specific stochastic processes and information sets.

To be concrete, as previously anticipated we assume that $S_t = (S_{1t}, S_{2t})$ and each component is a first order Markov chain with stationary transition probabilities that takes values on $\{0, 1\}$. We continue to restrict the information set to the canonical filtration associated with $\{S_t\}$. The state space of $S_t = (S_{1t}, S_{2t})$ is thus $\{(0,0), (1,0), (0,1), (1,1)\}$ for any $t \in \{1, ..., T\}$.

For a Markov chain, notice that we can decompose its transition probabilities as follows:

$$P(s_t|s_{t-1}) = P(s_{1t}, s_{2t}|s_{1t-1}, s_{2t-1}) = P(s_{1t}|s_{2t}, s_{1t-1}, s_{2t-1})P(s_{2t}|s_{1t-1}, s_{2t-1}), \tag{7}$$

and accordingly we can define Granger non-causality for a Markov chain:

Definition 3. *Strong one-step ahead non-causality for a Markov chain with stationary transition probabilities:* S_{2t-1} does not strongly cause S_{1t} one step ahead given S_{1t-1}, briefly $S_{1t} \nleftarrow S_{2t}$, if:

$$P(s_{1t}|s_{1t-1}, s_{2t-1}) = P(s_{1t}|s_{1t-1}) \qquad \forall t. \tag{8}$$

Similarly,

S_{1t-1} does not strongly cause S_{2t} one step ahead given S_{2t-1}, briefly $S_{1t} \nrightarrow S_{2t}$, if:

$$P(s_{2t}|s_{1t-1}, s_{2t-1}) = P(s_{2t}|s_{2t-1}) \qquad \forall t. \tag{9}$$

Definition 4. *Strong simultaneous independence for a Markov chain with stationary transition probabilities*: S_{1t} and S_{2t} are strongly simultaneously independent given S_{t-1}, briefly $S_{1t} \nLeftrightarrow S_{2t}$,if:

$$P(s_{1t}, s_{2t}|s_{1t-1}, s_{2t-1}) = P(s_{1t}|s_{1t-1}, s_{2t-1})P(s_{2t}|s_{1t-1}, s_{2t-1}) \qquad \forall t. \tag{10}$$

It is easy to show the equivalence between (4) and (8). In fact, note that under the Markov assumption, the conditional statement (4) implies $P(s_{1t}, s_{2t-1}/s_{1t-1}) =$

$P(s_{1t}/s_{1t-1})P(s_{2t-1}/s_{1t-1})$ $\forall t$, which in turn implies (8). Similarly, it is possible to proceed for the other definitions.

The non-causality definitions require the marginal distributions of S_{1t} and S_{2t} conditional on S_{t-1}. Then, to study causality it is necessary to consider the transition probabilities of the Markov process while to test for simultaneous independence we need to specify the joint distribution and compare it to the product of the marginal distributions.

Let us now show how we can construct a transition matrix starting from the non-causality definitions. Note that the state of the system in $(t-1)$ can be defined by the four possibilities of the joint Markov chain S_{t-1} :

$$B_t = (1, s_{1t-1}, s_{2t-1}, s_{2t-1}s_{1t-1})' = (1, s_{2t-1})' \otimes (1, s_{1t-1})',$$

where $\otimes$ denotes the Kronecker product[5]. Given the decomposition in (7), we can consider the logistic function to represent the two probabilities and then it is simple to verify that we can represent the joint probability of S_{1t} and S_{2t} as follows:

[5] In fact, B_t is an invertible linear transformation of:

$$B_t^* = [(1 - s_{1t-1})(1 - s_{2t-1}), s_{1t-1}(1 - s_{2t-1}), (1 - s_{1t-1}) s_{2t-1}, s_{1t-1}s_{2t-1}]',$$

where B_t^* is characterized by four mutually exclusive dummies representing the four states of the process in $(t-1)$. As in Mosconi and Seri (2006), we employ B_t instead of B_t^* to describe the state in $(t-1)$ since by means of this specification non-causality restrictions are more easily written and interpreted.

