Causality and Regime Inference in a Markov Switching VAR

Anders Warne

September 16, 2000

Abstract: This paper analyses three Granger noncausality hypotheses within a conditionally Gaussian MS-VAR model. Noncausality in mean is based on Granger's original concept for linear predictors by defining noncausality from the 1-step ahead forecast error variance for the conditional expectation. Noncausality in mean-variance concerns the conditional forecast error variance, while noncausality in distribution refers to the conditional distribution of the forecast errors. Necessary and sufficient parametric conditions for noncausality are presented for all hypotheses. As an illustration, the hypotheses are tested using monthly postwar U. S. data on money and income. We find that money is not Granger causal in mean for income, but Granger causal in mean-variance, i.e. there is unique information in money for predicting the next period regime and the regime affects the uncertainty about the income forecast.

Keywords: Granger causality, Markov process, regime switching, vector autoregression.

JEL Classification Numbers: C32.

1. Introduction

The most widely used concept of causality in time series econometrics is due to Granger (1969). Founded on the linear least squares predictor for some information set $\mathcal{X}$, Granger's definition of causality states that a variable m is causal for a variable y if the variance of the 1-step ahead forecast error for y is smaller when the history of m is included in $\mathcal{X}$ than when it is excluded. Moreover, if these forecast error variances are equal, then m is said to be noncausal for y.

Granger causality has primarily been studied within linear vector autoregressive (VAR) models. For such models, the necessary and sufficient condition for m to be noncausal for y is that all coefficients on lags of m are zero in the equation describing y. If the roots to the VAR model are stable (outside the unit disc), then the Wald, LM, and LR statistics have their usual limiting χ^2 distribution (see e.g. Lütkepohl, 1991), while the case of some unit

---

Remarks: This research was initiated while I was visiting the Institute of Economics at the University of Copenhagen. I have benefited greatly from discussions with Michael Bergman, Henrik Hansen and received valuable comments from seminar participants at the University of Copenhagen, the University of Aarhus, IIES, the Institute of Statistics and Econometrics, Humboldt University, Berlin, and ESEM96 in Istanbul, Turkey. Part of this work was undertaken when I was at the IIES, Stockholm University, and financially supported by *Bankforskningsinstitutet* and *Jan Wallanders och Tom Hedelius' stiftelse för samhällsvetenskaplig forskning*. The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

roots implies that the limiting distribution can be nonstandard; see Sims, Stock, and Watson (1990) and Toda and Phillips (1993).

For linear statistical models, Granger noncausality in the above sense means that *all* moments of the 1-step ahead forecast errors are equal for y conditional on $\mathcal{X}$ and for y conditional on $\mathcal{X}$ less the history of m. Once we turn our attention to nonlinear models, however, this result is no longer true. For instance, if the error process is conditionally heteroskedastic, then m may contain unique information for predicting the conditional but not the unconditional variance of the forecast errors; see e.g. Granger, Engle, and Robins (1986).

Conditional variances have attracted considerable interest from, in particular, the literature on asset pricing in finance (see e.g. Ferson, 1993), but also from the literatures on central bank independence (see e.g. Rogoff, 1985) and on inflation versus price level targeting (see e.g. Svensson, 1999 and references therein). While the latter bodies of literature typically examine the variances of certain economic aggregates under different behavioral and/or institutional assumptions, the result that changes in the parameters reflecting these assumptions can affect the variances suggests that, from a time series perspective, the conditional variances can depend on the probability that these parameters change at some future date(s).

Since the publication of Hamilton's (1989) paper, where the mean growth rate of U. S. real GNP varies according to a latent 2-state Markov process, there has been a growing literature on regime switching in applied macroeconomic and financial time series analysis; see, for instance, Cecchetti, Lam, and Mark (1990), Diebold and Rudebusch (1996), Garcia and Perron (1996), Ravn and Sola (1995), and Sola and Driffill (1994). There are several theoretical reasons why such time series may be subject to switching regimes, e.g. changes in economic policy, as in Vredin and Warne (2000) and Weise (1999), or, as suggested by Jacobson, Lindh, and Warne (1998), fixed costs associated with the implementation of new technologies.

Although we may have good a priori reasons to suspect that recurring regime shifts have influenced the data, we often have considerably less information about when such shifts have occurred or how many shifts there have been. A Markov switching model may often be a good starting point under such circumstances.

In this paper, a q-state Markov switching vector autoregression (MS-VAR) model is used to study three related Granger noncausality concepts. Apart from the unconditional and conditional 1-step ahead forecast error variance versions mentioned above, the case of conditional independence will also be examined (see Chamberlain, 1982, and Florens and Mouchart, 1982). While it is difficult to justify the interest in conditional independence from economic theory, it is nevertheless a useful concept for, at least, two reasons. First, it can serve as a benchmark when we attempt to interpret parametric restrictions that imply e.g. noncausality in terms of the conditional mean and variance (typically the noncausality restrictions will not be unique). Second, the parametric restrictions can depend on other properties of the statistical model which, in turn, may contribute to making statistical inference about these restrictions highly hazardous.[1]

The remainder of the paper is organized as follows. The basic assumptions and the three noncausality concepts are presented in the next section, while the MS-VAR model is introduced in Section 3. The main theoretical results on regime inference and noncausality are offered and discussed in Section 4. As an illustration, the results are applied to monthly U. S. data on money and income in Section 5. The main findings of the paper are summarized in Section 6 and, finally, proofs of the Propositions are given in the Mathematical Appendix.

2. Granger Causality

2.1. *Notation and Basic Assumptions*

To be concrete, let m_t and y_t denote money and income, respectively, and let the time series of these variables up to and including period t be given by $\mathcal{M}_t \equiv \{m_\tau : \tau = t, t-1, \ldots, 1-p\}$ and $\mathcal{Y}_t \equiv \{y_\tau : \tau = t, t-1, \ldots, 1-p\}$, where p is a nonnegative integer. Also, let z_t denote a vector of other variables and $\mathcal{Z}_t \equiv \{z_\tau : \tau = t, t-1, \ldots, 1-p\}$ its time series. We decompose z_t into two vectors, $z_{1,t}$ and $z_{2,t}$, and define the n dimensional vector x_t such that $x_t \equiv [x_{1,t}' \; x_{2,t}']'$, with $x_{1,t} \equiv [y_t \; z_{1,t}']'$ and $x_{2,t} \equiv [m_t \; z_{2,t}']'$ being n_i dimensional ($i = 1, 2$). Hence, the time series of x_t up to and including period t can, for instance, be written $\mathcal{X}_t \equiv \{\mathcal{Z}_t, \mathcal{Y}_t, \mathcal{M}_t\} \equiv \{\mathcal{X}_{1,t}, \mathcal{X}_{2,t}\}$.

Suppose $\mathcal{X}_T$ is a vector valued time series of random variables and that there exists a density (probability) function $f_t(x_t|\mathcal{X}_{t-1}; \theta)$ for each $t \in \{1, 2, \ldots, T\}$. The parameters and the parameter space are denoted by θ and Θ, where Θ is a subset of $\mathbb{R}^K$, the K dimensional Euclidean space, and it is assumed that the density (probability) function is measurable in $x_t|\mathcal{X}_{t-1}$ for every $\theta \in \Theta$ and continuous in θ for every $x_t|\mathcal{X}_{t-1}$ in the sample space. The true value of θ is denoted by $\theta^* \in \Theta$. Finally, suppose that the conditional mean $E[x_t|\mathcal{X}_{t-1}; \theta^*]$ is finite and that the conditional covariance matrix $E[(x_t - E[x_t|\mathcal{X}_{t-1}; \theta^*])(x_t - E[x_t|\mathcal{X}_{t-1}; \theta^*])'|\mathcal{X}_{t-1}; \theta^*]$ is finite and positive definite for all finite t.

[1] For instance, if the restrictions are nonlinear, then the first order partial derivatives can be linearly dependent under the null hypothesis.

2.2. *Definitions*

Let u_{t+1} denote the 1-step ahead forecast error for y_{t+1} conditional on $\mathcal{X}_t$ and θ^* when the predictor is given by the expectations operator. That is,

$$u_{t+1} \equiv y_{t+1} - E[y_{t+1}|\mathcal{X}_t;\theta^*]. \tag{1}$$

By assumption u_{t+1} has conditional mean zero and positive and finite conditional variance σ_t^2. The concept of causality formulated by Granger (1969) concerns the optimal (minimum MSE) unbiased 1-step ahead linear least squares predictor. Although the notion cannot directly be translated into nonlinear models, three causality concepts, inspired by Granger's ideas, which have been proposed in the literature will be presented below.

A coarse version of Granger noncausality in nonlinear models is the following:

Definition 1: *m is said to be noncausal in mean for y if and only if for all t*

$$E[u_{t+1}^2;\theta^*] = E[\tilde{u}_{t+1}^2;\theta^*] < \infty, \tag{2}$$

where $\tilde{u}_{t+1} \equiv y_{t+1} - E[y_{t+1}|\mathcal{Z}_t,\mathcal{Y}_t;\theta^*]$.

Analogously, we say that m is Granger causal in mean for y if $E[u_{t+1}^2;\theta^*]$ is smaller than $E[\tilde{u}_{t+1}^2;\theta^*]$ for some t.

A logical refinement is to measure causal effects from the behavior of the conditional forecast error variance. It is argued in the introduction that economic behavior can influence the time profile of the conditional variance and, hence, it is possible that a variance based causality measure is unfit for detecting certain relevant causal effects. Let us therefore consider the following noncausality concept (see Granger, Engle, and Robins, 1986):

Definition 2: *m is said to be noncausal in mean-variance for y if and only if for all t*

$$E[u_{t+1}^2|\mathcal{X}_t;\theta^*] = E[\tilde{u}_{t+1}^2|\mathcal{Z}_t,\mathcal{Y}_t;\theta^*] < \infty. \tag{3}$$

Alternatively, m is Granger causal in mean-variance for y if the two random variables in (3) are different for some t.

Further refined noncausality concepts can be introduced by requiring a sequence of moments (e.g. the first to the fourth) of the distributions for u_{t+1} conditional on $\mathcal{X}_t$ and for $\tilde{u}_{t+1}$ conditional on $\mathcal{Z}_t,\mathcal{Y}_t$ to be equal. Since all such moments need not exist and economic theory rarely provides statements about the behavior of moments higher than the second, we will only consider the "limit" case. That is, noncausality in terms of the density (probability) functions for u_{t+1} given $\mathcal{X}_t$, denoted by $g_{t+1}(u_{t+1}|\mathcal{X}_t;\theta)$, and for $\tilde{u}_{t+1}$ given $\mathcal{Z}_t,\mathcal{Y}_t$, denoted by $h_{t+1}(\tilde{u}_{t+1}|\mathcal{Z}_t,\mathcal{Y}_t;\theta)$.

Definition 3: *m is said to be noncausal in distribution for y if and only if for all t and all but a finite number of probability zero events in the sample space*

$$g_{t+1}(u_{t+1}|\mathcal{X}_t;\theta^*) = h_{t+1}(\tilde{u}_{t+1}|\mathcal{Z}_t,\mathcal{Y}_t;\theta^*). \tag{4}$$

In other words, m is Granger causal in distribution for y if the conditional distribution for the 1-step ahead forecast error is not invariant with respect to the history of money. This definition of noncausality is equivalent to the definition in Chamberlain (1982) and Florens and Mouchart (1982), where m is said to be noncausal for y if $y_{t+1}|\mathcal{Z}_t,\mathcal{Y}_t$ is independent of $\mathcal{M}_t$; see also Florens and Mouchart (1985) for additional noncausality concepts.

While noncausality in distribution gives a natural measure of causal effects from a statistical perspective, it is difficult to envision an economic mechanism which implies that m is noncausal in mean-variance for y but not in distribution. However, since the three causality concepts are nested, noncausality in distribution provides a means for interpreting parametric restrictions implied by, say, noncausality in mean-variance.[2] Moreover, for some nonlinear models noncausality in mean and in mean-variance can, as will be shown below, be highly difficult to test when certain sets of restrictions depend in complicated ways on other properties of the model. Practical considerations can therefore force a researcher to focus on a stronger hypothesis. Finally, it is worth emphasizing that under noncausality in distribution, all predictors of y based on the distribution of y conditional on $\mathcal{X}$ are invariant with respect to the history of m. Hence, possibly biased predictors, such as the median, do not depend on $\mathcal{M}_t$.

3. A Markov Switching VAR Model

In this section, I shall present a Markov switching vector autoregressive (MS-VAR) model which is nested within the class of autoregressive models studied in Hamilton (1990) and Krishnamurthy and Rydén (1998). A version of the model is discussed in more detail in Warne (1999). We shall then examine the causality hypotheses in the next section.

To establish notation, let x_t be generated by the following MS-VAR model:

$$x_t = \mu_{s_t}D_t + \sum_{k=1}^{p} A_{s_t}^{(k)}x_{t-k} + \varepsilon_t, \qquad t = 1,2,\ldots,T, \tag{5}$$

where p is finite, $\varepsilon_t|s_t \sim N(0,\Omega_{s_t})$ and Ω_{s_t} is positive definite. The vector D_t is d dimensional and deterministic, e.g. a constant and centered seasonal dummies. The initial values $x_0,\ldots,x_{1-p}$ are taken as fixed.

The random state or regime variable s_t is unobserved, conditional on s_{t-1} independent of past x, and assumed to follow a q-state Markov process. In other words, $\Pr[s_t = j|s_{t-1} =$

[2] By nested it is understood that (4) implies (3) which implies (2). Alternatively, let $H^{(i)}$ be the set of all parametric functions of θ which are consistent with Definition i, for $i = 1,2,3$. Then, $H^{(3)} \subseteq H^{(2)} \subseteq H^{(1)}$.

$i, s_{t-2} = h_2, \ldots, \mathcal{X}_{t-1}] = \Pr[s_t = j | s_{t-1} = i] = p_{ij}$, for all t and $h_l, i, j = 1, 2, \ldots, q$, $l \geq 2$. The Markov transition probabilities satisfy $\sum_{j=1}^{q} p_{ij} = 1$ for all i. It is assumed that the Markov process is irreducible (no absorbing states) and we collect its parameters into

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1q} \\ \vdots & & \vdots \\ p_{q1} & \cdots & p_{qq} \end{bmatrix}. \tag{6}$$

By construction one eigenvalue of P is always equal to unity and to ensure that s_t is ergodic the remaining eigenvalues are assumed to lie inside the unit circle. The ergodic probabilities, $\Pr[s_t = j] = \pi_j$, are collected into π, where $P'\pi = \pi$.

The random matrices μ_{s_t}, $A_{s_t}^{(k)}$ and Ω_{s_t} depend only on the regime variable s_t. Specifically, if $s_t = j$, then $\mu_{s_t} = \mu_j$, $A_{s_t}^{(k)} = A_j^{(k)}$ while $\Omega_{s_t} = \Omega_j$. Since s_t is ergodic it follows that μ_{s_t}, $A_{s_t}^{(k)}$ and Ω_{s_t} are ergodic as well.

Karlsen (1990) establishes a sufficient condition for x_t to be covariance stationary. His condition applies directly here when D_t does not include any deterministically trending variables and it is noteworthy that it allows for unit and explosive roots within states as long as some of the $A_{s_t}^{(k)}$ matrices vary across states. Furthermore, the condition is also valid when D_t includes trending variables and the random matrices μ_{s_t} and $A_{s_t}^{(k)}$ satisfy certain (nonlinear) restrictions. The interested reader is also referred to Holst, Lindgren, Holst, and Thuvesholmen (1994).

