EFFICIENT TESTS OF LONG-RUN CAUSATION IN TRIVARIATE VAR PROCESSES WITH A ROLLING WINDOW STUDY OF THE MONEY-INCOME RELATIONSHIP

JONATHAN B. HILL

Abstract. This paper develops a simple sequential multiple horizon non-causation test strategy for trivariate VAR models (with one auxiliary variable). We apply the test strategy to a rolling window study of money supply and real income, with the price of oil, the unemployment rate and the spread between the Treasury bill and commercial paper rates as auxiliary processes. Ours is the first study to control simultaneously for common stochastic trends, sensitivity of causality tests to chosen sample period, null hypothesis over-rejection, sequential test size bounds, and the possibility of causal delays. Evidence suggests highly significant direct or indirect causality from M1 to real income, in particular through the unemployment rate and M2 once we control for cointegration.

1. Introduction

We are interested in testing for linear causal patterns over multiple horizons within aggregate measures of income and money supply, with macroeconomic control variables. In particular, we test for the precise horizon at which money growth causes disposable income growth, controlling for test sensitivity to chosen sample period and common stochastic trends of unknown form. In order to do so, we develop a recursive technique for characterizing typically nonlinear causality chains for a trivariate process X, Y and Z in terms of linear parametric restrictions, which lead to simple compound hypotheses for tests of multiple horizon non-causation when the auxiliary variable Z is scalar-valued. In principle X and Y can be of any vector dimension.

Noncausation from Y to X in a bivariate system implies non-causation at all horizons, irrespective of the vector dimensions of X and Y[1]. However, following the seminal studies of Granger (1969), cf. Wiener (1956), and Sims (1972), Sims (1980), Lütkepohl (1993) and Renault and Szafarz (1991) point out that indirect multi-step ahead causality from Y to X is possible in multivariate systems with

Date: May 2005.
Dept. of Economics, Florida International University, Miami, Fl; jonathan.hill@fiu.edu; www.fiu.edu/~hilljona.

Keywords and phrases: multiple horizon causality; Wald tests; parametric bootstrap; money-income causality; rolling windows; cointegration.
JEL classification: C12; C32; C53; E47.

I would like to thank two anonymous referees and the editor M Hashem Pesaran for helpful comments that lead to substantial improvements. All errors, if any, are of course mine.

[1]See Proposition 2.3 of Dufour and Renault (1998). See, also, Florens and Mouchart (1982).

auxiliary variables Z. Dufour and Renault (1998) set out a broad non-parametric and parametric theory of general horizon Granger-Wiener causality for discreet-valued processes, and demonstrate the nonlinear nature of parametric conditions for non-causality in multivariate VAR processes.

A simple, efficient test procedure for multi-step ahead causation that can be employed to characterize causality chains and causal neutralization[2], however, has yet to be established. The fundamental problem lies in the inherently nonlinear nature of parametric conditions for non-causality in VAR models, and the potential for asymptotic degeneracy of test statistics. Lütkepohl and Müller (1994) and Lütkepohl and Burda (1997) tackle the problem of degenerate Wald statistics, however at a severe cost of empirical power.

Using a more intuitive approach, Dufour *et al* (2003) regress W_{t+h} on $(W_t, ..., W_1)$, and a direct in-sample test of zero coefficient restrictions is all that is required. However, the method allows for a test of non-causality only at one horizon at a time; a new VAR model must be estimated for each horizon making cross-horizon comparisons particularly difficult; the method usually cannot itself be used to distinguish between simple non-causation (the total absence of indirect causal routes) and causal neutralization; and non-causation over horizons $1...h$ followed by causality at $h+1$ can occur *only if* an indirect causality chain exists: the procedure of Dufour *et al* (2003) does not ensure such a logical outcome is deduced[3]. Nonetheless, attractive features of this procedure are its relative ease of implementation, and the fact that it can be used on a multivariate VAR process of arbitrary dimension.

Chao *et al* (2001) and Corradi and Swanson (2002) consider linear and non-linear out-of-sample tests of non-causality. Like Dufour *et al* (2003), this method can be applied to vector processes of arbitrary dimension, only tests for non-causality at a particular horizon, and cannot be used in a simple efficient fashion to address causality chains[4].

In this paper, we develop recursive parametric representations of causality chains for trivariate, discreet-time VAR processes in the case of one scalar-valued auxiliary variable Z. The recursions developed here imply greatly simplified sequential linear restrictions for performing in-sample tests of non-causality up to arbitrary time horizons.