$$P(s_{1t}, s_{2t}|s_{1t-1}, s_{2t-1}) = P(s_{1t}|s_{2t}, s_{1t-1}, s_{2t-1})P(s_{2t}|s_{1t-1}, s_{2t-1}) \quad (11)$$
$$= \frac{\exp(\alpha' A_t)}{1+\exp(\alpha' A_t)} * \frac{\exp(\beta' B_t)}{1+\exp(\beta' B_t)},$$

where

$$A_t = (1, s_{2t})' \otimes (1, s_{1t-1})' \otimes (1, s_{2t-1})'$$
$$= (1, s_{2t-1}, s_{1t-1}, s_{1t-1}s_{2t-1}, s_{2t}, s_{2t}s_{2t-1}, s_{2t}s_{1t-1}, s_{2t}s_{1t-1}s_{2t-1})'$$

or alternatively

$$P(s_{1t}, s_{2t}|s_{1t-1}, s_{2t-1}) = P(s_{2t}|s_{1t}, s_{1t-1}, s_{2t-1})P(s_{1t}|s_{1t-1}, s_{2t-1}) \quad (12)$$
$$= \frac{\exp(\alpha^{*\prime} A_t^*)}{1+\exp(\alpha^{*\prime} A_t^*)} * \frac{\exp(\beta^{*\prime} B_t)}{1+\exp(\beta^{*\prime} B_t)},$$

where

$$A_t^* = (1, s_{1t})' \otimes (1, s_{2t-1})' \otimes (1, s_{1t-1})'$$
$$= (1, s_{1t-1}, s_{2t-1}, s_{2t-1}s_{1t-1}, s_{1t}, s_{1t}s_{1t-1}, s_{1t}s_{2t-1}, s_{1t}s_{2t-1}s_{1t-1})',$$

and the vectors α and β (or α^* and β^*) have dimensions (8×1) and (4×1), respectively[6], and B_t has already been defined. Note that α and β represent 8 and 4 parameters, respectively. Then, we simply have an alternative parameterization of the 12 elements of the transition matrix.

The parameterizations (11)-(12) are very useful since they allow us to impose the non-causality restrictions in a very simple way by easily restricting the transition matrix to be described by a number of parameters comprised between 4 and 12.

[6] In the following we denote with α_j and β_j the $j^{'}th$ element of the vectors α and β, respectively.

It is important to note that the parameters (α, β) and (α^*, β^*) are bijective transformations of the transition probabilities. That is

$$P_{ij} = f(A_t, B_t, \alpha, \beta) = f^*(A_t^*, B_t, \alpha^*, \beta^*). \tag{13}$$

Hence, we can obtain the parameters (α^*, β^*) as:

$$(\alpha^*, \beta^*) = f^{*-1}(P_{ij}) = f^{*-1}(f(A_t, B_t, \alpha, \beta)). \tag{14}$$

Consequently we can obtain the values of (α^*, β^*) from (α, β).

3.2 Parameter restrictions for Granger-causality

Given the parametrizations (11)-(12), the conditions for strong one step ahead non-causality and strong simultaneous independence are easily stated as restrictions on the parameter space.

The restriction $H_{1 \nRightarrow 2}$, related to the non-causality of of S_{1t} towards S_{2t}, implies that the dependence on S_{1t-1} disappears in the second term of the decomposition (11), thus it simply requires that the parameters of the terms of B_t depending on S_{1t-1} are equal to zero:

$$H_{1 \nRightarrow 2}(S_{1t} \nrightarrow S_{2t}): \quad \beta_2 = \beta_4 = 0. \tag{15}$$

Under $H_{1 \nRightarrow 2}$, S_{1t-1} does not strongly cause one-step ahead S_{2t} given S_{2t-1}. The terms s_{1t-1} and $s_{2t-1}s_{1t-1}$ are excluded from B_t, thus $p(s_{2t}/s_{1t-1}, s_{2t-1}) = p(s_{2t}/s_{2t-1})$.