In the next section we shall consider Markov chains that can be split into two independent processes (where one can be a single regime process). This allows coefficients in two subsystems of equation (5) to vary with the regime and, at the same time, be independent. Let $s_{1,t}$ and $s_{2,t}$ be a q_1 and a q_2 state Markov process, respectively, with $q = q_1 q_2$ and $s_{1,t}$ and $s_{2,t}$ independent, i.e. $p_{ij} = p_{i_1 j_1}^{(1)} p_{i_2 j_2}^{(2)}$ where $\Pr[s_{l,t} = j_l | s_{l,t-1} = i_l] = p_{i_l j_l}^{(l)}$ and $\sum_{j_l=1}^{q_l} p_{i_l j_l}^{(l)} = 1$ for $i_l = 1, \ldots, q_l$. Collecting the parameters into $P^{(1)}$ and $P^{(2)}$ and defining $s_t \equiv s_{2,t} + q_2(s_{1,t} - 1)$ for the pair $(s_{1,t}, s_{2,t})$ we have that $P = (P^{(1)} \otimes P^{(2)})$. While the restrictions implied by independence appear to be nonlinear, they are in fact linear. The reason is that the elements of each row of $P^{(l)}$ sum to unity.[3] Moreover, the Markov process $s_{l,t}$ is serially independent if and only if, for all $i_l, j_l = 1, \ldots, q_l$, $p_{i_l j_l}^{(l)} = \pi_{j_l}^{(l)}$. Hence, a serially uncorrelated Markov process is also equivalent to a set of linear constraints on the transition probabilities.

[3] The restriction are directly available from inspection of the $q \times q_1$ matrix $P\left(I_{q_1} \otimes \imath_{q_2}\right)$ and the $q \times q_2$ matrix $P\left(\imath_{q_1} \otimes I_{q_2}\right)$, where $\imath_{q_l}$ is a q_l dimensional unit vector.

4. Noncausality and Regime Inference

The issue of making inference about the Markov regime is closely linked with Granger causality. For instance, suppose that y_t is equal to μ_{1,s_t} plus an iid Gaussian residual. In this case, m will be noncausal in mean for y if and only if the history of m does not contain any unique information about the next period regime. This is trivially satisfied when the Markov process is serially uncorrelated but other, more interesting, possibilities also exist.

Although we focus on a conditionally Gaussian ε_t vector, the main result on regime inference (cf. Proposition 1 below) can be extended to a set of general conditions that the distribution function for (ε_t, s_t) must satisfy. This extension, which retains the assumptions about the regime process, is presented in the Mathematical Appendix and is used as a means towards proving Proposition 1 below.

Based on the $x_{i,t}$ vectors that were introduced in Section 2.1, consider the following partition of (5):

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} \delta_{1,s_t} \\ \delta_{2,s_t} \end{bmatrix} D_t + \sum_{k=1}^{p} \begin{bmatrix} \alpha_{11,s_t}^{(k)} & \alpha_{12,s_t}^{(k)} \\ \alpha_{21,s_t}^{(k)} & \alpha_{22,s_t}^{(k)} \end{bmatrix} \begin{bmatrix} x_{1,t-k} \\ x_{2,t-k} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}. \tag{7}$$

Furthermore, let us partition Ω_{s_t} conformably with $\epsilon_{i,t}$, i.e. $\Omega_{ij,s_t} \equiv E[\epsilon_{i,t}\epsilon_{j,t}' | s_t]$. We also have use for a finer partition of (5). For that partition, it suffices to give the income equation:

$$y_t = \mu_{1,s_t} D_t + \sum_{k=1}^{p} \left(a_{11,s_t}^{(k)} y_{t-k} + a_{12,s_t}^{(k)} z_{1,t-k} + a_{13,s_t}^{(k)} m_{t-k} + a_{14,s_t}^{(k)} z_{2,t-k} \right) + \varepsilon_{1,t}, \tag{8}$$

while ω_{11,s_t} is the variance of $\varepsilon_{1,t}$ conditional on s_t. We shall use the first partition when we discuss regime predictions while the second partition is used in conjunction with Granger causality.

4.1. *Regime Inference*

For expositional reasons, let us first assume that all regimes are known. The prediction of next period's income conditional on s_{t+1} and $\mathcal{X}_t$ is then[4]

$$E[y_{t+1} | s_{t+1}, \mathcal{X}_t] = y_{t+1} - \varepsilon_{1,t+1}. \tag{9}$$

Accordingly, the forecast error is given by $\varepsilon_{1,t+1}$ and the conditional forecast error variance by $\omega_{11,s_{t+1}}$. The necessary and sufficient condition for money not to Granger cause income in mean, mean-variance, and in distribution is that $a_{13,s_t}^{(k)}$ in equation (8) is equal to zero for all k and t.

[4] Formally, the expectation is also taken with respect to θ^* and $\mathcal{D}_T \equiv \{D_\tau : \tau = 1, \ldots, T\}$. To simplify notation, however, $\mathcal{D}_T$ and, sometimes, θ^* are not explicitly expressed as parts of the conditioning information.

Let us now drop the assumption that the regimes are known. While the regime variable s_t is independent of past x conditional on s_{t-1}, it can be predicted using only past x. Let $\Pr[s_{t+1}|\mathcal{X}_t]$ denote the probability of a particular state occurring at $t+1$ conditional on the information available at t. The prediction of next period's income is then given by

$$E[y_{t+1}|\mathcal{X}_t] = \sum_{s_{t+1}} E[y_{t+1}|s_{t+1}, \mathcal{X}_t] \Pr[s_{t+1}|\mathcal{X}_t]. \tag{10}$$

The role for money is different in (10) relative to (9) in that the history of money can now predict income by containing information which helps predict next period's state.

Since s_{t+1} is independent of $\mathcal{X}_t$ conditional on s_t it follows that

$$\Pr[s_{t+1}|\mathcal{X}_t] = \sum_{s_t} \Pr[s_{t+1}|s_t] \Pr[s_t|\mathcal{X}_t]. \tag{11}$$

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

This discussion presumes that the coefficients in the income equation vary freely with the regime. It is possible, however, that these coefficients vary with $s_{1,t+1}$ but not with $s_{2,t+1}$. Similarly, there may be information in $\mathcal{M}_t$ for predicting $s_{2,t+1}$ but not for predicting $s_{1,t+1}$. In such situations it may still be the case that the prediction of income in (10) does not depend on the history of money. This leads us to the first result.

Proposition 1: *The regime forecasts of $s_{1,t+1}$ and $s_{2,t+1}$ are independent and there is no information in $\mathcal{X}_{2,t}$ for predicting $s_{1,t+1}$, i.e.*

$$\Pr[(s_{1,t+1}, s_{2,t+1}) = (j_1, j_2)|\mathcal{X}_t; \theta^*] = \Pr[s_{1,t+1} = j_1|\mathcal{X}_{1,t}; \theta^*] \Pr[s_{2,t+1} = j_2|\mathcal{X}_t; \theta^*], \tag{12}$$

for all $j_1 \in \{1, \dots, q_1\}$ with $q_1 \geq 2$, $j_2 \in \{1, \dots, q_2\}$, and $t \in \{1, \dots, T\}$, if and only if either

(A1): (I) $P = (P^{(1)} \otimes P^{(2)})$, $\delta_{i,s_t} = \delta_{i,s_{i,t}}$, $\alpha^{(k)}_{ij,s_t} = \alpha^{(k)}_{ij,s_{i,t}}$, $\Omega_{ii,s_t} = \Omega_{ii,s_{i,t}}$, *and* $\Omega_{12,s_t} = 0$ *for all* $i, j \in \{1, 2\}$, $k \in \{1, \dots, p\}$, $s_t \in \{1, \dots, q\}$, *and*

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

(A2): $P = (\imath_{q_1} \pi^{(1)\prime} \otimes P^{(2)})$,

is satisfied.

Notice first that all restrictions in (A1) and (A2) are linear. Furthermore, if we change the restrictions in (A1.II) to $\alpha^{(k)}_{21,s_{2,t}} = 0$, then there is no information in $\mathcal{X}_{1,t}$ for predicting $s_{2,t+1}$. Moreover, in the Appendix it is shown that

Corollary 1: *If and only if condition* (A1.I) *is satisfied, then*

$$\Pr[(s_{1,t}, s_{2,t}) = (i_1, i_2)|\mathcal{X}_\tau; \theta^*] = \Pr[s_{1,t} = i_1|\mathcal{X}_\tau; \theta_1^*] \Pr[s_{2,t} = i_2|\mathcal{X}_\tau; \theta_2^*],$$

for all $i_1 \in \{1, \dots, q_1\}$ *with* $q_1 \geq 2$, $i_2 \in \{1, \dots, q_2\}$ *with* $q_2 \geq 2$, *and* $t, \tau \in \{1, \dots, T\}$, *with* $\theta^* = (\theta_1^*, \theta_2^*)$.

Hence, for the predictions of $s_{1,t}$ and $s_{2,t}$ to be independent, it is not sufficient that the Markov processes are independent. In fact, the joint distribution for x_t conditional on s_t (and $\mathcal{X}_{t-1}$) being equal to the product between the marginal distributions for $x_{l,t}$ conditional on $s_{l,t}$ (and $\mathcal{X}_{t-1}$) for $l = 1, 2$ must also be satisfied. Under these additional restrictions forecasting, filtering and smoothing inference about the two regime variables can be conducted independently. Additionally,

Corollary 2: *If and only if condition* (A1) *is satisfied, then*

$$\Pr[(s_{1,t}, s_{2,t}) = (i_1, i_2)|\mathcal{X}_\tau; \theta^*] = \Pr[s_{1,t} = i_1|\mathcal{X}_{1,\tau}; \theta_1^*] \Pr[s_{2,t} = i_2|\mathcal{X}_\tau; \theta_2^*],$$

for all $i_1 \in \{1, \dots, q_1\}$ *with* $q_1 \geq 2$, $i_2 \in \{1, \dots, q_2\}$, *and* $t, \tau \in \{1, \dots, T\}$, *with* $\theta^* = (\theta_1^*, \theta_2^*)$.

In the Appendix (see Lemma 2 and Lemma 3) I present necessary and sufficient conditions for conducting optimal inference on $s_{1,t}$ and $s_{2,t}$ independently in Markov switching models when the density function for $\varepsilon_t|s_t$ meets the criteria for conducting optimal inference on s_t using the algorithm in Hamilton (1994) and Kim (1994).

The intuition behind condition (A1) is, in fact, straightforward. Suppose $p = D_t = 1$, $n = q = q_1 = 2$, while $\epsilon_{2,t}$ is iid. The restrictions on Ω_{s_t} in (A1) are sufficient for the money equation residual to be iid. Now consider the experiment of drawing two m_t's, one for each regime, when y_{t-1} and m_{t-1} are fixed. The difference between these two draws is:

$$m_{t|s_t=2} - m_{t|s_t=1} = (\delta_{2,2} - \delta_{2,1}) + (\alpha_{21,2} - \alpha_{21,1})y_{t-1} + (\alpha_{22,2} - \alpha_{22,1})m_{t-1}. \tag{13}$$

The right hand side of (13) is zero for all vectors (y_{t-1}, m_{t-1}) when the coefficients in the money equation are constant across states. Accordingly, if these restrictions are satisfied, then $\Pr[s_t|\mathcal{Y}_t, \mathcal{M}_t] = \Pr[s_t|\mathcal{Y}_t, \mathcal{M}_{t-1}]$ and all information about s_t is found in the income equation. If the coefficient on money in that equation is zero for both states, then lags of money play no role for predicting regime switches.

4.2. *Granger Noncausality*

Before we present the next result, some additional notation is needed. Specifically, let

$$\bar{\mu}_{1,t} \equiv E[\mu_{1,s_{t+1}}|\mathcal{X}_t; \theta^*], \tag{14}$$

while

$$\bar{a}^{(k)}_{1r,t} \equiv E[a^{(k)}_{1r,s_{t+1}}|\mathcal{X}_t;\theta^*], \tag{15}$$

for all $r \in \{1,\ldots,4\}$ and $k \in \{1,\ldots,p\}$. The 1-step ahead forecast error for y is then given by $u_{t+1} = v_{t+1} + \varepsilon_{1,t+1}$, where

$$\begin{aligned} v_{t+1} \equiv \Big(\mu_{1,s_{t+1}} - \bar{\mu}_{1,t}\Big)D_{t+1} &+ \sum_{k=1}^{p}\Big(a^{(k)}_{11,s_{t+1}} - \bar{a}^{(k)}_{11,t}\Big)y_{t+1-k} \\ &+ \sum_{k=1}^{p}\Big(a^{(k)}_{12,s_{t+1}} - \bar{a}^{(k)}_{12,t}\Big)z_{1,t+1-k} + \sum_{k=1}^{p}\Big(a^{(k)}_{13,s_{t+1}} - \bar{a}^{(k)}_{13,t}\Big)m_{t+1-k} \\ &+ \sum_{k=1}^{p}\Big(a^{(k)}_{14,s_{t+1}} - \bar{a}^{(k)}_{14,t}\Big)z_{2,t+1-k}, \end{aligned} \tag{16}$$

is (conditional on $\mathcal{X}_t$) uncorrelated with $\varepsilon_{1,t+1}$.[5] A sufficient, but not necessary, condition for v_{t+1} to be mean zero stationary is that income is covariance stationary. Another possibility is that x_t is cointegrated. For the remainder of this section I shall assume that u_{t+1} is mean zero stationary.

PROPOSITION 2: *m is noncausal in mean for y if and only if either*

(B1): (A1)*; or*

(B2): (i) $\sum_{j=1}^{q}\mu_{1,j}p_{ij} = \bar{\mu}_1$, (ii) $\sum_{j=1}^{q}a^{(k)}_{1r,j}p_{ij} = \bar{a}^{(k)}_{1r}$, *and* (iii) $\bar{a}^{(k)}_{13} = 0$ *for all* $i \in \{1,\ldots,q\}$, $r \in \{1,\ldots,4\}$, *and* $k \in \{1,\ldots,p\}$,

is satisfied.

The nonlinear restrictions in condition (B2) state that given any regime i, the expected value of each random coefficient in the income equation is constant and that the expected values of the coefficients on lags of money are all zero. Hence, if we are willing to condition on P having full rank q, then condition (B2) implies that all coefficients in the income equation must be constant, while the coefficients on lags of money are all zero.

This observation can be generalized as follows:

COROLLARY 3: *Suppose that condition* (A2) *with* $\text{rank}[P^{(2)}] = q_2$ *is satisfied, then condition* (B2) *is equivalent to*

(B3): (i) $\sum_{j_1=1}^{q_1}\mu_{1,(j_1,j_2)}\pi^{(1)}_{j_1} = \bar{\mu}_1$, (ii) $\sum_{j_1=1}^{q_1}a^{(k)}_{1r,(j_1,j_2)}\pi^{(1)}_{j_1} = \bar{a}^{(k)}_{1r}$, *and* (iii) $\bar{a}^{(k)}_{13} = 0$ *for all* $j_2 \in \{1,\ldots,q_2\}$, $r \in \{1,\ldots,4\}$, *and* $k \in \{1,\ldots,p\}$.