Although we allow the vector innovation components to be contemporaneously correlated, we make no attempt to consider causality and causal chains from the perspective of impulse response functions, a la Wold decompositions and forecast error variance decompositions, and so-called instantaneous causality, cf. Granger

[2] *Causal neutralization* from Y to X occurs when multiple causal routes at some time horizon $h \geq 2$ exist through Z, yet cancel each other out such that noncausation holds.

[3] For example, in their study of monthly GDP (X), the federal funds rate (Y), the GDP deflator and non-borrowed reserves (Z), horizon specific tests suggest Y fails to cause X for horizons 1 and 2, and causes X at horizon $h = 3$. This is possible only if an indirect causal route $Y \to Z \to X$ exists. However, their test procedure reveals that Y fails to cause Z one-month, a characteristic that implies *noncausation at all horizons*, which contradicts their conclusion.

[4] In general, the exixstence and detection of patterns "causality" will be sensitive to functional form, causal moment, and in-sample versus out-of-sample techniques. See Comte and Lieberman (2000) for details on the various "orders" of causality. For a smooth-transition autoregression model of 1-step ahead causality, see Rothman *et al* (2001). For a consistent out-of-sample predictive accuracy Bierens-type test (usable as a test of non-linear causality), see Corradi and Swanson (2002).

(1988), Lütkepohl (1993) and Swanson and Granger (1997)[5]. Our methodology is grounded on discreet-time metric linear projection theory in Hilbert space, and closely follows Dufour and Renault's (1998) theory of linear Granger-Wiener causality with respect to observable information.

We focus entirely on trivariate processes with one scalar auxiliary variable for expository and empirical reasons. Low dimension models are still employed in many empirical studies of causality (e.g. bivariate volatility spillover: see Hong, 2001; Hiemstra and Jones, 1994; and Brooks 1998). Indeed, bivariate applications using VARs still abound: see, for example, Coe and Nason (2004). Moreover, a causal chain $Y \to Z \to X$ implies Y will eventually cause X if Z is univariate, and linear necessary and sufficient conditions for non-causation up to arbitrary horizons are available in all cases (see Theorem 2.1, below).

A related technique of sequential linear hypothesis construction for models with multiple auxiliary processes is certainly achievable. Apparently, however, rather restrictive assumptions must be imposed (e.g. Y causes only one component of Z), even for an analysis of 2-step ahead causation. We provide a special case when non-causation up to horizon h in a system with one scalar auxiliary variable is equivelant to non-causation when multiple auxiliary variables are included. In general, we leave the topic of multivariate auxiliary variables for future research.

We apply our test procedure to the classic question of whether monetary dynamics influence the growth of real income. We use monthly M1 and real disposable income, with the unemployment rate, M2, the price of oil and the spread between the Treasury bill rate and the commercial paper rate as auxiliary variables. We study the causal properties from money to income in the sample period Jan. 1959-Dec. 2002. In order to control for the possibility of parametric evolution with respect to detectable causal patterns and test sensitivity to the chosen sample period, we study causation over rolling sample windows of fixed and increasing length, with a minimum length of 324 months. Moreover, we employ conventional and bootstrap test techniques that are robust to unknown forms of cointegration, a la Toda and Yamamoto (1995), and derive an upper bound of the test size due to the sequential nature of the test method. This is the first such study (to the best of our knowledge) to control simultaneously for each method/issue just described.

Controlling only for integration and using the complete data sample, we find evidence for the presence of delayed causal effects from the growth of M1 to the growth of income, in particular through the unemployment rate. Using rolling windows, we find significant evidence of a causal delay of 1-3 months before growth in M1 anticipates income growth, with the longest delay occurring through the unemployment rate. Once cointegration is controlled, we find substantially significant evidence that money causes real income 1 or 2-months ahead through M2, the unemployment rate and the price of oil, a result that strongly supports the major findings of Swanson (1998).

---

[5]Prediction-based non-causation for observable variables (e.g. Y to X) in the sense of Granger (1969) does not typically generalize to non-causation with respect to innovations in an impulse-response framework (e.g ϵ_y to X): see Dufour and Tessier (1993) and Dufour and Renault (1998). See Swanson and Granger (1997) for issues related to Wold-forms, vector innovation decompositions, and contemporaneous causality through innovation correlatedness, and an insightful application to macroeconomic settings of the graph-theoretic approach (e.g. Studený and Bouckaert, 1998) to vector-innovation causal chains.