On the other hand, S_{2t-1} does not strongly cause one-step ahead S_{1t} given S_{1t-1}, if:

$$H_{1\not\Rightarrow 2}(S_{1t} \not\leftarrow S_{2t}): \quad \beta_3^* = \beta_4^* = 0. \tag{16}$$

The terms s_{2t-1} and $s_{2t-1}s_{1t-1}$ are excluded from B_t, thus $p(s_{1t}/s_{1t-1}, s_{2t-1}) = p(s_{1t}/s_{1t-1})$.

In the same way it is possible to define restrictions for the strong simultaneous independence:

$$H_{1\not\Leftrightarrow 2}(S_{1t} \not\Leftrightarrow S_{2t}): \quad \alpha_5 = \alpha_6 = \alpha_7 = \alpha_8 = 0, \tag{17}$$

hence, under this restriction, the joint distribution $P(s_{1t}, s_{2t}|s_{1t-1}, s_{2t-1})$ can be written as follows:

$$P(s_{1t}, s_{2t}|s_{1t-1}, s_{2t-1}) = P(s_{1t}|s_{1t-1}, s_{2t-1})P(s_{2t}|s_{1t-1}, s_{2t-1}).$$

In the present framework it is also possible to define restrictions for the independence of the Markov chains. When independence holds, the transition matrix of the joint Markov process is given by the Kronecker product of the transition matrix of each specific chain. In fact, we can write the transition probabilities as follows:

$$P(s_{1t}, s_{2t}|s_{1t-1}, s_{2t-1}) = P(s_{1t}|s_{1t-1})P(s_{2t}|s_{2t-1}), \tag{18}$$

and the transition matrix Π will be:

$$\Pi = \Pi_1 \otimes \Pi_2, \tag{19}$$

where Π_i, $i = 1, 2$ is the (2×2) transition matrix of the specific Markov chains. Thus restrictions for independence are given by:

$$H_{1 \perp\!\!\!\perp 2} \quad (S_{1t} \perp\!\!\!\perp S_{2t}): \quad \beta_2 = \beta_4 = 0; \alpha_2 = \alpha_4 = \alpha_5 = \alpha_6 = \alpha_7 = \alpha_8 = 0. \quad (20)$$

As discussed in Warne (2000), in empirical applications it is important to consider Granger non-causality not only for the underlying Markov chains but also for the observable processes Y_{1t} and Y_{2t}.

In this case, in our framework we can simply define that Y_{2t-1} does not Granger-cause Y_{1t} (given Y_{1t-1} and S_{1t-1}) if:

(i) $P(s_{1t}/s_{1t-1}, s_{2t-1}) = P(s_{1t}/s_{1t-1})$

(ii) $\mu_i(S_t) = \mu_i(S_{it})$

(iii) $\phi_{ij}^k(S_t) = \phi_{ij}^k(S_{it})$

(iv) $\Omega_{ii}(S_t) = \Omega_{ii}(S_{it})$

(v) $\Omega_{12}(S_t) = 0$

for all $i, j \in (1, 2)$, $k \in \{1, ..., p\}$, $S_t \in \{(0,0), (1,0), (0,1), (1,1)\}$ and

(vi) $\phi_{12}^k(S_{1t}) = 0$ for all $k \in \{1, ..., p\}$ and $S_{1t} \in \{0, 1\}$.

Restriction (i) states that the information s_{2t-1} is not useful for predicting s_{1t} and the test is possible using the procedure previously described. Restrictions (ii)-(iv) simply state that the parameters of the equation for Y_{1t} change only according to the process S_{1t}, and the parameters of the equation for Y_{2t} change only according to the process S_{2t}. Restriction (v) states the instantaneous non-causality between the two variables, Y_{1t} and Y_{2t}, defined as a zero correlation condition. Finally, restriction (vi) states the Granger non-causality for the VAR model. Note that, if we replace the restriction (vi) with $\phi_{21}^k(S_{2t}) = 0$ and restriction (i) with $P(s_{2t}/s_{1t-1}, s_{2t-1}) = P(s_{2t}/s_{2t-1})$, then Y_{1t-1} does not Granger-cause Y_{2t} (given Y_{2t-1} and S_{2t-1}).