This Corollary is of particular interest when $q = 2$. For such Markov processes, the rank of P can be either 1 or 2. In both cases, the conditions in Corollary 3 are satisfied and,

[5] Blix (1997) derives a general formula for the expectation of $x_{t+\tau}$, $\tau \geq 1$, conditional on $\mathcal{X}_t$ and applies it to rational expectations hypotheses.

hence, condition (B3) gives two sets of parameter constraints that are equivalent to the set of restrictions in (B2). If $q_2 = 1$, then the 2-state Markov process is serially uncorrelated, with $P = \imath_2 \pi^{(1)\prime}$ having rank 1. For this case, (B3) states that $\sum_{j_1=1}^{2} a_{13,j_1}^{(k)} \pi_{j_1}^{(1)} = 0$ for all k. On the other hand, if $q_2 = 2$, then the Markov process is serially correlated with $P = P^{(2)}$. Now, condition (B3) states that all coefficients in the income equation are constant across the regimes, and the coefficients on lags of money are zero.

The number of restrictions under (B2) and (B3) are typically very different. Suppose that $n = 2$ and $D_t = 1$ for all t in the 2-state MS-VAR model. Here we find that there are $3p + 1$ restrictions in (B2), whereas (B3) with $q_2 = 1$ has $p + 1$ restrictions (the p constraints on $a_{13,j}^{(k)}$ above, and the reduced rank constraint $p_{11} = p_{21}$), and (B3) with $q_2 = 2$ has $3p + 1$.

One difference between the two (B3) sets of restrictions is that the $q_2 = 1$ set contains nonlinear restrictions, while the $q_2 = 2$ set only contains linear constraints. For the purpose of testing and conducting statistical inference, the restrictions in (B2) are *too* general. For instance, suppose condition (B3) with $q_2 = 1$ is satisfied. It then follows that the $(3p + 1) \times (8p + 12)$ matrix with partial derivative of the (B2) constraints with respect to θ has rank equal to p or $p + 1$.[6] Under the assumption that the ML estimator of θ is asymptotically normal with a positive definite covariance matrix, the limiting distributions of the Wald, LM, and LR statistics are generally unknown. In fact, the only case when the limiting distributions of these statistics is known is when P has full rank 2, i.e. when (B3) with $q_2 = 2$ in Corollary 3 holds.

The problems with testing (B2) for general choices of q are even more severe. Unless (B3) with $q_2 = q$ is satisfied, the nonlinear restrictions in (B2) will always have a matrix with partial derivatives with respect to the parameters which has a reduced row rank. Moreover, the exact rank of this matrix depends not only on the rank of P, but also on the other parameters of the model. Let us therefore turn to more restrictive forms of noncausality.

PROPOSITION 3: *m is noncausal in mean-variance for y if and only if either*

(C1): (A1)*; or*

(C2): (i) (B2), (ii) $\sum_{j=1}^{q}[(\mu_{1,j} - \bar{\mu}_1) \otimes (\mu_{1,j} - \bar{\mu}_1)]p_{ij} = \sigma_\mu$, (iii) $\sum_{j=1}^{q}[(a_{1r,j}^{(k)} - \bar{a}_{1r}^{(k)}) \otimes (a_{1s,j}^{(l)} - \bar{a}_{1s}^{(l)})]p_{ij} = \sigma_{r,s}^{(k,l)}$, (iv) $\sum_{j=1}^{q}[(\mu_{1,j} - \bar{\mu}_1) \otimes (a_{1r,j}^{(k)} - \bar{a}_{1r}^{(k)})]p_{ij} = \sigma_{\mu,r}^{(k)}$, (v) $\sum_{j=1}^{q} \omega_{11,j} p_{ij} = \sigma_\omega$, *and* (vi) $a_{13,j}^{(k)} = 0$ *for all* $i, j \in \{1, \ldots, q\}$, $r, s \in \{1, 2, 4\}$, *and* $k, l \in \{1, \ldots, p\}$,

is satisfied.

Compared with the conditions in Proposition 2 we now have that the coefficients on money in the income equation must be zero for all lags and regimes when money is noncausal in mean-variance for income. Moreover, under (C2), the mean and the covariances

[6] The rank is p if and only if $\mu_{1,1} = \mu_{1,2}$ and $a_{1r,1}^{(k)} = a_{1r,2}^{(k)}$ for all $k \in \{1, \ldots, p\}$ and $r \in \{1, 3\}$.

of all random coefficients in the income equation are constant conditional on the previous regime. Notice that the covariance restrictions under this condition are algebraically equivalent to $\sum_{j=1}^{q}(\mu_{1,j} \otimes \mu_{1,j})p_{ij} = \sigma_\mu + (\bar{\mu}_1 \otimes \bar{\mu}_1)$, etc.

The problems of testing and conducting statistical inference that were discussed in the (B2) case above remain under the stricter condition (C2). If we are willing to condition on P having full rank q, however, then (C2) is equivalent to (B2) and the additional requirement that the variances of $\varepsilon_{1,t+1}$ are constant across states. The following Corollary generalizes this result.

COROLLARY 4: *Suppose that condition* (A2) *with* $\mathrm{rank}[P^{(2)}] = q_2$ *is satisfied, then condition* (C2) *is equivalent to*

(C3): (i) (B3), (ii) $\sum_{j_1=1}^{q_1}[(\mu_{1,(j_1,j_2)} - \bar{\mu}_1) \otimes (\mu_{1,(j_1,j_2)} - \bar{\mu}_1)]\pi_{j_1}^{(1)} = \sigma_\mu$, (iii) $\sum_{j_1=1}^{q_1}[(a_{1r,(j_1,j_2)}^{(k)} - \bar{a}_{1r}^{(k)}) \otimes (a_{1s,(j_1,j_2)}^{(l)} - \bar{a}_{1s}^{(l)})]\pi_{j_1}^{(1)} = \sigma_{r,s}^{(k,l)}$, (iv) $\sum_{j_1=1}^{q_1}[(\mu_{1,(j_1,j_2)} - \bar{\mu}_1) \otimes (a_{1r,(j_1,j_2)}^{(k)} - \bar{a}_{1r}^{(k)})]\pi_{j_1}^{(1)} = \sigma_{\mu,r}^{(k)}$, (v) $\sum_{j_1=1}^{q_1} \omega_{11,(j_1,j_2)}\pi_{j_1}^{(1)} = \sigma_\omega$, *and* (vi) $a_{13,j}^{(k)} = 0$ *for all* $r,s \in \{1,2,4\}$, $j \in \{1,\ldots,q\}$, $j_2 \in \{1,\ldots,q_2\}$, *and* $k,l \in \{1,\ldots,p\}$.

In the case when $q = 2$, we now have that condition (C3) only contains linear restrictions. To see this, note that $q_2 = 1$ is equivalent to the restrictions $p_{11} = p_{21}$, and $a_{13,j}^{(k)} = 0$ for all j and k. Hence, compared with condition (B3) we now have p additional restrictions. Similarly, the case when $q_2 = 2$ means that all coefficients in the income equation are constant across states, that the coefficients on lags of money are zero, and that the conditional variance of $\varepsilon_{1,t+1}$ is constant across states and over time.

For $q \geq 3$, however, it is difficult to formulate restrictions which are equivalent to (C2) conditional on various choices of $\mathrm{rank}[P]$, but which do not suffer from linearly dependent partial derivatives. Let us therefore consider the limiting case.

PROPOSITION 4: *m is noncausal in distribution for y if and only if either*

(D1): (A1)*; or*

(D2): (i) (A2), (ii) $\mu_{1,j} = \mu_{1,j_1}$, (iii) $a_{1r,j}^{(k)} = a_{1r,j_1}^{(k)}$, (iv) $a_{13,j}^{(k)} = 0$, *and* (v) $\omega_{11,j} = \omega_{11,j_1}$ *for all* $j \in \{1,\ldots,q\}$, $r \in \{1,2,4\}$, *and* $k \in \{1,\ldots,p\}$,

is satisfied.

Hence, when the distribution for u_{t+1} conditional on $\mathcal{X}_t$ is invariant with respect to $\mathcal{M}_t$, then the parameters of the MS-VAR model satisfies linear restrictions. Accordingly, the matrix with partial derivatives of the constraints with respect to the parameters has full row rank. In addition, statistical inference can be conducted with the usual limiting distributions provided that the ML estimator of θ is asymptotically normal with a positive definite covariance matrix; for some preliminary results on ML estimation and the asymptotic distribution, the reader is referred to Bickel and Ritov (1996) and Bickel, Ritov, and Rydén (1998).

The assumption that u_{t+1} is (weakly) stationary means that noncausality in distribution implies noncausality in mean-variance. In other words, [(D1) or (D2)] is a sufficient condition for [(C1) or (C2)]. Moreover, by Corollary 4 we have that the reverse is also true when $q = 2$ since (D2) is equivalent to (C3) with $\text{rank}[P^{(2)}] = q_2$ for such models. In fact, this observation can be generalized as follows:

COROLLARY 5: *Suppose* $\text{rank}[P] \in \{1, q\}$*, then* m *is noncausal in distribution for* y *if and only if it is noncausal in mean-variance for* y*.*

For general $\text{rank}[P]$, however, noncausality in mean-variance does not imply noncausality in distribution. To understand this better, consider the following example. Suppose that $q_1 = q_2 = 2$, with P satisfying condition (A2), and $\pi_1^{(1)} = \pi_2^{(1)} = .5$, while $P^{(2)}$ has full rank. Moreover, suppose that $y_t = \varepsilon_{1,t}$, where $\omega_{11,(1,1)} = \omega_{11,(2,2)} = 1$ and $\omega_{11,(1,2)} = \omega_{11,(2,1)} = 2$. For this model it can easily be verified that condition (C3) in Corollary 4 is satisfied, with $\sum_{j_1=1}^{2} \omega_{11,(j_1,j_2)} \pi_{j_1}^{(1)} = 1.5$ for both $j_2 \in \{1, 2\}$. Hence, m is noncausal in mean-variance for y. However, m can be causal in distribution for y since $\omega_{11,(j_1,j_2)}$ depends on $s_{2,t+1}$ (i.e. given j_1, the value of $\omega_{11,(j_1,j_2)}$ varies with j_2) and $\mathcal{M}_t$ can contain unique information for predicting $s_{2,t+1}$.

This example illustrates the intuition behind the linear restrictions in (D2). That is, noncausality in distribution requires that the income equation coefficients are invariant with respect to those regimes which $\mathcal{M}_t$ can help predict. The moment restrictions in (C2) are not strict enough to rule out all cases when $\mathcal{M}_t$ is useful for predicting those regimes which influence the income equation coefficients.

Finally, while the concepts of weak and strong exogeneity in Engle, Hendry, and Richard (1983) have primarily been applied to linear time series models, they may also be examined here since the definition of noncausality in distribution corresponds to their definition of Granger noncausality. It is then straightforward to verify that condition (D1) is sufficient for x_1 (x_2) to be weakly (strongly) exogenous for x_2 (x_1). Condition (D2), however, has no such implications because the sequential cut condition need not be satisfied.

5. An Application to Money and Income

In this section we shall analyse the causality restrictions for monthly U. S. data on money and income. The variables are M1 and industrial production for the sample period 1959:1 to 1995:2. Both series are seasonally adjusted (as in many previous studies using these data; e.g. Christiano and Ljungqvist, 1988) and taken from Citibase.[7] To avoid numerical problems in the estimation the log levels of the variables are multiplied by 1200. Moreover, D_t is a constant.

Initially we look at a single regime VAR model with 12 lags for the levels and 11 lags for a first differences. Some of the evidence is presented in Table 1. From the top panel it can be seen that we can reject that money is noncausal (in mean) for income at the 5 percent level of marginal significance for the log levels, but not for the first differences. This result is thus consistent with what has been found by e.g. Christiano and Ljungqvist (1988). Moreover, if we include a linear trend in the vector of deterministic variables (as suggested by Stock and Watson, 1989), the p-value for the levels model is unaffected, while it falls to 6.5 percent for the first difference specification.

In the bottom panel we have computed the likelihood ratio statistics (trace) for the hypotheses of 2 and at least 1 unit root against the alternatives of fewer unit roots. For the levels model we cannot reject the hypothesis of 2 unit roots,[8] while for the first differences we can reject the hypothesis of at least 1 unit root against the alternative that the variables are stationary. Looking at the modulus of the eigenvalues for the levels system we find that the largest eigenvalue is .999 while the second largest is .987. Hence, by the results in e.g. Sims et al. (1990) the limiting distribution for the F statistic in the levels model with a constant is not $F(12, \infty)$ if there are 2 unit roots and the actual p-value may very well be higher than that reported in Table 1.[9]

Still, these two models both seem to be misspecified. In particular, there are strong signs of conditional heteroskedastocity for the residuals in both systems and in the levels specification there are also signs of serial correlation. If an MS-VAR model with a serially correlated regime process has generated the data, then we would expect to find such forms of misspecification. If this is indeed the case then, as we have seen from the theoretical analysis above, the hypothesis of Granger noncausality (in mean) needs to be parametrically respecified.

[7] The use of seasonally adjusted data is far from ideal. I have chosen to use such data here to facilitate comparisons with other studies. For more recent studies on the issue of money-income causality and the use of nonlinear time series models, see e.g. Swanson (1998) and Rothman, Dijk, and Franses (1999).

[8] The p-values have been calculated using the software developed in connection with MacKinnon, Haug, and Michelis (1999) and can be downloaded from http://qed.econ.queensu.ca/pub/faculty/mackinnon/johtest/ .

[9] See e.g. Warne (1997) for an example of how the limiting distribution of the Wald statistic compares with the χ^2 distribution once unit roots are taken into account.

In Table 2 some stylized facts[10] about the behavior of money and income growth are presented for an MS-VAR model with two states and two lags.[11] On average, income grows at about 3 percent per year with a standard deviation of approximately 11 percent. The average growth in money is somewhat higher than that of income, while the standard deviation is about 5 percent. Moreover, income and money growth do not seem to be contemporaneously correlated.

Turning to the two states, we find that the estimates of the MS-VAR model roughly suggests a zero average income growth in State 1 and a 50 percent higher volatility than on average. In State 2, income growth is roughly 4 percent and the volatility is quite modest. Money, on the other hand, grows at about the same rate in both states and is somewhat more volatile in State 1. Concerning the contemporaneous correlations, there is a negative correlation in State 1 and a positive in State 2. Both correlations are, however, very small.

The estimated smooth probabilities of being in State 1 are graphed in the upper box of Figure 1. The shaded regions represent the periods from peak to trough according to the NBER dating scheme. In the lower box, the maximum posterior estimates of the regime process are depicted, i.e.

$$\hat{s}_t = \arg \max_{j \in \{1,2\}} \Pr\left[s_t = j \,\middle|\, \mathcal{X}_T; \hat{\theta} \right].$$

Here we see that the estimated regimes tend to switch from State 2 (1) to State 1 (2) around the NBER peaks (troughs). Summing up, these simple moments and plots suggest that we can interpret State 1 as “the bad state” and State 2 as “the good state”.[12]

With $q = 2$ it follows that either q_1 or q_2 is equal to unity. Accordingly, the coefficients in the money and income equations cannot both vary over time and be independent. The ML estimates of all 28 parameters are reported in Table 3.[13] The estimates are computed via the EM algorithm; for more details the reader is referred to Lindgren (1978), Hamilton (1990, 1994), and references therein. Standard errors for the point estimates are calculated from the conditional scores (as in Hamilton, 1996) and are reported within parentheses. The point estimates for the bad state (Regime 1) are generally more uncertain than those for the

---

[10] Formulas for computing the mean and the autocovariance conditional on the state are given in Warne (1999); see also Timmermann (2000). Standard errors are given within parenthesis. The latter statistics are computed using the delta method, with numerical partials of the moment expressions. The covariance matrix for the ML estimator of θ has been estimated from the conditional scores.