Studies of statistical causal relationships within aggregate measures of money and real income have expanded substantially since the seminal investigations of Sims (1972, 1980). The bivariate methodologies of Sims (1972, 1980) and Christiano and Ljungqvist (1988) were augmented in numerous subsequent studies. See the seminal money-income study of Stock and Watson (1989), and see, e.g., and Friedman and Kuttner (1993). These now classic studies do not consider common stochastic trends, data dependent model selection, the evolution of causal patterns over time, causal delays, nor the tendency for Wald statistics to lead to over-rejections of the null of non-causation. Sims *et al* (1990), Toda and Phillips (1993,1994) and Toda and Yamamoto (1995) each consider problems associated with testing 1-step ahead causation in VAR systems with integrated processes or cointegrating relationships. Thoma (1994) and Swanson (1998) consider rolling windows of data samples in order to control for the dependence of test results on the chosen data period. Thoma (1994) does not use a data dependent technique for VAR order selection, does not consider the possibility of cointegration; and Swanson (1998) controls for common stochastic trends and test sensitivity to chosen sample period using a rolling window approach[6]. Neither study controls for causal delays and the possibility that the asymptotic distribution may be a poor proxy for the true small sample distribution.

Finally, we do not consider real-time data: we only use the latest time series available, and do not control for the fact that several time series used in this study are periodically updated. See Amato and Swanson (2001) who find that money fails to Granger-cause (1-step ahead) output when real-time data is used in VARs and VECMs, using standard in-sample and out-of-sample test procedures.

The rest of the paper contains the following topics. In Section 2 we define h-step ahead causation, and provide parametric characterizations of causal chains for trivariate and multivariate VAR processes in Section 3. Section 4 develops the test strategy for causation at or up to general horizons $h \geq 1$. Section 5 contains the empirical study. Section 6 concludes with parting comments. Appendix 1 contains all tables, a simulation study is performed in Appendix 2, and all proofs are contained in Appendix 3.

Throughout, we employ the following notation conventions. We write $U_t \perp V_t$ for m-vector processes U_t and V_t to denote orthogonality between all scalar components for all t, $u_{i,t} \perp v_{j,t}$, $i,j = 1...m$, which in $L_2(\Omega, \mathfrak{F}_t, Q)$ implies $E(u_{i,t}v_{j,t}) = 0$ for every $i,j = 1...k$ and every t. For an m-vector-process $\{W_t : t \in \mathbb{Z}\}$, let $W(-\infty, t]$ denote the Hilbert space spanned by the components $W_{i,s} : i = 1...m$, $s \leq t$. For Hilbert spaces A and B, we write $A + B$ to denote the Hilbert space spanned by all components of A and B.

[6]Swanson's (1998) fixed window lengths are set at 10 and 15 years. A window of length 10 years has 120 observations before lagging, and a 4-vector VAR(12) model of first differences, optimally selected in many windows considered in Swanson's (1998) study, implies a degrees of freedom of $120 - 12 - 13 \times 4 = 56$, assuming intercepts are included and controlling for sample truncation due to the 12 lags. Despite such a potentially low degrees of freedom, the chi-squared distribution is used for all Wald tests, for all models and for all sample periods. In the present paper, we use windows of a minimal length of 324 months, as well as a bootstrap test method: both the larger degrees-of-freedom and the small sample test method will in principle improve inference accuracy.

2. Causality Preliminaries

We define non-causality in the manner of Granger (1969), which was augmented to a multiple horizon Hilbert space framework by Dufour and Renault (1998). Consider some m-vector processes $\{W_t\}$ with trivariate representation $W_t = (X_t', Y_t', Z_t')'$, where X_t, Y_t, and Z_t have dimensions $m_x \geq 1$, $m_y \geq 1$ and $m_z \geq 0$ respectively, and $m = m_x + m_y + m_z \geq 2$. We assume W_t is defined in $L_2(\Omega, \mathfrak{F}_t, Q)$, where $\mathfrak{F}_t$ denotes an increasing σ-field of all past and present information at time t. Denote by $I(t)$ the information universe at time t which contains $X(-\infty, t]$ and any information available in all periods $H = \cap_{t\in\mathbb{Z}} I(t)$ (e.g. starting conditions and constants), and let $I_{XZ} = I_{XZ}(t) = H + X(-\infty, t] + Z(-\infty, t]$ for arbitrary t.