Finally, if we replace the restriction (i) with $\Pi = \Pi_1 \otimes \Pi_2$, i.e. we impose independent Markov chains, the set of non-causality restrictions coincide exactly with those proposed by Warne (2000).

Granger non-causality restrictions described in this section can be tested with a Wald, Lagrange Multiplier (LM) or Likelihood Ratio (LR) test. Under sufficient regularity conditions, the Maximum Likelihood estimator of the MS model is consistent and converges to a Normal distribution (see Lindgren, 1978; Kiefer, 1978; Psaradakis and Sola, 1998): according all test statistics are χ^2-distributed. However, in some empirical applications, the regularity conditions may not hold and the conventional Gaussian asymptotic approximations can poorly perform (see Psaradakis and Sola, 1998). To study the quality of asymptotic approximations, we performed a simple Monte Carlo experiment[7] using as data-generating process the model proposed by Hamilton and Lin (1996), which is described in the following section. We found that for $T \geqslant 100$ and a nominal size of 0.05, the tests have empirical rejection frequencies that are insignificantly different from 0.05 and a power ranging from 87% to 93%.

4 An application to the relationship between financial and business cycles

To illustrate our methodology, we consider the model proposed by Hamilton and Lin (1996).

Hamilton and Lin propose the following model specification for the business

[7] See Psaradakis and Sola (1998), Di Sanzo and Perez-Alonso (2011) and Di Sanzo (2009) for details about Monte Carlo and bootstrap simulations in MS models and nonlinear time-series.

cycle:

$$Y_t - \mu_{S_t} = \phi(Y_{t-1} - \mu_{S_{t-1}}) + \sigma\epsilon_t \tag{21}$$

where Y_t is the monthly growth rate of industrial production. S_t is an unobserved latent variable that takes values $(0, 1)$, represents the state of the business cycle and it is described as a two state Markov chain. When $S_t = 0$, the average growth rate of industrial production is given by the population parameter μ_0 whereas when $S_t = 1$, the average growth rate is μ_1. ϕ is an unknown constant, ϵ_t is assumed to be i.i.d. $N(0, 1)$ and σ represents the standard deviation of the error term.

The model for stock returns takes the form:

$$R_t = \delta_0 + \delta_1 R_{t-1} + e_t \tag{22}$$

$$e_t = \sqrt{g_{Z_t}} u_t \tag{23}$$

$$u_t = \sqrt{h_t} w_t \tag{24}$$

$$h_t = \zeta + \xi u_{t-1}^2 + \varphi u_{t-1}^2 I_{t-1} \tag{25}$$

where R_t denotes the monthly excess return on the S&P 500 stocks over the Treasury bill yield. The random variable w_t is assumed to be i.i.d. $N(0, 1)$, Z_t is an unobserved latent variable that reflects the volatility phase of the stock market and $(\delta_0, \delta_1, g_{z_t}, \zeta, \xi, \varphi)$ are unknown parameters. As before, Z_t follows a two-state first-order Markov process.

If the parameter g_{Z_t} in equation (23) does not switch between regimes, it is simply equals to unity for all t and in this case equations (22)-(25) describe stock returns with an autoregression whose residual e_t follows a first-order ARCH-L process. The "L" in ARCH-L denotes the leverage effect, which indicates that stock price increases and decreases can produce asymmetric effects on subsequent volatility (see Nelson, 1991). The indicator variable I_{t-1} in equation (25) takes value 1 if e_{t-1} is negative and zero otherwise. This means that if $\varphi > 0$, a stock price decrease has a greater effect on subsequent volatility that would a stock price increase of the same magnitude. Hamilton and Lin normalize $g_0 = 1$, thus g_1 can be interpreted as the ratio of the average variance of stock returns when $Z_t = 1$ compared to that observed when $Z_t = 0$.