[11] Based on the misspecification tests discussed in Hamilton (1996) this model is consistent with the data; see Warne (1999) for details.

[12] The interpretation of the two regimes is discussed in more detail by Warne (1999).

[13] The statistic $\max|\mathrm{eig}(\hat{A})|$ refers to Karlsen’s (1990) condition for stationarity. If the true value of this statistic is less than unity, then x_t is stationary.

good state. This is related to the good state being roughly 3 times as likely (unconditionally) as the bad state.[14]

The sets of necessary and sufficient conditions for money to be conditionally uninformative about the regime process that influences income and to be noncausal for income, respectively, are given in Table 4 for the case when $n = q = 2$. Since the constraints in (A1) when $q_1 = 1$ and $q_2 = 2$ is a special case of (C3) (and thus of (D2)), we shall not test this hypothesis.

Let us first consider the hypothesis that, conditional on the state, money does not Granger cause income. That is, $\alpha^{(1)}_{12,s_t} = \alpha^{(2)}_{12,s_t} = 0$ for each state separately. The results from using Wald and F-statistics are reported in Table 5. For the good state, this hypothesis is rejected at the 5 percent level, and for the bad state at the 60 percent level. Hence, money does not cause income when we know that next period's state is the bad state, while it may cause income if next period's state is the good state.

Next, consider the hypothesis that the history of money is uninformative about next period's state. According to the evidence in Table 5, both sets of restrictions, (A1) with $q_1 = 2, q_2 = 1$ and (A2) (see Table 4), are strongly rejected by the data. Moreover, if (A2) is false, then the Markov process is serially correlated. Based on the point estimates of the transition probabilities in Table 3, good (bad) states are typically followed by good (bad) states. Together with the conclusion from testing (A2), this suggests that the regime process is subject to positive serial correlation.

While various aspects of the noncausality hypotheses have already been examined, the tests of all sufficient conditions have not been addressed. From Proposition 2 and Corollary 3 we have that if and only if [(B1) or (B3) and $q_2 = 1$ or (B3) and $q_2 = 2$] is true, then money is noncausal in mean for income. Given the evidence about (A2) it is not surprising that [(B3) and $q_2 = 1$] is strongly rejected as well. Now, if (A2) is false then P must have full rank 2 and, thus, $q_2 = 2$. According to the results in Table 5 the hypothesis [(B3) and $q_2 = 2$] is rejected at the 10 percent level, but not at the 5 percent level. Hence, there is some evidence that the conditional mean of income is invariant with respect to the history of money.

Finally, from Proposition 3, Proposition 4, and Corollary 4 we know that money does not Granger cause income in mean-variance and in distribution if and only if [(C1) or (C3) and $q_2 = 1$ or (C3) and $q_2 = 2$] is true. These three sets of restrictions are, individually,

[14] The estimated number of observation for State 1 is approximately 111 and for State 2 about 320. These estimates are calculated as $\sum_{t=1}^{T} \Pr[s_t = j | \mathcal{X}_T; \hat{\theta}]$. Note that this number is approximately equal to $\hat{\pi}_j T$, where $\hat{\pi}_j$ is the ML estimator of π_j; for details see Warne (1999). 95 percent confidence intervals for the estimated number of observations are given within parenthesis. These bands have been computed from the estimated standard error for $\hat{\pi}_j$.

rejected at conventional levels of marginal significance, suggesting that there is information in money about the conditional forecast error variance of income.

Reestimating the MS-VAR model under the noncausality in mean restrictions (B3) with $q_1 = 1$ and $q_2 = 2$ (i.e. the hypothesis that cannot be rejected at the 5 percent level of marginal significance) we obtain the results presented in Tables 6 and 7. The conditional mean of income growth is now, by assumption, equal for the two states. The remaining estimated moments (cf. Table 6) are, however, very similar to those reported in Table 4. Moreover, from Table 7 it can be seen that neither the estimated money equation parameters nor the Markov probabilities are greatly affected by the imposed restrictions on the income equation. Hence, it seems reasonable to conclude that an MS-VAR model of money and income growth where the income equation coefficients are constant across states, the coefficients on lags of money are zero, but where the variance of the income equation residual varies with the regime, is a suitable representation of the data.

To examine the effects on the standard Granger noncausality tests for single regime VARs (cf. Table 1) when data has actually been generated by an MS-VAR model, a small Monte Carlo study has been undertaken. As the null model, I selected the MS-VAR specification in Table 7, i.e. when the growth rate of money is noncausal in mean for the growth rate of income. The alternative model is given by the MS-VAR model in Table 3. Residuals have been drawn from a standard normal bivariate distribution and the regimes have been generated from a uniform distribution over the interval $[0, 1]$, with the initial regime determined by the ergodic probability.[15]

In the top panel of Table 8 the estimated critical values at the 10, 5, and 1 percent level are given along with p-values taken from the estimated distributions. Since data have been generated by a model with 2 unit roots it is not surprising that we can no longer reject the null hypothesis that money does not Granger causes income in the single regime VAR system for the levels at the 5 percent level (the estimated p-value is roughly 12 percent). Looking at the size distortions in the bottom panel we find that the levels test is highly oversized. Part of this is most likely due to not using a reference distribution in Table 1 that takes the number of unit roots into account, but the first difference test is also oversized and, hence, the effect of misspecification may also be very important.[16] Still, despite size distortions, the size adjusted power of these tests is decent under the alternative model. At the 5 percent level, we reject a false null in about 62 percent of the cases for the levels VAR and in 71 percent of the cases for the first difference VAR.

[15] The total number of replications is 10,000 for each data model, the initial values from the actual series have been used, and the levels series have been directly generated from the first differences. RATS code for the simulation experiment is available from my web site at: http://www.research.texlips.org/.

[16] For samples of the size we are dealing with here (421 usable observations), we may expect the reference distribution to provide a good approximation when the estimated model is consistent with the data.

Summing up, if data is generated by a bivariate 2-state conditionally Gaussian MS-VAR(2) model, then money Granger causes income for monthly U. S. data in log growth rates. There is some evidence that money is noncausal in mean for income, but the hypotheses implying noncausality in mean-variance (distribution) are all strongly rejected at conventional marginal levels of significance. This can be contrasted with single regime VAR's where the opposite conclusion is reached in first difference models. In fact, noncausality in such models is very robust with respect to the selection of sample period and lag order (except for very short lag orders, when the estimated residuals do not pass standard serial correlation diagnostics). For example, Christiano and Ljungqvist (1988) show that money Granger causes income in a bivariate single regime VAR (with 12 lags) for log levels, but not for log growth rates. They argue, based on evidence from bootstrapped empirical distributions, that the conclusion for the first difference model is wrong. The evidence from the 2-state MS-VAR(2) model supports Christiano and Ljungqvist's conclusion that money does Granger cause income in the bivariate case, although for a very different reason. Specifically, the results in this paper suggest that the "cause" stems from a signal extraction problem where the history of money is useful for predicting the regime variable.

The suggestion that money and income are correlated because of a signal extraction problem rather than a causal link from money to income goes back to (at least) Lucas (1972) in the macroeconomics literature. The signal extraction problem in the (reduced form) regime switching VAR is primarily about the uncertainty of the regimes, but it may also, at a deeper level, reflect a Lucas type problem through the parameters.

6. Concluding Remarks

The concern in this paper is the determination of a set of (economically and statistically meaningful) parametric Granger noncausality restrictions for a q-state MS-VAR model. To this end we examine three related types of noncausality that have been suggested elsewhere in the literature. The starting point for these concepts is the 1-step ahead forecast errors for the conditional expectations operator.

The weakest form of noncausality that we consider is based on Granger's original version, called noncausality in mean, where the forecast error variances are compared for two information sets, with one set being a strict subset of the other. The second form compares both the conditional means and the conditional variances (noncausality in mean-variance), while the third form deals with the conditional distribution of the forecast errors (noncausality in distribution or conditional independence).

It is shown that noncausality in any one of the three forms is not associated with a unique set of restrictions on the parameters of the MS-VAR. For each noncausality concept, however, the number of such sets is finite and depends on the dimension of the observable variable vector and on the number of Markov regimes. The noncausality in mean and in mean-variance cases generally result in some of these 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, thereby making these concepts difficult to deal with in applied analysis. For noncausality in mean-variance, however, the restrictions are always linear when we condition on either a full or a unit rank, as we can e.g. do when the number of regimes is exactly two.

Noncausality in distribution, on the other hand, always means that some set of linear restrictions is satisfied by the MS-VAR model. For each such set, the number of restrictions is known and does not depend on the rank of the Markov transition matrix. Accordingly, statistical inference may be conducted using standard methods as long as the ML estimator of the parameters is asymptotically normal; see e.g. Bickel and Ritov (1996), Bickel, Ritov, and Rydén (1998), and Giudici, Rydén, and Vandekerkhove (1999)

In the process of discussing Granger noncausality in MS-VAR models, we have also studied regime inference in some detail. Specifically, one quickly realizes that if, say, money is noncausal in mean for income, then the history of money cannot contain any unique information for predicting those next period states that influence the conditional mean of income. In the case of two regimes, this means that (i) the Markov regimes do not influence the conditional mean of income, (ii) money is conditionally independent of the regimes, or (iii) the regime process is serially uncorrelated.

To illustrate the concepts we have tested the implied restrictions on U. S. monthly data on the first differences of money and income for the 2 regime case as well as for the levels and first differences in single regime VARs. Here we have found some evidence supporting the view that money is not Granger causal in mean for income. In particular, the income equation may very well have constant coefficients, with those on lags of money being zero. Still, the conditional variance of income seems to vary with the regime process and the history of money appears to contain unique information for predicting the next period regime. Accordingly, the hypothesis that money is Granger causal in mean-variance (and in distribution) for income is strongly rejected by the data.

To examine the effects of not taking the regime process into account when testing Granger noncausality in standard VAR models a small Monte Carlo study has been undertaken. We find the F-test in the single regime levels model is highly oversized and that the F-test for the single regime first difference model is also oversized (although to a lesser extent) when data is generated from a 2-state MS-VAR model for the first differences. An explanation for the poor performance of the levels test is, of course, that the reference distribution does not take unit roots into account and is therefore asymptotically invalid. But since the test in the first difference model is also oversized, model misspecification also seems to play an important role. Furthermore, the (size corrected) power of the single regime tests is decent against the alternative we consider, but given the sample size, far from spectacular.

TABLE 1: Evidence from single regime VAR models on money and income causality and cointegration in the U. S. for the sample 1959:1–1995:2.

Granger Noncausality

System	Hypothesis	# d. f.	F	p-value
y	$a_{12}^{(k)} = 0$	(12,397)	2.034	.020
m	$k = 1, \dots, 12$			
Δy	$a_{12}^{(k)} = 0$	(11,399)	1.534	.117
Δm	$k = 1, \dots, 11$			

Cointegration Tests

System	Hypothesis	eigenvalue	LR_{tr}	p-value
y	# unit roots = 2	.022	9.38	.33
m	# unit roots ≥ 1	.000	.01	.92
Δy	# unit roots = 2	.076	53.59	.00
Δm	# unit roots ≥ 1	.047	20.45	.00

TABLE 2: ML estimates and standard errors of conditional and unconditional means and covariances for a 2-state MS-VAR(2) model of money and income growth in the US for the sample 1959:1–1995:2.

Unconditional moments			
variable	mean	variance	covariance
Δy	3.32	118.07	
	(.98)	(17.05)	.57
Δm	5.98	34.48	(5.21)
	(.54)	(3.83)	
Conditional moments			
	Regime 1		
Δy	.63	286.86	
	(3.22)	(41.70)	−10.75
Δm	5.21	59.24	(18.89)
	(1.04)	(9.85)	
	Regime 2		
Δy	4.22	58.22	
	(.69)	(6.67)	3.44
Δm	6.24	29.91	(3.17)
	(.61)	(3.51)	

TABLE 3: ML estimates and standard errors for a 2-state MS-VAR(2) model of money and income growth in the US for the sample 1959:1–1995:2.

Income equation			*Money equation*		
coefficient	$s_t = 1$	$s_t = 2$	coefficient	$s_t = 1$	$s_t = 2$
δ_{1,s_t}	−1.006	1.648	δ_{2,s_t}	4.742	2.410
	(2.897)	(.759)		(1.444)	(.446)
$\alpha^{(1)}_{11,s_t}$	.447	.274	$\alpha^{(1)}_{21,s_t}$	.067	−.025
	(.128)	(.057)		(.070)	(.035)
$\alpha^{(1)}_{12,s_t}$	.378	−.104	$\alpha^{(1)}_{22,s_t}$	.232	.476
	(.359)	(.112)		(.122)	(.056)
$\alpha^{(2)}_{11,s_t}$	−.084	.127	$\alpha^{(2)}_{21,s_t}$	−.031	−.064
	(.144)	(.054)		(.083)	(.033)
$\alpha^{(2)}_{12,s_t}$	−.167	.274	$\alpha^{(2)}_{22,s_t}$	−.149	.203
	(.312)	(.110)		(.137)	(.056)
Ω_{11,s_t}	242.546	46.535	Ω_{22,s_t}	55.940	14.483
	(32.767)	(5.297)		(9.766)	(1.426)
Ω_{12,s_t}	−19.168	2.872			
	(15.268)	(1.734)			
Markov probabilities					
p_{11}	.744		p_{22}	.914	
	(.071)			(.030)	

max $|\text{eig}(\hat{A})| = .408$ $\ln L(\mathcal{X}_T;\hat{\theta}) = -2852.6$ $\hat{\pi}_1 = .251$ (.060)

Estimated # obs: $s_t = 1: 111$ (59, 163); $s_t = 2: 320$ (268, 372)

TABLE 4: Noncausality and conditional regime independence restrictions in a 2-state MS-VAR(p) model with $x_t = (\Delta y_t, \Delta m_t)$.

Hypothesis	Restrictions	# restrictions
(A1); $q_1 = 1, q_2 = 2$	$\delta_{1,s_t} = \delta_1$, $\alpha^{(k)}_{11,s_t} = \alpha^{(k)}_{11}$, $\alpha^{(k)}_{12,s_t} = 0$, $\Omega_{11,s_t} = \Omega_{11}$, $\Omega_{12,s_t} = 0$	3p+4
(A1); $q_1 = 2, q_2 = 1$	$\delta_{2,s_t} = \delta_2$, $\alpha^{(k)}_{21,s_t} = \alpha^{(k)}_{21}$, $\alpha^{(k)}_{22,s_t} = \alpha^{(k)}_{22}$, $\Omega_{22,s_t} = \Omega_{22}$, $\Omega_{12,s_t} = 0$, $\alpha^{(k)}_{12,s_t} = 0$	4p+4
(A2); $q_1 = 2, q_2 = 1$	$p_{11} = p_{21}$	1
(B3); $q_1 = 1, q_2 = 2$	$\delta_{1,s_t} = \delta_1$, $\alpha^{(k)}_{11,s_t} = \alpha^{(k)}_{11}$, $\alpha^{(k)}_{12,s_t} = 0$	3p+1
(B3); $q_1 = 2, q_2 = 1$	$p_{11} = p_{21}$, $\sum_{j=1}^{2} \alpha^{(k)}_{12,j}\pi_j = 0$	p+1
(C3); $q_1 = 1, q_2 = 2$	$\delta_{1,s_t} = \delta_1$, $\alpha^{(k)}_{11,s_t} = \alpha^{(k)}_{11}$, $\alpha^{(k)}_{12,s_t} = 0$, $\Omega_{11,s_t} = \Omega_{11}$	3p+2
(C3); $q_1 = 2, q_2 = 1$,	$p_{11} = p_{21}$, $\alpha^{(k)}_{12,s_t} = 0$	2p+1

TABLE 5: Wald and F-tests of the Granger noncausality restrictions in the 2-state MS-VAR(2) model of money and income growth in the US for the sample 1959:1–1995:2.