In principle, none of the following results rely on stationarity assumptions. For example, we may allow time to be bounded in the finite past. For brevity, however, we consider only an unbounded past.

We say the subvector Y "*does not cause*" X *at horizon* $h > 0$ in some Hilbert space (denoted $Y \overset{h}{\nrightarrow} X|I_{XZ}$) if $P(X_{t+h}|I_{XZ}(t) + Y(-\infty, t]) = P(X_{t+h}|I_{XZ}(t))$: the inclusion of $Y(-\infty, t]$ does not improve the L_2-orthogonal projection of X_{t+h} for all t; Y "*does not cause*" X *up to horizon* $h > 0$ (denoted $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$) if $P(X_{t+k}|I_{XZ}(t) + Y(-\infty, t]) = P(X_{t+h}|I_{XZ}(t))$ for each $k = 1...h$ and for all t; and Y "*does not cause*" X *at any horizon* $h > 0$ (denoted $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$) if $P(X_{t+h}|I_{XZ}(t) + Y(-\infty, t]) = P(X_{t+h}|I_{XZ}(t))$ for every $h > 0$ and for all t.

It is important to point out that causation $Y \overset{h}{\rightarrow} X$ occurs *if and only if* at least one scalar component of the closed linear span of $Y_{i,s}$, $i = 1...m_y$, $s \leq t$, improves a forecast of at least one scalar component $X_{j,t+h}$, $j = 1...m_x$.

2.1. Non-parametric Preliminaries.

The following results will be useful for subsequent discourse regarding causality chains and the dimension of Z. Results (i)-(iii) follow straightforwardly from Propositions 2.3 and 2.4 of Dufour and Renault (1998). We provide a proof only for (iv). Each process X, Y and Z are of arbitrary dimension unless otherwise noted.

Theorem 2.1 *i. If* $Y \overset{1}{\nrightarrow} (X, Z)|I_{XZ}$, *or* $(Y, Z) \overset{1}{\nrightarrow} X|I_{XZ}$, *then* $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$; *ii. If* $m_z \geq 2$ *and* $Z = (Z_1', Z_2')'$ *for arbitrary sub-vectors* Z_i, *and if* $(Y, Z_2) \overset{1}{\nrightarrow} (X, Z_1)|I_{XZ_1}$, *then* $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$; *iii. In order for non-causation* $Y \overset{1}{\nrightarrow} X|I_{XZ}$ *to be followed by causation* $Y \overset{h}{\rightarrow} X|I_{XZ}$, *for any* $h > 1$, *it is necessary for* $Y \overset{1}{\rightarrow} Z \overset{1}{\rightarrow} X$; *iv. If* Z *is scalar-valued and* $Y \overset{1}{\rightarrow} Z \overset{1}{\rightarrow} X$, *then* $Y \overset{h}{\rightarrow} X|I_{XZ}$ *for some* $h \geq 1$.

Remark 1: If the auxiliary process Z affords the partition $Z = (Z_1', Z_2')'$ such that $(Y, Z_2) \overset{1}{\nrightarrow} (X, Z_1)|I_{XZ_1}$, then no form of causal chain can exist, and $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$: even if $Y \overset{1}{\rightarrow} Z_1$ and/or $Z_2 \overset{1}{\rightarrow} X$, causal-chains are broken by $Y \overset{1}{\rightarrow} Z_1 \overset{1}{\nrightarrow} X$ or $Y \overset{1}{\nrightarrow} Z_2 \overset{1}{\rightarrow} X$ or $Y \overset{1}{\rightarrow} Z_1 \overset{1}{\nrightarrow} Z_2 \overset{1}{\rightarrow} X$, etc. Similarly, if non-causation $Y \overset{1}{\nrightarrow} X|I_{XZ}$ holds, and $Y \overset{1}{\nrightarrow} Z|I_{XZ}$ and/or $Z \overset{1}{\nrightarrow} X|I_{XZ}$ holds, then a broken causal chain exists, and non-causation for all horizons exists.