Hamilton and Lin study the relationship between financial and business cycles by putting the processes (21)-(22) into a first-order MCMS-VAR model:

$$\begin{bmatrix} Y_t \\ R_t \end{bmatrix} = \begin{bmatrix} \mu_{S_t} - \phi\mu_{S_{t-1}} \\ \delta_0 \end{bmatrix} + \begin{bmatrix} \phi & 0 \\ 0 & \delta_1 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ R_{t-1} \end{bmatrix} + \begin{bmatrix} \sigma & 0 \\ 0 & \sqrt{g_{Z_t} h_t} \end{bmatrix} v_t. \quad (26)$$

The error term v_t is assumed to be *i.i.d.* $N(0, I_2)$ with I_2 standing for the (2×2) identity matrix.

It is worth nothing that in model (26) the variables Y_t and R_t are linked only through variables S_t and Z_t. This specification highlights clearly the main feature of our strategy test that allows testing Granger-causality by working directly with Markov chains.

It also possible to allow for dynamic linkages through more conventional coefficients as well, i.e. adding lagged R_{t-j} to the autoregression explaining Y_t, for example[8]. In this case it is important to consider the complete non-causality restrictions set described in the previous section.

In their work, Hamilton and Lin show that economic variables may be useful

[8]However, based on the specification tests for MS models proposed by Hamilton (1996), we settle on this parsimonious specification of the model. See Smith (2008) for an evaluation of the finite sample properties of these tests .

in forecasting stock returns volatility, but no rigorous test based on Granger's ideas is provided. We thus employ the Granger-causality tests described in the previous section for investigating, in a formal way, whether business cycle has a predicting power for financial variables and/or vice versa.

4.1 Estimation results and testing

We employ the same data set as Hamilton and Lin (1996) to facilitate the comparison. The output variable Y is obtained as 100 times the monthly change in the natural logarithm of the Federal Reserve Board's index of industrial production from 1965: M1 to 1993: M6. The excess stock return R is 100 times the change in the natural logarithm of the S&P 500 stock index (plus the dividend yield on the S&P 500) minus the 3-month Treasury bill yield, both quoted at monthly rates. All data were taken from Citibase.

Maximum Likelihood estimates for the parameters and transition probabilities are reported in Tables 1 and 2, respectively.

As in Hamilton and Lin, the parameters μ_{s_t} are large in absolute value and statistically significant suggesting two distinct growth states. The sign of the average growth rate of the output is positive during regime 1, $\widehat{\mu}_1 = 0.293$, and negative during regime 0, $\mu_0 = -0.687$. Therefore output growth is well characterized by recurrent shifts between positive and negative growth periods.

The estimated value $\widehat{g}_1 = 0.118$ indicates that the unforecastable component of stock returns (the residual e_t in equation (22)) has a variance that is over nine times as large in regime 0 as it is in regime 1.

To explore the nature of the link between economic variables and stock returns it is crucial to study the transition matrix of the joint process since it contains information on the relationship among the phases of the different series.

The transition matrix produces useful information on the most probable scenario at time t, given the scenario at time $t-1$.

Analyzing table 2, we observe that when both variables are in the same state at time $t-1$, it is probable to maintain this scenario in the next period. In fact, we observe that once we get $(S_{t-1} = 0, Z_{t-1} = 0)$ at time $t-1$, the most probable scenario at time t is $(S_t = 0, Z_t = 0)$ with a 80% probability, while from $(S_{t-1} = 1, Z_{t-1} = 1)$ the most probable scenario at time t is $(S_t = 1, Z_t = 1)$ with probability equals to 70%. These results lead to the conclusion that the stock returns volatility and the business cycle are typically driven by the same events and thus, the former may be useful in forecasting the latter, and *vice versa.*