Hypothesis	# rest.	W	p-value	F	p-value
$\alpha_{12,1}^{(k)} = 0$	2	1.113	.573	.540	.583
$\alpha_{12,2}^{(k)} = 0$	2	6.600	.037	3.201	.042
Money and Regime Forecasts					
(A1); $q_1 = 2$, $q_2 = 1$	12	60.390	.000	4.939	.000
(A2); $q_1 = 2$, $q_2 = 1$	1	53.696	.000	51.951	.000
Noncausality in mean					
(B1); $q_1 = 2$, $q_2 = 1$	12	60.390	.000	4.939	.000
(B3); $q_1 = 2$, $q_2 = 1$	3	55.267	.000	17.910	.000
(B3); $q_1 = 1$, $q_2 = 2$	7	13.261	.066	1.850	.076
Noncausality in mean-variance					
(C1); $q_1 = 2$, $q_2 = 1$	12	60.390	.000	4.939	.000
(C3); $q_1 = 2$, $q_2 = 1$	5	71.542	.000	13.910	.000
(C3); $q_1 = 1$, $q_2 = 2$	8	74.204	.000	9.060	.000
Noncausality in distribution					
(D1); $q_1 = 2$, $q_2 = 1$	12	60.390	.000	4.939	.000
(D2); $q_1 = 2$, $q_2 = 1$	5	71.542	.000	13.910	.000
(D2); $q_1 = 1$, $q_2 = 2$	8	74.204	.000	9.060	.000

NOTES: For the Wald statistic (W) the reference distribution is χ^2 with degrees of freedom equal to the number of restrictions, and for the F statistic the distribution is approximated by an F distribution with degrees of freedom given by the number of restrictions and the number of observations, $T = 431$, minus the average number of free parameters per equation.

TABLE 6: ML estimates and standard errors of conditional and unconditional means and covariances for a 2-state MS-VAR(2) model of money and income growth in the US for the sample 1959:1–1995:2 under the linear noncausality in mean restrictions (B3) with $q_1 = 1$ and $q_2 = 2$.

	Unconditional moments		
variable	mean	variance	covariance
Δy	3.85	117.88	
	(1.44)	(17.53)	−3.18
Δm	5.99	34.94	(5.28)
	(.59)	(3.79)	
	Conditional moments		
	Regime 1		
Δy	3.85	303.70	
	(1.44)	(38.44)	−13.04
Δm	5.25	59.74	(18.44)
	(1.09)	(9.56)	
	Regime 2		
Δy	3.85	58.19	
	(1.44)	(8.92)	−.01
Δm	6.23	26.74	(2.81)
	(.66)	(3.65)	

TABLE 7: ML estimates and standard errors for a 2-state MS-VAR(2) model of money and income growth in the US for the sample 1959:1–1995:2 under the linear noncausality in mean restrictions (B3) with $q_1 = 1$ and $q_2 = 2$.

Income equation			*Money equation*		
coefficient	$s_t = 1$	$s_t = 2$	coefficient	$s_t = 1$	$s_t = 2$
δ_{1,s_t}	2.193	2.193	δ_{2,s_t}	4.640	2.514
	(.436)	(.436)		(1.490)	(.446)
$\alpha^{(1)}_{11,s_t}$	.323	.323	$\alpha^{(1)}_{21,s_t}$	.084	−.028
	(.048)	(.048)		(.068)	(.035)
$\alpha^{(1)}_{12,s_t}$	0	0	$\alpha^{(1)}_{22,s_t}$	.249	.475
				(.129)	(.056)
$\alpha^{(2)}_{11,s_t}$	.107	.107	$\alpha^{(2)}_{21,s_t}$	−.044	−.062
	(.046)	(.046)		(.076)	(.033)
$\alpha^{(2)}_{12,s_t}$	0	0	$\alpha^{(2)}_{22,s_t}$	−.162	.188
				(.135)	(.055)
Ω_{11,s_t}	272.670	46.212	Ω_{22,s_t}	55.537	15.012
	(32.750)	(5.007)		(9.288)	(1.463)
Ω_{12,s_t}	−17.791	2.599			
	(15.826)	(1.764)			
		Markov probabilities			
p_{11}	.708		p_{22}	.906	
	(.079)			(.031)	

$\max|\text{eig}(\hat{A})| = .488$ $\quad \ln L(\mathcal{X}_T;\hat{\theta}) = -2860.2$ $\quad \hat{\pi}_1 = .243 \quad (.055)$

Estimated # obs: $\quad s_t = 1 : 107 \quad (60, 154); \quad s_t = 2 : 324 \quad (277, 371)$

TABLE 8: Monte Carlo evidence on tests for Granger noncausality (in mean) in single regime VAR models with 12 (11) lags for the levels (first differences) when data has been generated by a 2-state MS-VAR(2) model for the first differences.

Granger Noncausality

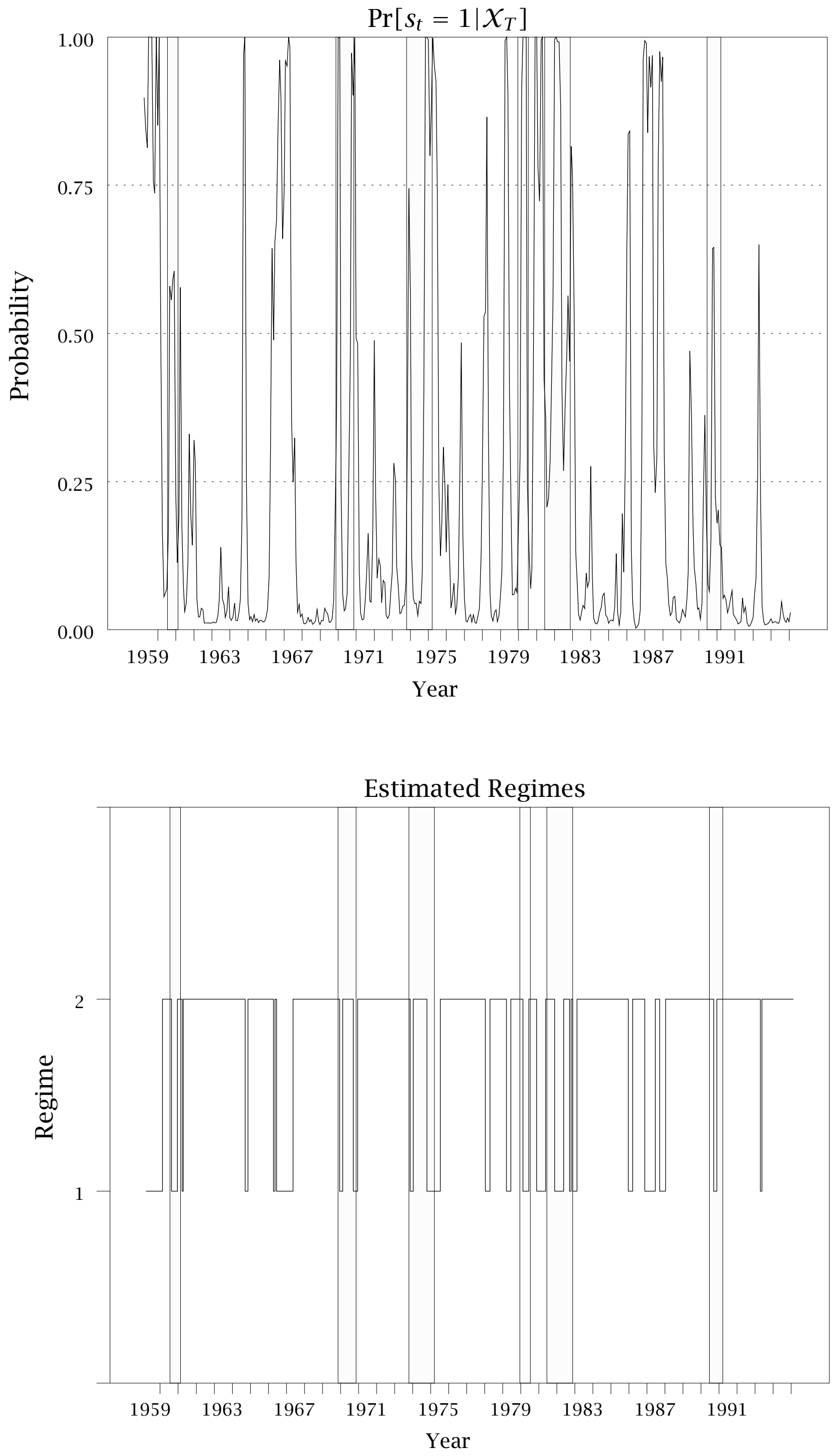
System	F_{obs}	p-value	(B3), $q_1 = 1, q_2 = 2$ 90 %	95 %	99 %
(y, m)	2.034	.116	2.089	2.365	2.919
$(\Delta y, \Delta m)$	1.534	.166	1.735	1.986	2.524
			Unrestricted		
(y, m)	2.034	.723	4.626	5.340	6.703
$(\Delta y, \Delta m)$	1.534	.846	4.698	5.423	6.946

Size & Size Adjusted Power

	size			power		
System	10 %	5 %	1 %	10 %	5 %	1 %
(y, m)	.33	.21	.07	.709	.619	.441
$(\Delta y, \Delta m)$	.14	.08	.02	.789	.708	.538

NOTES: For each bootstrap model 10,000 samples have been generated for first differences and levels. To calculate the size, the critical values from the $F(12, 397)$ and $F(11, 399)$ distributions have been used, i.e. $(1.56, 1.78, 2.23)$ at the 10, 5, and 1 percent level in the former case, and $(1.59, 1.81, 2.29)$ in the latter.

FIGURE 1: Estimated smooth probabilities (top) and regimes (bottom) for a 2-state MS-VAR(2) model of money and income growth in the US for the sample 1959:1–1995:2



Proof of Proposition 1

It is straightforward to show that (A2) implies that there is no information in $\mathcal{X}_{2,t}$ for predicting $s_{1,t+1}$ since it implies that $\Pr[s_{1,t+1}|\mathcal{X}_t] = \Pr[s_{1,t+1}]$. Let us therefore focus on the only remaining possibility, i.e. that $\Pr[s_{1,t}|\mathcal{X}_t] = \Pr[s_{1,t}|\mathcal{X}_{1,t}]$. To prove that condition (A1) is necessary and sufficient for this to hold, we shall proceed in two steps. The first step involves finding a general condition for predictions of $s_{1,t}$ (and $s_{2,t}$) to be invariant with respect to alternative information sets. In the second step we show that when $\varepsilon_t|s_t$ is Gaussian, then the parameter restrictions in (A1) are necessary and sufficient for the invariance condition in the first step to be satisfied under the two information sets of interest.

Let $\xi_{t|\tau}(j) = \Pr[s_t = j|x_\tau, \mathcal{W}_\tau]$, where x_t is a vector of variables and W_τ is the history of an observable vector w_t up to and including period τ. The vector w_t can, for example, be defined such that it contains x_{t-1} and various exogenous variables observable at time t. Furthermore, let $\eta_t(j) = f_{x_j}(x_t|s_t = j, \mathcal{W}_t)$ be the density function for x_t conditional on the state and the history of w. We stack these functions into $q \times 1$ vectors $\xi_{t|\tau}$ and η_t, respectively. From e.g. Hamilton (1994) we have that $\xi_{t|t}$, $\xi_{t|t-1}$, and η_t are related according to:

$$\xi_{t|t} = \frac{(\xi_{t|t-1} \odot \eta_t)}{\imath_q'(\xi_{t|t-1} \odot \eta_t)}, \qquad t = 1, 2, \dots, \tag{A.1}$$

while

$$\xi_{t|t-1} = P'\xi_{t-1|t-1}, \qquad t = 2, 3, \dots, \tag{A.2}$$

and $\xi_{1|0} = \rho$, a $q \times 1$ vector of positive constants summing to unity. Here, $\odot$ denotes the Hadamard (element-by-element) product and $\imath_q$ the $q \times 1$ unit vector.

Let s_t be represented by two Markov processes, $s_{1,t}$ and $s_{2,t}$, which are not necessarily independent. Define j such that $j \equiv j_2 + q_2(j_1 - 1)$ when $(s_{1,t}, s_{2,t}) = (j_1, j_2)$, where $q_1, q_2 \geq 1$ and $q = q_1 q_2 \geq 2$. Then $\xi_{t|\tau}(j) = \xi_{t|\tau}(j_1, j_2) = \Pr[s_{1,t} = j_1, s_{2,t} = j_2|x_\tau, \mathcal{W}_\tau]$, while $\xi_{t|\tau}^{(1)}(j_1) = \sum_{j_2=1}^{q_2} \xi_{t|\tau}(j_1, j_2)$ and similarly for $\xi_{t|\tau}^{(2)}(j_2)$. More compactly, this means that $\xi_{t|\tau}^{(1)} = [I_{q_1} \otimes \imath_{q_2}']\xi_{t|\tau}$ and $\xi_{t|\tau}^{(2)} = [\imath_{q_1}' \otimes I_{q_2}]\xi_{t|\tau}$. The following result about Hadamard and Kronecker products will prove useful below:

LEMMA 1: *If and only if $\eta_t = (\eta_t^{(1)} \otimes \eta_t^{(2)})$ with $\eta_t^{(l)}$ being $q_l \times 1$ for $l = 1, 2$, then*

$$\left(I_{q_1} \otimes \imath_{q_2}'\right)\left(\xi_{t|t-1} \odot \eta_t\right) = \left(\left[I_{q_1} \otimes \eta_t^{(2)\prime}\right]\xi_{t|t-1}\right) \odot \eta_t^{(1)}, \tag{A.3}$$

while

$$\left(\imath_{q_1}' \otimes I_{q_2}\right)\left(\xi_{t|t-1} \odot \eta_t\right) = \left(\left[\eta_t^{(1)\prime} \otimes I_{q_2}\right]\xi_{t|t-1}\right) \odot \eta_t^{(2)}. \tag{A.4}$$

PROOF: The j:th element of $(\xi_{t|t-1} \odot \eta_t)$ is given by $\xi_{t|t-1}(j_1, j_2)\eta_t^{(1)}(j_1)\eta_t^{(2)}(j_2)$. Premultiplying this $q \times 1$ vector by $[I_{q_1} \otimes \imath_{q_2}']$ we obtain a $q_1 \times 1$ vector whose j_1:th element is

$$\eta_t^{(1)}(j_1) \sum_{j_2=1}^{q_2} \xi_{t|t-1}(j_1, j_2)\eta_t^{(2)}(j_2).$$

Now define

$$\gamma_{t|t-1}(j_1) \equiv \begin{bmatrix} \xi_{t|t-1}(j_1, 1) \\ \vdots \\ \xi_{t|t-1}(j_1, q_2) \end{bmatrix}, \qquad j_1 = 1, \dots, q_1. \tag{A.5}$$