We test in a formal way this issue using the Granger-causality tests described in section 3.2. The non-causality relationships between S_t and Z_t are tested employing Wald type tests. We consider the standard p-value, p_{STD}, based on the Normal approximation and the bootstrap p-value, p_b, calculated using bootstrap simulations. Results are depicted in Table 3:

1) The hypothesis $H_{1\nRightarrow 2}$, related to the non-causality of S_t towards Z_t, is strongly rejected. Therefore, macroeconomic variables, such as the index of industrial production, have a predicting power for financial variables. This result supports the empirical evidence provided by Chen (1991) and Billio and Pelizzon (2003) among others, that macroeconomic variables are key determinants of stock returns.

2) The hypothesis $H_{1\nLeftarrow 2}$, related to the non-causality of Z_t towards S_t, is also rejected. This means that stock volatility is useful for forecasting the direction of aggregate economic activity. Our finding is in accordance with the empirical

works of Perez-Quiros and Timmermann (2001) and Chauvet (1999), among others, which show that financial variables lead the business cycle and seem to be generated from expectations about changes in future economic activity.

3) The hypothesis $H_{1 \perp\!\!\!\perp 2}$ and $H_{1 \not\Leftrightarrow 2}$, concerning the independence between Z_t and S_t, are rejected. This result is in contrast with the general attitude in empirical literature studying relationships between financial and economic cycles, to impose a priori the independence of the Markov chains (see Chauvet, 1999, among others).

5 Conclusions

In this paper we propose a new technique to test Granger-causality in Multiple chain Markov switching VAR models with correlated Markov chains. Our strategy is based on a particular decomposition of the transition probabilities that allows to test directly if a Markov chain causes another Markov chain in the Granger sense, that is, if one Markov chain helps to predict another one. Test for independence are also provided. We finally illustrate our methodology with an empirical application. In particular, we investigate the causality and interdependence between financial and economic cycles in U.S. using the bivariate (MS) model proposed by Hamilton and Lin (1996). The causality tests suggest that financial variables are useful in forecasting the direction of aggregate economic activity, and vice versa.

References

Billio M., and L. Pelizzon (2003). Volatility and shocks spillover before and after EMU in Europe stock markets, *Journal of Multinational Financial Management*, 13, 323-340.

Boldin, M.D. (1996). A check on the robustness of Hamilton's Markov Switching model approach to the economic analysis of the business cycle. *Studies in Nonlinear Dynamics & Econometrics,* 1, 35-46.

Breunig, R., Najarian S. and Pagan A., (2003). Specification Testing of Markov Switching Models. *Oxford Bulletin of Economics and Statistics*, 65, 703-725.

Chauvet M., (1999). Stock Market Fluctuations and the Business cycle, *Journal of Economic and Social Measurement*, 25, 235-57.

Chen N., (1991). Financial Economic Opportunities and the Macroeconomy, *Journal of Finance*, 46, 529-554.

Di Sanzo, S. (2009). Testing for Linearity in Markov Switching Models: A Bootstrap Approach, *Statistical Methods & Applications*, 18, 153-168.

Di Sanzo S., and A. Pérez-Alonso, (2011). Unemployment and hysteresis: A nonlinear unobserved components approach , *Studies in Nonlinear Dynamics & Econometrics,* 15, Issue 1, 24–50.

Florens, J.P. and D. Fougère (1996). Non-Causality in Continuous Time, *Econometrica*, 1195-1212.

Florens, J. P. and M. Mouchart (1982). A Note on Non-Causality, *Econometrica*, 583-91.

Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica*, 357-84.

Hamilton, J.D. (1996). Specification Testing in Markov-switching Time Series Models, *Journal of Econometrics*,70, 127-157.

Hamilton, J.D. and J. Lin (1996). Stock Market Volatility and Business Cycle, *Journal of Applied Econometrics*, 573-93.

Hamilton, J.D. (1994). Time Series Analysis, Princeton, NJ: Princeton University Press.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica,* 357-84.

Kiefer, N.M., (1978). Discrete parameter variation: efficient estimation of a switching regression model, *Econometrica* 46, 427-434.

Kim, C.J. and C.R. Nelson (1999). *S*tate-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, Cambridge, Mass.: MIT Press.*

Krolzig, H.M.(1997). Markov-switching vector autoregressions. Modelling, statistical inference and applications to business cycle analysis, Springer, Berlin.

Lindgren, G., (1978). Markov Regime Models for Mixed Distributions and Switching Regressions. *Scandinavian Journal of Statistics*, 5, pp. 81-91.

Lutkephol , H. (1991). Introduction to Multiple Time Series Analysis, *Berlin: Springer-Verlag.*

McCulloch, R. E. and R. S. Tsay (1994). Statistical analysis of economic time series via markov switching models. Journal of Time Series Analysis, 15, 523–539.

Mosconi, R.and R. Seri (2006). Non-Causality in Bivariate Binary Time Series, J*ournal of Econometrics,* 132, 379-407.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, 347-70.

Perez-Quiros, G. and A. Timmermann (2001). Business cycle asymmetries in stock returns: evidence from higher order moments and conditional densities, *Journal of Econometrics* ,103, 259-306.

Psaradakis, Z., and M. Sola (1998). Finite-sample properties of the maximum likelihood estimator in autoregressive models with Markov switching, *Journal of Econometrics* 86, 369-86.

Psaradakis, Z., O.R Morten. and M. Sola (2005). Markov-Switching causality and the Money-Output Relationship, *Journal of Applied Econometrics*, 665-83.

Sims, C.A., J.H. Stock and M.W. Watson (1990). Inference in Linear Time Series Models with Some Unit Roots, *Econometrica*, 113-44.

Smith, D.R. (2008). Evaluating Specification Tests for Markov-Switching Time-Series Models, *Journal of Time Series Analysis*, Vol. 29, Issue 4, 629-652.

Toda, H.Y. and P.C.B Phillips (1993). Vector Autoregression and Causality, *Econometrica*, 1367-93.

Warne A., (2000). Causality and Regime Inference in a Markov Switching VAR, WP 18, Sveriges Riksbank, Stockholm.

Tables

TABLE 1

Maximum likelihood estimates

$\widehat{\mu}_0$	-0.687	(0.126)	$\widehat{g}_1$	0.118	(0.033)
$\widehat{\mu}_1$	0.293	(0.043)	$\widehat{\zeta}$	21.30	(2.8)
$\widehat{\phi}$	0.255	(0.053)	$\widehat{\xi}$	0.009	(0.011)
$\widehat{\sigma}^2$	0.450	(0.116)	$\widehat{\varphi}$	0.143	(0.013)
$\widehat{\delta}_0$	0.423	(0.125)			
$\widehat{\delta}_1$	0.186	(0.055)			

Note : **standard errors in parentheses.**

TABLE 2

Transition Matrix

	$S_{t-1}=0, Z_{t-1}=0$	$S_{t-1}=0, Z_{t-1}=1$	$S_{t-1}=1, Z_{t-1}=0$	$S_{t-1}=1, Z_{t-1}=1$
$S_t=0, Z_t=0$	0.810	0.021	0.950	0.033
$S_t=0, Z_t=1$	0.113	0.050	0.024	0.223
$S_t=1, Z_t=0$	0.015	0.019	0.015	0.063
$S_t=1, Z_t=1$	0.062	0.910	0.011	0.681

TABLE 3

Causality and Independence Tests

Hypothesis	p_{STD}	p_b
$H_{1 \nrightarrow 2}$	0.000	0.005
$H_{1 \nleftarrow 2}$	0.000	0.002
$H_{1 \perp\!\!\!\perp 2}$	0.000	0.006
$H_{1 \nLeftrightarrow 2}$	0.000	0.008