Then

$$\gamma_{t|t-1}(j_1)' \eta_t^{(2)} = \sum_{j_2=1}^{q_2} \xi_{t|t-1}(j_1, j_2) \eta_t^{(2)}(j_2).$$

Collecting these results we find that

$$\left[I_{q_1} \otimes \iota_{q_2}'\right]\left[\xi_{t|t-1} \odot \left(\eta_t^{(1)} \otimes \eta_t^{(2)}\right)\right] = \begin{bmatrix} \gamma_{t|t-1}(1)' \eta_t^{(2)} \\ \vdots \\ \gamma_{t|t-1}(q_1)' \eta_t^{(2)} \end{bmatrix} \odot \eta_t^{(1)}. \tag{A.6}$$

Define the $q_2 \times q_1$ matrix $\gamma_{t|t-1}$ according to $\gamma_{t|t-1} \equiv [\gamma_{t|t-1}(1) \;\; \cdots \;\; \gamma_{t|t-1}(q_1)]$. It then follows that

$$\gamma_{t|t-1}' \eta_t^{(2)} = \begin{bmatrix} \gamma_{t|t-1}(1)' \eta_t^{(2)} \\ \vdots \\ \gamma_{t|t-1}(q_1)' \eta_t^{(2)} \end{bmatrix}. \tag{A.7}$$

Moreover, $\xi_{t|t-1} = \mathrm{vec}(\gamma_{t|t-1})$, with vec being the column stacking operator. Next,

$$\begin{aligned} \gamma_{t|t-1}' \eta_t^{(2)} &= \left[\eta_t^{(2)\prime} \otimes I_{q_1}\right] \mathrm{vec}(\gamma_{t|t-1}') \\ &= \left[\eta_t^{(2)\prime} \otimes I_{q_1}\right] K_{q_2,q_1} \mathrm{vec}(\gamma_{t|t-1}) \\ &= K_{q_1,1}\left[I_{q_1} \otimes \eta_t^{(2)\prime}\right] \xi_{t|t-1} \\ &= \left[I_{q_1} \otimes \eta_t^{(2)\prime}\right] \xi_{t|t-1}, \end{aligned} \tag{A.8}$$

where $K_{m,n}$ is the $mn \times mn$ commutation matrix, $K_{m,1} = I_m$, and the third equality follows by Theorem 3.9 in Magnus and Neudecker (1988). Collecting these last results we have established (A.3). The result (A.4) follows by similar arguments. □

If $s_{1,t}$ and $s_{2,t}$ are independent, it follows that

$$\begin{aligned} \xi_{t|t-1}^{(1)} &= \left[I_{q_1} \otimes \iota_{q_2}'\right]\left[P^{(1)\prime} \otimes P^{(2)\prime}\right] \xi_{t-1|t-1} \\ &= P^{(1)\prime} \xi_{t-1|t-1}^{(1)}, \end{aligned} \tag{A.9}$$

since $P^{(2)} \iota_{q_2} = \iota_{q_2}$. Similarly, $\xi_{t|t-1}^{(2)} = P^{(2)\prime} \xi_{t-1|t-1}^{(2)}$. However, this does not mean that $\xi_{t|t-1}^{(1)}$ and $\xi_{t|t-1}^{(2)}$ are independent since $\xi_{t-1|t-1}^{(1)}$ and $\xi_{t-1|t-1}^{(2)}$ need not be independent.

LEMMA 2: *If and only if* (i) $\eta_t = \varphi_t(\eta_t^{(1)} \otimes \eta_t^{(2)})$ *where* φ_t *is a scalar and* $\eta_t^{(l)}$ *a* $q_l \times 1$ *vector,* (ii) $\eta_t^{(1)}$ *and* $\eta_t^{(2)}$ *are vectors of density functions for independent random variables, and* (iii) $s_{1,t}$ *and* $s_{2,t}$ *are* *independent, then for all $t = 1, \dots, T$*

$$\xi_{t|t}^{(l)} = \frac{\left(\xi_{t|t-1}^{(l)} \odot \eta_t^{(l)}\right)}{\iota_{q_l}'\left(\xi_{t|t-1}^{(l)} \odot \eta_t^{(l)}\right)}, \qquad l = 1, 2, \tag{A.10}$$

with $\xi_{t|\tau} = (\xi_{t|\tau}^{(1)} \otimes \xi_{t|\tau}^{(2)})$, where $\xi_{t|\tau}^{(1)}$ and $\xi_{t|\tau}^{(2)}$ are independent for $\tau = t, t-1$.

PROOF: Note first that $\iota_q' = \iota_{q_1}'(I_{q_1} \otimes \iota_{q_2}') = \iota_{q_2}'(\iota_{q_1}' \otimes I_{q_2})$. For $l = 1$ we know that $\xi_{t|t}^{(1)} = [I_{q_1} \otimes \iota_{q_2}']\xi_{t|t}$. From equation (A.1) we thus have that

$$\begin{aligned}\xi_{t|t}^{(1)} &= \left[I_{q_1} \otimes \iota_{q_2}'\right]\left[\xi_{t|t-1} \odot \eta_t\right]\left[\iota_{q_1}'\left(I_{q_1} \otimes \iota_{q_2}'\right)\left(\xi_{t|t-1} \odot \eta_t\right)\right]^{-1} \\ &= \left[\left(\left[I_{q_1} \otimes \eta_t^{(2)\prime}\right]\xi_{t|t-1}\right) \odot \eta_t^{(1)}\right]\left[\iota_{q_1}'\left(\left[\left(I_{q_1} \otimes \eta_t^{(2)\prime}\right)\xi_{t|t-1}\right] \odot \eta_t^{(1)}\right)\right]^{-1},\end{aligned} \tag{A.11}$$

by Lemma 1 and since the scalar φ_t cancels. A similar expression is obtained for $\xi_{t|t}^{(2)}$. Let $\rho = (\rho^{(1)} \otimes \rho^{(2)})$ where the elements of $\rho^{(l)}$ are positive and sum to unity. Then

$$\begin{aligned}\xi_{1|1}^{(1)} &= \left[\left(\rho^{(1)} \otimes \eta_1^{(2)\prime}\rho^{(2)}\right) \odot \eta_1^{(1)}\right]\left[\iota_{q_1}'\left(\left[\rho^{(1)} \otimes \eta_1^{(2)\prime}\rho^{(2)}\right] \odot \eta_1^{(1)}\right)\right]^{-1} \\ &= \left[\rho^{(1)} \odot \eta_1^{(1)}\right]\left[\iota_{q_1}'\left(\rho^{(1)} \odot \eta_1^{(1)}\right)\right]^{-1},\end{aligned} \tag{A.12}$$

and similarly for $\xi_{1|1}^{(2)}$. By (ii) it follows that $\xi_{1|1}^{(1)}$ and $\xi_{1|1}^{(2)}$ are independent. Thus, $\xi_{1|1} = (\xi_{1|1}^{(1)} \otimes \xi_{1|1}^{(2)})$. Moreover, by (iii) we have that $\xi_{2|1}^{(l)} = P^{(l)\prime}\xi_{1|1}^{(l)}$, which are also independent for $l = 1, 2$. Thus, $\xi_{2|1} = (\xi_{2|1}^{(1)} \otimes \xi_{2|1}^{(2)})$ and so on for $t = 2, 3, \dots, T$, thereby establishing sufficiency.

To prove necessity, suppose (i) is not true. Let $\eta_t = (\eta_t^{(1)} \otimes \eta_t^{(2)}) \odot \psi_t$, where $\psi_t \neq (\psi_t^{(1)} \otimes \psi_t^{(2)})$ for $q_l \times 1$ vectors $\psi_t^{(l)}$. Then, for example

$$\begin{aligned}\xi_{t|t}^{(1)} &= \left[\left(I_{q_1} \otimes \eta_t^{(2)\prime}\right)\left(\xi_{t|t-1} \odot \psi_t\right) \odot \eta_t^{(1)}\right]\left[\iota_{q_1}'\left(\left[I_{q_1} \otimes \eta_t^{(2)\prime}\right]\left[\xi_{t|t-1} \odot \psi_t\right] \odot \eta_t^{(1)}\right)\right]^{-1} \\ &\neq \left[\left(\left[I_{q_1} \otimes \eta_t^{(2)\prime}\right]\xi_{t|t-1}\right) \odot \eta_t^{(1)}\right]\left[\iota_{q_1}'\left(\left[\left(I_{q_1} \otimes \eta_t^{(2)\prime}\right)\xi_{t|t-1}\right] \odot \eta_t^{(1)}\right)\right]^{-1}.\end{aligned} \tag{A.13}$$

The only case when the inequality can be replaced with an equality is if $\psi_t = (\psi_t^{(1)} \otimes \psi_t^{(2)})$. Next, if (ii) does not hold, then for instance $\xi_{1|1}^{(1)}$ and $\xi_{1|1}^{(2)}$ cannot be independent. Finally, if (iii) does not hold, then $\xi_{t|t-1}^{(1)} \neq P^{(1)\prime}\xi_{t-1|t-1}^{(1)}$ and depends on $\xi_{t-1|t-1}^{(2)}$ as well. Thus, $\xi_{2|1}^{(1)}$ and $\xi_{2|1}^{(2)}$ cannot be independent even if $\xi_{1|1}^{(1)}$ and $\xi_{1|1}^{(2)}$ are. □

Note that assumptions (i) and (ii) are often closely related. For the Gaussian distribution, for example, (i) implies (ii) and vice versa. However, there may exist some perverse distributions which can satisfy (i) but not (ii) unless additional parametric conditions hold.

We can always select $\rho = (\rho^{(1)} \otimes \rho^{(2)})$ when the parameters are assumed to be known. However, Hamilton (1990) shows that when parameters are unknown, the ML estimator of ρ is given by the estimate of $\xi_{1|T}$. The following Lemma ensures that the results in Lemma 2 also hold when the ML estimator of θ is consistent; see Krishnamurthy and Rydén (1998).

LEMMA 3: *If and only if the conditions in Lemma 2 are satisfied, then*

$$\xi_{t|\tau} = \left(\xi_{t|\tau}^{(1)} \otimes \xi_{t|\tau}^{(2)}\right), \tag{A.14}$$

for all $t, \tau = 1, \ldots, T$, *with* $\xi_{t|\tau}^{(1)}$ *and* $\xi_{t|\tau}^{(2)}$ *being independent.*

PROOF: Let us first prove this for all $\tau < t$. We have already established in Lemma 2 that $\xi_{\tau|\tau}^{(1)}$ and $\xi_{\tau|\tau}^{(2)}$ are independent for all τ. By equation (22.3.13) in Hamilton (1994) we have that $\xi_{t|\tau} = (P')^{t-\tau}\xi_{\tau|\tau}$ for $\tau = 1, \ldots, t-1$. By independence of $s_{1,t}$ and $s_{2,t}$ and of $\xi_{\tau|\tau}^{(1)}$ and $\xi_{\tau|\tau}^{(2)}$ we obtain $\xi_{t|\tau} = [(P^{(1)\prime})^{t-\tau}\xi_{\tau|\tau}^{(1)} \otimes (P^{(2)\prime})^{t-\tau}\xi_{\tau|\tau}^{(2)}] = (\xi_{t|\tau}^{(1)} \otimes \xi_{t|\tau}^{(2)})$, which are thus independent.

To show (A.14) for $\tau > t$ it is sufficient to consider $\tau = T$ since the algorithm for computing smooth probabilities is valid for any $\tau > t$. From Kim (1994) (see also Lindgren, 1978; Hamilton, 1994) we get

$$\xi_{t|T} = \xi_{t|t} \odot \left[P\left(\xi_{t+1|T} \ominus \xi_{t+1|t}\right)\right], \qquad t = 1, \ldots, T-1, \tag{A.15}$$

where $\ominus$ denotes element-by-element division. To show that $\xi_{t|T} = (\xi_{t|T}^{(1)} \otimes \xi_{t|T}^{(2)})$, with $\xi_{t|T}^{(l)}$ independent for $l = 1, 2$, we begin with $t = T-1$. By Lemma 2 we have that $\xi_{T|\tau} = (\xi_{T|\tau}^{(1)} \otimes \xi_{T|\tau}^{(2)})$ for $\tau = T, T-1$. Accordingly,

$$\left[\xi_{T|T} \ominus \xi_{T|T-1}\right] = \left[\left(\xi_{T|T}^{(1)} \ominus \xi_{T|T-1}^{(1)}\right) \otimes \left(\xi_{T|T}^{(2)} \ominus \xi_{T|T-1}^{(2)}\right)\right]. \tag{A.16}$$

Let $\psi_T^{(l)} \equiv P^{(l)}(\xi_{T|T}^{(l)} \ominus \xi_{T|T-1}^{(l)})$ for $l = 1, 2$. We then obtain

$$P\left[\xi_{T|T} \ominus \xi_{T|T-1}\right] = \left[\psi_T^{(1)} \otimes \psi_T^{(2)}\right] \equiv \psi_T. \tag{A.17}$$

Hence, $\xi_{T-1|T} = (\xi_{T-1|T-1} \odot \psi_T)$. With $\xi_{t|T}^{(1)} = [I_{q_1} \otimes \iota_{q_2}']\xi_{t|T}$ it follows by Lemma 1 and Lemma 2 that

$$\begin{aligned} \xi_{T-1|T}^{(1)} &= \left[\left(I_{q_1} \otimes \psi_T^{(2)\prime}\right)\xi_{T-1|T-1}\right] \odot \psi_T^{(1)} \\ &= \psi_T^{(2)\prime}\xi_{T-1|T-1}^{(2)}\left(\xi_{T-1|T-1}^{(1)} \odot \psi_T^{(1)}\right), \end{aligned} \tag{A.18}$$

since $\xi_{T-1|T-1} = (\xi_{T-1|T-1}^{(1)} \otimes \xi_{T-1|T-1}^{(2)})$. From the definition of $\psi_T^{(2)}$ we find that

$$\begin{aligned} \psi_T^{(2)\prime}\xi_{T-1|T-1}^{(2)} &= \left(\xi_{T|T}^{(2)} \ominus \xi_{T|T-1}^{(2)}\right)' P^{(2)\prime}\xi_{T-1|T-1}^{(2)} \\ &= \left(\xi_{T|T}^{(2)} \ominus \xi_{T|T-1}^{(2)}\right)' \xi_{T|T-1}^{(2)} \\ &= \sum_{j_2=1}^{q_2} \xi_{T|T}^{(2)}(j_2). \end{aligned} \tag{A.19}$$

This is equal to unity and we thus get

$$\xi_{T-1|T}^{(1)} = \xi_{T-1|T-1}^{(1)} \odot \left[P^{(1)}\left(\xi_{T|T}^{(1)} \ominus \xi_{T|T-1}^{(1)}\right)\right]. \tag{A.20}$$

Proceeding with $\xi_{T-1|T}^{(2)}$, the above arguments imply that

$$\xi_{T-1|T}^{(2)} = \xi_{T-1|T-1}^{(2)} \odot \left[P^{(2)}\left(\xi_{T|T}^{(2)} \ominus \xi_{T|T-1}^{(2)}\right)\right], \tag{A.21}$$

and, hence, by Lemma 2, $\xi^{(l)}_{T-1|T}$ are independent for $l = 1, 2$ and $\xi_{T-1|T} = (\xi^{(1)}_{T-1|T} \otimes \xi^{(2)}_{T-1|T})$. For the remaining t, backwards recursions, using the above arguments, implies the result. Necessity follows by the arguments in Lemma 2. □

Note that conditions (i) and (ii) are only sufficient in forecast situations. If s_t is serially uncorrelated, then $P' = \pi\imath_q'$, with π being the vector of ergodic probabilities. Accordingly, for all $\tau < t$, $\xi_{t|\tau} = (P')^{t-\tau}\xi_{\tau|\tau} = \pi$ since $\imath_q'\pi = \imath_q'\xi_{\tau|\tau} = 1$. Hence, if $s_{1,t}$ and $s_{2,t}$ are independent and serially uncorrelated, then $\xi_{t|\tau} = (\xi^{(1)}_{t|\tau} \otimes \xi^{(2)}_{t|\tau}) = (\pi^{(1)} \otimes \pi^{(2)})$ for all $\tau < t$.

This completes step one in the proof of Proposition 1. We have established necessary and sufficient conditions for how the information used to predict s_t can be split into information valuable for predicting $s_{1,t}$ but not $s_{2,t}$, and vice versa, and when information can be "thrown away" without affecting the regime predictions. Note that the conditions in Lemma 2 are very general in the sense that they apply to any vector of density functions η_t. For example, the functional form can vary over t as well as over states. The crucial underlying assumption is that s_t is independent of information available at time $t-1$ conditional on s_{t-1}. If this assumption is violated, then the algorithms for computing regime predictions are no longer valid.

The assumption that $s_{1,t}$ and $s_{2,t}$ are independent, in fact, increases the level of generality of the results. For example, it allows $q_2 = 1$ in which case $\eta_t = \varphi_t\eta^{(1)}_t$ (with the scalar φ_t being invariant with respect to s_t) is necessary and sufficient for regime predictions based on the vector densities η_t and $\eta^{(1)}_t$ to be equivalent. The scalar φ_t can, for instance, be a marginal density.

When $q_1, q_2 \geq 2$ we allow for the possibility that two subsystems of the model can contain information for predicting one independent regime process each but not the other regime process, while a third subsystem is completely noninformative about regimes. By considering r independent Markov chains, these results can be generalized further. For my purposes, however, the above results are sufficient.

Now let us return to the MS-VAR with conditionally Gaussian residuals. Here we find that for each $j \in \{1, \ldots, q\}$ the joint log density is

$$\ln(\eta_t(j)) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\Bigl(\det[\Omega_j]\Bigr) - \frac{1}{2}\varepsilon_{t|j}'\Omega_j^{-1}\varepsilon_{t|j}, \tag{A.22}$$

where $\varepsilon_{t|j} = x_t - \mu_j D_t - \sum_{k=1}^{p} A_j^{(k)} x_{t-k}$. Let n_1 and n_2 be the number of $x_{1,t}$ and $x_{2,t}$ variables, respectively, with $n_1 + n_2 = n$. The marginal density for $x_{2,t}$, conditional on $s_t = j$ and $\mathcal{X}_{t-1}$, is

$$\ln\Bigl(\eta^{(2)}_t(j)\Bigr) = -\frac{n_2}{2}\ln(2\pi) - \frac{1}{2}\ln\Bigl(\det[\Omega_{22,j}]\Bigr) - \frac{1}{2}\epsilon_{2,t|j}'\Omega_{22,j}^{-1}\epsilon_{2,t|j}. \tag{A.23}$$

If this density is invariant with respect to $s_{1,t}$, then (i) $\Omega_{22,(j_1,j_2)} = \Omega_{22,j_2}$, $\delta_{2,(j_1,j_2)} = \delta_{2,j_2}$, and $\alpha^{(k)}_{2r,(j_1,j_2)} = \alpha^{(k)}_{2r,j_2}$ for all $j_1 \in \{1, \ldots, q_1\}$, $j_2 \in \{1, \ldots, q_2\}$, $r \in \{1, 2\}$, and $k \in \{1, \ldots, p\}$. For $q_2 = 1$ these restrictions imply that the parameters in the marginal density for $x_{2,t}$ are constant across states.

Under these restrictions, the density for $x_{1,t}$, conditional on $s_t = j = j_2 + q_2(j_1 - 1)$, $x_{2,t}$, and $\mathcal{X}_{t-1}$, is

$$\begin{aligned}\ln\Big(\eta_t^{(1)}(j)\Big) = &-\frac{n_1}{2}\ln(2\pi) - \frac{1}{2}\ln\Big(\det[\tilde{\Omega}_{11,j}]\Big) + \epsilon'_{2,t|j_2}\Omega^{-1}_{22,j_2}\Omega'_{12,j}\tilde{\Omega}^{-1}_{11,j}\epsilon_{1,t|j}\\ &-\frac{1}{2}\epsilon'_{1,t|j}\tilde{\Omega}^{-1}_{11,j}\epsilon_{1,t|j} - \frac{1}{2}\epsilon'_{2,t|j_2}\Omega^{-1}_{22,j_2}\Omega'_{12,j}\tilde{\Omega}^{-1}_{11,j}\Omega_{12,j}\Omega^{-1}_{22,j_2}\epsilon_{2,t|j_2},\end{aligned} \tag{A.24}$$

where $\tilde{\Omega}_{11,j} \equiv \Omega_{11,j} - \Omega_{12,j}\Omega^{-1}_{22,j_2}\Omega'_{12,j}$. If this density function is invariant with respect to $s_{2,t}$ for $q_2 \geq 2$, then (ii) $\Omega_{11,(j_1,j_2)} = \Omega_{11,j_1}$, $\delta_{1,(j_1,j_2)} = \delta_{1,j_1}$, and $\alpha^{(k)}_{1r,(j_1,j_2)} = \alpha^{(k)}_{1r,j_1}$ for all $j_1 \in \{1,\ldots,q_1\}$, $j_2 \in \{1,\ldots,q_2\}$, $r \in \{1,2\}$, and $k \in \{1,\ldots,p\}$; and (iii) $\Omega_{12,j} = 0$ for all $j \in \{1,\ldots,q\}$. Under (i) to (iii) we find that $\eta_t = (\eta_t^{(1)} \otimes \eta_t^{(2)})$ for all t, with $\eta_t^{(l)}$ being the marginal density of $x_{l,t}$ conditional on $s_{l,t}$ and $\mathcal{X}_{t-1}$. If these linear restrictions are not satisfied, then η_t cannot be decomposed into the (Kronecker) product between a q_1 and a q_2 vector density. For $q_2 = 1$, restrictions (iii) can be dispensed with. In that case, $\eta_t = \varphi_t\eta_t^{(1)}$, with φ_t being given by the marginal density for $x_{2,t}$.

To satisfy the remaining two conditions in Lemma 2 we only need to let $s_{1,t}$ and $s_{2,t}$ be independent. For $q_2 \geq 2$ we have that $\eta_t^{(1)}$ and $\eta_t^{(2)}$ are vectors of densities for independent random variables ($\epsilon_{1,t}|s_{1,t}$ and $\epsilon_{2,t}|s_{2,t}$) from, in particular, restrictions (iii), and for $q_2 = 1$ this is not needed since φ_t is just a scalar which cancels in (A.1). By Lemma 2 it then follows that

$$\Pr[s_t = j|\mathcal{X}_t;\theta^*] = \Pr[s_{1,t} = j_1\,|\,\mathcal{X}_{1,t},\mathcal{X}_{2,t};\theta_1^*]\Pr[s_{2,t} = j_2\,|\,\mathcal{X}_{1,t-1},\mathcal{X}_{2,t};\theta_2^*].$$

When $q_2 \geq 2$ it also follows that $\Pr[s_{1,t} = j_1|\mathcal{X}_{1,t},\mathcal{X}_{2,t};\theta_1^*] = \Pr[s_{1,t} = j_1|\mathcal{X}_{1,t},\mathcal{X}_{2,t-1};\theta_1^*]$.

The final stage is now straightforward. Since $\mathcal{X}_{2,t}$ is assumed to be noninformative about $s_{1,t}$, (iii) must also hold for $q_2 = 1$, and (iv) $\alpha^{(k)}_{12,j_1} = 0$ for all $j_1 \in \{1,\ldots,q_1\}$ and $k \in \{1,\ldots,p\}$ for $q_2 \geq 1$. Hence, we have shown that

$$\Pr\big[(s_{1,t},s_{2,t}) = (j_1,j_2)\,|\,\mathcal{X}_t;\theta^*\big] = \Pr[s_{1,t} = j_1\,|\,\mathcal{X}_{1,t};\theta_1^*]\Pr[s_{2,t} = j_2\,|\,\mathcal{X}_t;\theta_2^*],$$

implies that (A1) is satisfied. To prove the reverse is straightforward. Q.E.D.

Proof of Proposition 2

Given that u_{t+1} is mean zero stationary we know that $E[u^2_{t+1}] \leq E[\tilde{u}^2_{t+1}]$ since $(\mathcal{Y}_t,\mathcal{Z}_t) \subset \mathcal{X}_t$ for all t. In particular,

$$E\Big[\tilde{u}^2_{t+1}\Big] = E\Big[u^2_{t+1}\Big] + E\Big[\big(E[y_{t+1}\,|\,\mathcal{X}_t] - E[y_{t+1}\,|\,\mathcal{Y}_t,\mathcal{Z}_t]\big)^2\Big]. \tag{A.25}$$

Accordingly, the variances of u_{t+1} and $\tilde{u}_{t+1}$ are equal if and only if $E[y_{t+1}|\mathcal{X}_t] = E[y_{t+1}|\mathcal{Y}_t,\mathcal{Z}_t]$ for all t.

The prediction of y_{t+1} conditional on $\mathcal{X}_t$ is given by

$$\begin{aligned}E\Big[y_{t+1}|\mathcal{X}_t\Big] = \bar{\mu}_{1,t}D_{t+1} + \sum_{k=1}^{p}\Big(&\bar{a}^{(k)}_{11,t}y_{t+1-k} + \bar{a}^{(k)}_{12,t}z_{1,t+1-k}\\ &+ \bar{a}^{(k)}_{13,t}m_{t+1-k} + \bar{a}^{(k)}_{14,t}z_{2,t+1-k}\Big).\end{aligned} \tag{A.26}$$

The necessary and sufficient conditions for this expression to be invariant with respect to $\mathcal{M}_t$ are, for all t, given by

$$
\begin{aligned}
&\text{(i)} && \bar{\mu}_{1,t} = E\Big[\mu_{1,s_{t+1}} \Big| \mathcal{Y}_t, \mathcal{Z}_t\Big], && \\
&\text{(ii)} && \bar{a}_{1r,t}^{(k)} = E\Big[a_{1r,s_{t+1}}^{(k)} \Big| \mathcal{Y}_t, \mathcal{Z}_t\Big], && r \in \{1, \ldots, 4\} \text{ and } k \in \{1, \ldots, p\}, \\
&\text{(iii)} && \bar{a}_{13,t}^{(k)} = 0, && k \in \{1, \ldots, p\}.
\end{aligned}
$$

To prove the claim in Proposition 2 we therefore have to show that (i)–(iii) are equivalent to [(B1) or (B2)].

NONCAUSALITY IN MEAN $\Rightarrow$ [(B1) OR (B2)]

From the definitions of $\bar{\mu}_{1,t}$ and $\bar{a}_{1r,t}^{(k)}$ in equations (14) and (15) we find that these random matrices can be expressed as

$$
\bar{\mu}_{1,t} = \sum_{i=1}^{q} \sum_{j=1}^{q} \mu_{1,j} p_{ij} \Pr[s_t = i \,|\, \mathcal{X}_t], \tag{A.27}
$$

and

$$
\bar{a}_{1r,t}^{(k)} = \sum_{i=1}^{q} \sum_{j=1}^{q} a_{1r,j}^{(k)} p_{ij} \Pr[s_t = i \,|\, \mathcal{X}_t]. \tag{A.28}
$$

From these two equations it can be seen that $\bar{\mu}_{1,t}$ and $\bar{a}_{1r,t}^{(k)}$ depend on t, and thus potentially on $\mathcal{M}_t$, only via the filter probabilities $\Pr[s_t = i|\mathcal{X}_t]$.

Suppose first that $(\bar{\mu}_{1,t}, \bar{a}_{1r,t}^{(k)})$ indeed varies with t. It now follows that noncausality in mean implies that

$$
\Pr[(s_{1,t}, s_{2,t}) = (i_1, i_2) \,|\, \mathcal{X}_t] = \Pr[s_{1,t} = i_1 \,|\, \mathcal{X}_{1,t}] \Pr[s_{2,t} = i_2 \,|\, \mathcal{X}_t], \tag{A.29}
$$

must hold for all i_1, i_2, and t, while $(\mu_{1,(j_1,j_2)}, a_{1r,(j_1 j_2)}^{(k)})$ only depends on j_2. By Corollary 2 we know that equation (A.29) can only be satisfied under (A1) and, thus, under (B1). The remaining parameter restrictions, $p_{ij} = p_{i_1 j_1}^{(1)} p_{i_2 j_2}^{(2)}$, are also satisfied under (B1).

Notice that the formulation in (A.29) covers the case when $n_2 = 1$, i.e. $\mathcal{X}_{1,t} = (\mathcal{Y}_t, \mathcal{Z}_t)$, as well as the cases when $n_2 \geq 2$. It is therefore more general than one where $\Pr[s_{1,t} = i_1|\mathcal{X}_{1,t}]$ is replaced with $\Pr[s_{1,t} = i_1|\mathcal{Y}_t, \mathcal{Z}_t]$.

It remains to examine the case when $(\bar{\mu}_{1,t}, \bar{a}_{1r,t}^{(k)})$ is invariant with respect to t. From equations (A.27)–(A.28) we now have that $\sum_{j=1}^{q} \mu_{1,j} p_{ij} = \bar{\mu}_1$, $\sum_{j=1}^{q} a_{1r,j}^{(k)} p_{ij} = \bar{a}_{1r}^{(k)}$, with $\bar{a}_{13}^{(k)} = 0$ for all i, r, and k. Hence, condition (B2) is satisfied.

[(B1) OR (B2)] $\Rightarrow$ NONCAUSALITY IN MEAN

Evaluating equation (A.26) under (B1) and (B2), respectively, gives the result. Q.E.D.

Proof of Proposition 3

Let $B_{s_t}^{(k)} \equiv [a_{11,s_t}^{(k)}\ a_{12,s_t}^{(k)}\ a_{14,s_t}^{(k)}]$ for all k and s_t, $w_t \equiv [y_t\ z_t']'$, while $\bar{B}_t^{(k)} \equiv E[B_{s_{t+1}}^{(k)} | \mathcal{X}_t]$ for all t and k. Since v_{t+1} and $\varepsilon_{1,t+1}$ are uncorrelated conditional on $\mathcal{X}_t$, it follows that the conditional variance of u_{t+1} is

$$E[u_{t+1}^2 \,|\, \mathcal{X}_t] = \sigma_{v,t}^2 + \sum_{j=1}^{q} \omega_{11,j} \Pr[s_{t+1} = j \,|\, \mathcal{X}_t], \tag{A.30}$$

where

$$\begin{aligned}
\sigma_{v,t}^2 \equiv & \sum_{j=1}^{q} \left(\mu_{1,j} - \bar{\mu}_{1,t}\right) D_{t+1} D_{t+1}' \left(\mu_{1,j} - \bar{\mu}_{1,t}\right)' \Pr[s_{t+1} = j \,|\, \mathcal{X}_t] \\
& + \sum_{j=1}^{q} \sum_{k=1}^{p} \sum_{l=1}^{p} \left(B_j^{(k)} - \bar{B}_j^{(k)}\right) w_{t+1-k} w_{t+1-l}' \left(B_j^{(l)} - \bar{B}_j^{(l)}\right)' \Pr[s_{t+1} = j \,|\, \mathcal{X}_t] \\
& + \sum_{j=1}^{q} \sum_{k=1}^{p} \sum_{l=1}^{p} \left(a_{13,j}^{(k)} - \bar{a}_{13,t}^{(k)}\right) \left(a_{13,j}^{(l)} - \bar{a}_{13,t}^{(l)}\right) m_{t+1-k} m_{t+1-l} \Pr[s_{t+1} = j \,|\, \mathcal{X}_t] \\
& + 2 \sum_{j=1}^{q} \sum_{k=1}^{p} \left(\mu_{1,j} - \bar{\mu}_{1,t}\right) D_{t+1} w_{t+1-k}' \left(B_j^{(k)} - \bar{B}_j^{(k)}\right)' \Pr[s_{t+1} = j \,|\, \mathcal{X}_t] \\
& + 2 \sum_{j=1}^{q} \sum_{k=1}^{p} \left(a_{13,j}^{(k)} - \bar{a}_{13,t}^{(k)}\right) \left(\mu_{1,j} - \bar{\mu}_{1,t}\right) D_{t+1} m_{t+1-k} \Pr[s_{t+1} = j \,|\, \mathcal{X}_t] \\
& + 2 \sum_{j=1}^{q} \sum_{k=1}^{p} \sum_{l=1}^{p} \left(a_{13,j}^{(k)} - \bar{a}_{13,t}^{(k)}\right) \left(B_j^{(l)} - \bar{B}_j^{(l)}\right) w_{t+1-l} m_{t+1-k} \Pr[s_{t+1} = j \,|\, \mathcal{X}_t].
\end{aligned} \tag{A.31}$$

NONCAUSALITY IN MEAN-VARIANCE $\Rightarrow$ [(C1) OR (C2)]

From Definition 1 and Definition 2 we know that noncausality in mean-variance implies noncausality in mean when u_{t+1} is stationary, i.e. that either (B1) or (B2) is satisfied.

If (B1) holds, then by construction (C1) is satisfied. On the other hand, if (B2) holds, then $\bar{\mu}_{1,t} = \bar{\mu}_1$, $\bar{B}_t^{(k)} = \bar{B}^{(k)}$, and $\bar{a}_{13,t}^{(k)} = 0$ for all t and k. Evaluating equation (A.30) under these restrictions, we find that the third term in equation (A.31) is

$$\sum_{i=1}^{q} \sum_{k=1}^{p} \sum_{l=1}^{p} \left(\sum_{j=1}^{q} a_{13,j}^{(k)} a_{13,j}^{(l)} p_{ij} \right) m_{t+1-k} m_{t+1-l} \Pr[s_t = i \,|\, \mathcal{X}_t].$$

Under noncausality in mean-variance, this term is invariant with respect to $\mathcal{M}_t$. For each triple (i, k, l) with $k = l$, this means that

$$\sum_{j=1}^{q} \left(a_{13,j}^{(k)}\right)^2 p_{ij} = 0. \tag{A.32}$$

Since $(a_{13,j}^{(k)})^2 \geq 0$ and $p_{ij} \geq 0$, with strict inequality for some j for each i, the restrictions in (A.32) can only be satisfied when $a_{13,j}^{(k)} = 0$ for all j and k.

Let us now turn to the first term in equation (A.31). Under (B2), it can be rewritten as

$$\sum_{i=1}^{q}\left(\sum_{j=1}^{q}\left[(\mu_{1,j}-\bar{\mu}_1)\otimes(\mu_{1,j}-\bar{\mu}_1)\right]p_{ij}\right)(D_{t+1}\otimes D_{t+1})\Pr[s_t=i\,|\,\mathcal{X}_t].$$

Noncausality in mean-variance now implies that either

$$\sum_{j=1}^{q}\left[(\mu_{1,j}-\bar{\mu}_1)\otimes(\mu_{1,j}-\bar{\mu}_1)\right]p_{ij}=\sigma_{\mu}, \tag{A.33}$$

or

$$\sum_{j_1=1}^{q_1}\sum_{j_2=1}^{q_2}\left[(\mu_{1,(j_1,j_2)}-\bar{\mu}_1)\otimes(\mu_{1,(j_1,j_2)}-\bar{\mu}_1)\right]p_{i_1j_1}^{(1)}p_{i_2j_2}^{(2)}=\sigma_{\mu,i_1},$$

and equation (A.29) is satisfied. The latter case means that condition (C1) is satisfied. Let us therefore continue with the case when equation (A.33) holds.

Evaluating the remaining terms in equation (A.31) and the second term on the right hand side of (A.30) in the same manner gives us that noncausality in mean-variance implies that either (C1) or an additional set of restrictions from (C2) is satisfied. Once the last term has been examined, the conclusion follows.

[(C1) OR (C2)] ⇒ NONCAUSALITY IN MEAN-VARIANCE

Evaluating equation (A.30) under (C1) and (C2), respectively, gives the result. Q.E.D.

Proof of Proposition 4

The density function for u_{t+1} conditional on $\mathcal{X}_t$ (and θ^*) can be expressed as

$$g_{t+1}(u_{t+1}\,|\,\mathcal{X}_t)=\sum_{j=1}^{q}f(u_{t+1}\,|\,s_{t+1}=j,\mathcal{X}_t)\Pr[s_{t+1}=j\,|\,\mathcal{X}_t]. \tag{A.34}$$

With $u_{t+1}=\nu_{t+1}+\varepsilon_{1,t+1}$, and $\nu_{t+1|j}\equiv(\nu_{t+1}|s_{t+1}=j)$ being given by ν_{t+1} in equation (16) evaluated at $s_{t+1}=j$, we have that

$$u_{t+1}\,|\,(s_{t+1}=j,\mathcal{X}_t)\sim N(\nu_{t+1|j},\omega_{11,j}). \tag{A.35}$$

for each $j\in\{1,\ldots,q\}$.

NONCAUSALITY IN DISTRIBUTION ⇒ [(D1) OR (D2)]

Since m is assumed to be noncausal in distribution for y, we know that the density function for u_{t+1} conditional on $\mathcal{X}_t$ is invariant with respect to $\mathcal{M}_t$. This means that the density function $f(u_{t+1}|s_{t+1}=j,\mathcal{X}_t)$ is also invariant with respect to $\mathcal{M}_t$. Moreover, this density function does not depend on those values of s_{t+1} for which $\mathcal{M}_t$ provides unique forecasting information. That is, noncausality in distribution is equivalent to

$$
\begin{aligned}
g_{t+1}(u_{t+1} \mid \mathcal{X}_t) &= \sum_{j_1=1}^{q_1} \sum_{j_2=1}^{q_2} f(u_{t+1} \mid s_{1,t+1} = j_1, \mathcal{Y}_t, \mathcal{Z}_t) \Pr[s_{1,t+1} = j_1 \mid \mathcal{Y}_t, \mathcal{Z}_t] \\
&\quad \times \Pr[s_{2,t+1} = j_2 \mid \mathcal{X}_t] \\
&= \sum_{j_1=1}^{q_1} f(u_{t+1} \mid s_{1,t+1} = j_1, \mathcal{Y}_t, \mathcal{Z}_t) \Pr[s_{1,t+1} = j_1 \mid \mathcal{Y}_t, \mathcal{Z}_t] \\
&= h_{t+1}(\tilde{u}_{t+1} \mid \mathcal{Y}_t, \mathcal{Z}_t).
\end{aligned} \tag{A.36}
$$

From Proposition 1 we know that either (A1) or (A2) must hold for the regime forecasts of $s_{1,t+1}$ and $s_{2,t+1}$ to be independent and for $\mathcal{M}_t$ to be noninformative about $s_{1,t+1}$ once we have conditioned on $(\mathcal{Y}_t, \mathcal{Z}_t)$. If (A1) holds, then condition (D1) is by construction satisfied. On the other hand, if (A2) holds, then $\Pr[s_{1,t+1} = j_1 | \mathcal{X}_t] = \pi_{j_1}^{(1)}$ for all j_1 and t. The restrictions on the density function $f(\cdot)$ in equation (A.36) for the Gaussian case imply that $\mu_{1,j} = \mu_{1,j_1}$, $a_{1r,j}^{(k)} = a_{1r,j_1}^{(k)}$, $a_{13,j}^{(k)} = 0$, and $\omega_{11,j} = \omega_{11,j_1}$ for all $j \in \{1, \dots, q\}$, $r \in \{1, 2, 4\}$, and $k \in \{1, \dots, p\}$. Hence, condition (D2) is satisfied.

[(D1) OR (D2)] $\Rightarrow$ NONCAUSALITY IN DISTRIBUTION

Evaluating equation (A.34) under (D1) and (D2), respectively, gives the result. Q.E.D.

References

Bickel, P. J., and Ritov, Y. (1996). "Inference in Hidden Markov Models I: Local Asymptotic Normality in the Stationary Case." *Bernoulli*, *2*, 199–228.

Bickel, P. J., Ritov, Y., and Rydén, T. (1998). "Asymptotic Normality of the Maximum Likelihood Estimator for General Hidden Markov Models." *Annals of Statistics*, *26*, 1614–1635.

Blix, M. (1997). "Rational Expectations in a VAR with Markov Switching." In *Rational Expectations and Regime Shifts in Macroeconometrics.* PhD Thesis, Monograph Series No. 31, Institute for International Economic Studies, Stockholm University, Sweden.

Cecchetti, S. G., Lam, R.-S., and Mark, N. C. (1990). "Mean Reversion in Equilibrium Asset Prices." *American Economic Review*, *80*, 398–418.

Chamberlain, G. (1982). "The General Equivalence of Granger and Sims Causality." *Econometrica*, *50*, 569–581.

Christiano, L. J., and Ljungqvist, L. (1988). "Money does Granger-Cause Output in the Bivariate Money-Output Relation." *Journal of Monetary Economics*, *22*, 217–235.

Diebold, F. X., and Rudebusch, G. D. (1996). "Measuring Business Cycles: A Modern Perspective." *Review of Economics and Statistics*, *78*, 67–77.

Engle, R. F., Hendry, D. F., and Richard, J. F. (1983). "Exogeneity." *Econometrica*, *51*, 277–304.

Ferson, W. E. (1993). "Theory and Empirical Testing of Asset Pricing Models." In R. A. Jarrow, W. T. Ziemba, and V. Maksimovic (Eds.), *The Finance Handbook.* North Holland. (Forthcoming)

Florens, J. P., and Mouchart, M. (1982). "A Note on Noncausality." *Econometrica*, *50*, 583–591.

Florens, J. P., and Mouchart, M. (1985). "A Linear Theory for Noncausality." *Econometrica*, *53*, 157–175.

Garcia, R., and Perron, P. (1996). "An Analysis of the Real Interest Rate under Regime Shifts." *Review of Economics and Statistics*, *78*, 111–125.

Giudici, P., Rydén, T., and Vandekerkhove, P. (1999). "Likelihood Ratio Tests for Hidden Markov Models." *Biometrics.* (Forthcoming)

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

Granger, C. W. J., Engle, R. F., and Robins, R. P. (1986). "Wholesale and Retail Prices: Bivariate Time Series Modelling with Forecastable Error Variances." In D. Belsey and E. Kuh (Eds.), *Model Reliability.* Cambridge, MA.: MIT Press. (1–17)

Hamilton, J. D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica*, *57*, 357–384.

Hamilton, J. D. (1990). "Analysis of Time Series Subject to Changes in Regime." *Journal of Econometrics*, *46*, 39–70.

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

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

Holst, U., Lindgren, G., Holst, J., and Thuvesholmen, M. (1994). "Recursive Estimation in Switching Autoregressions with a Markov Regime." *Journal of Time Series Analaysis*, *15*, 489–506.

Jacobson, T., Lindh, T., and Warne, A. (1998). *Growth, Savings, Financial Markets and Markov Switching Regimes* (Working Paper No. 69). Sveriges Riksbank.

Karlsen, H. (1990). *A Class of Non-Linear Time Series Models.* PhD Thesis, Department of Mathematics, University of Bergen, Norway.

Kim, C. J. (1994). "Dynamic Linear Models with Markov-Switching." *Journal of Econometrics*, *60*, 1–22.

Krishnamurthy, V., and Rydén, T. (1998). "Consistent Estimation of Linear and Non-Linear Autoregressive Models with Markov Regime." *Journal of Time Series Analysis*, *19*, 291–307.

Lindgren, G. (1978). "Markov Regime Models for Mixed Distributions and Switching Regressions." *Scandinavian Journal of Statistics*, *6*, 81–91.

Lucas, R. E., Jr. (1972). “Expectations and the Neutrality of Money.” *Journal of Economic Theory*, *4*, 103–124.

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

MacKinnon, J. G., Haug, A. A., and Michelis, L. (1999). “Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration.” *Journal of Applied Econometrics*, *14*, 563–577.

Magnus, J., and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Chichester: John Wiley.

Ravn, M. O., and Sola, M. (1995). “Stylized Facts and Regime Changes: Are Prices Procyclical?” *Journal of Monetary Economics*, *36*, 497–526.

Rogoff, K. (1985). “The Optimal Degree of Commitment to an Intermediate Monetary Target.” *Quarterly Journal of Economics*, *100*, 1169–1190.

Rothman, P., Dijk, D. van, and Franses, P. H. (1999). *A Multivariate STAR Analysis of the Relationship Between Money and Output* (Research Report No. EI-9945/A). Econometric Institute.

Sims, C. A., Stock, J. H., and Watson, M. W. (1990). “Inference in Linear Time Series Models with Some Unit Roots.” *Econometrica*, *58*, 113–144.

Sola, M., and Driffill, E. J. (1994). “Testing the Term Structure of Interest Rates from a Stationary Markov-Switching var.” *Journal of Economic Dynamics and Control*, *18*, 601–628.

Stock, J. H., and Watson, M. W. (1989). “Interpreting the Evidence on Money-Income Causality.” *Journal of Econometrics*, *40*, 161–181.

Svensson, L. E. O. (1999). “Price Level Targeting versus Inflation Targeting: A Free Lunch?” *Journal of Money, Credit and Banking*, *31*, 277–295.

Swanson, N. R. (1998). “Money and Output viewed through a Rolling Windows.” *Journal of Monetary Economics*, *41*, 455–473.

Timmermann, A. (2000). “Moments of Markov Switching Models.” *Journal of Econometrics*, *96*, 75–111.

Toda, H. Y., and Phillips, P. C. B. (1993). “Vector Autoregressions and Causality.” *Econometrica*, *61*, 1367–1393.

Vredin, A., and Warne, A. (2000). *Unemployment and Inflation Regimes* (Working Paper No. 107). Sveriges Riksbank.

Warne, A. (1997). “Inference in Cointegrated VAR Systems.” *Review of Economics and Statistics*, *LXXIX*, 508–511.

Warne, A. (1999). *Estimation of Means and Autocovariances in a Markov Switching VAR Model.* (Manuscript, Research Department, Sveriges Riksbank, Sweden)

Weise, C. L. (1999). “The Asymmetric Effects of Monetary Policy: A Nonlinear Vector Autoregression Approach.” *Journal of Money, Credit and Banking*, *31*, 85–108.

Anders Warne, Research Department, Sveriges Riksbank, 103 37 Stockholm, Sweden

E-mail address: anders.warne@riksbank.se

URL: http://www.riksbank.com/

http://www.research.texlips.org/

http://www.econometrics.texlips.org/