Remark 2: Because $Y \overset{1}{\nrightarrow} (X, Z)$ or $(Y, Z) \overset{1}{\nrightarrow} X$ are sufficient for $Y \overset{(\infty)}{\nrightarrow} X$, non-causation $Y \overset{1}{\nrightarrow} X|I_{XZ}$ followed by causation $Y \overset{h}{\rightarrow} X|I_{XZ}$, $h \geq 2$, can only occur if a causality chain exists, $Y \overset{1}{\rightarrow} Z \overset{1}{\rightarrow} X$. However, except in the univariate Z-case, a causality chain $Y \overset{1}{\rightarrow} Z \overset{1}{\rightarrow} X$ is generally not sufficient for causation Y

$\stackrel{h}{\rightarrow} X|I_{XZ}$, $h \geq 2$, due to the possibility that multiple causal routes through the auxiliary variables Z may cancel each other out (causal neutralization). When Z is scalar-valued, then a causality chain $Y \stackrel{1}{\rightarrow} Z \stackrel{1}{\rightarrow} X$ implies Y will eventually cause X.

2.2. **Parametric Preliminaries.** Denote by $W(-\infty, t]$ the closed linear span of $W_{i,s} : i = 1...m,\ s \leq t$, and let H be empty for brevity. Assume W_t has an autoregressive representation

$$W_t = \sum\nolimits_{i=1}^{\infty} \pi_i W_{t-i} + \epsilon_t, \quad \epsilon_t \perp W(-\infty, t-1] = 0, \tag{2.1}$$

where ϵ_t denotes a mean-zero m-vector L_2-orthogonal to the subspace $W(-\infty, t-1]$, with non-singular moment matrix $E[\epsilon_t \epsilon_t']$. The coefficients π_i are real-valued $m \times m$ matrices for each i, and the infinite series $\sum_{i=1}^{\infty} \pi_i W_{t-i}$ is assumed to converge in mean-square. In what follows, we explicitly ignore the issue of cointegration and VECM's, however only slight modifications to (2.1) and the following discourse is required to include this case.

Note that we allow the innovations to be contemporaneously correlated across components $\epsilon_{i,t}$, but not serially correlated ($\epsilon_t \perp \epsilon_s,\ s \neq t$). Thus, the distributed lag $\sum_{i=1}^{\infty} \pi_i W_{t-i}$ represents the best *linear* 1-step ahead forecast of W_t, $P(W_t|W(-\infty, t-1])$, but not necessarily the best 1-step ahead forecast, $P(W_t|I(t-1))$, although the two coincide for Gaussian vector processes. The setup in (2.1) is fairly standard (e.g. Lütkepohl, 1991), but does not preclude the possibility of nonlinear causal relationships (nor, indeed, of second-order causal patterns). In the following, we therefore use the notation $Y \stackrel{(h)}{\nrightarrow} X|I_{XZ}$ strictly to imply "linear predictive" non-causation: $Y \stackrel{(h)}{\nrightarrow} X|I_{XZ}$ *if and only if* $P(X_{t+h}|I_{XZ}(t) + Y(-\infty, t-1]) = P(X_{t+h}|I_{XZ}(t))$. See Dufour and Renault (1998) and Comte and Lieberman (2000)[7].

By Hilbert projection operator linearity, and error orthogonality, the h-step ahead projection $\hat{W}_{t+h}|I_W(t)$ of W_{t+h} onto the linear sub-space $W(-\infty, t]$ satisfies the recursion

$$\hat{W}_{t+h}|I_W(t) = \sum\nolimits_{i=1}^{\infty} \pi_i \hat{W}_{t+h-i} = \sum\nolimits_{i=1}^{\infty} \pi_i^{(h)} W_{t+1-i}, \tag{2.2}$$

where $\hat{W}_{t+h-i}|I_W(t) \equiv W_{t+h-i} \forall i \geq h$, and the coefficient matrix sequence $\{\pi_i^{(h)}\}_{i=1}^{\infty}$ is defined by the recursive relationship

$$\pi_1^{(0)} = I_m, \quad \pi_j^{(1)} = \pi_j, \quad \pi_j^{(h+1)} = \pi_{j+1}^{(h)} + \pi_1^{(h)} \pi_j. \tag{2.3}$$

See, e.g., Dufour and Renault (1998: eq. 3.8).

---

[7]Most $L_2(\Omega, \mathfrak{F}_t, Q)$ processes of interest will have a representation (1) either in levels, or after some standard transformation, e.g. first differencing. Nonetheless, in Hill (2004) we find that several processes used in our empirical study of money and income demonstrate highly significant patterns of smooth-transition autoregressive nonlinearity. See, also, Rothman *et al* (2001). Despite the inherent shortcomings associated with linear time series models, however, nonlinear models do not typically afford straightforward recursive parametric causal chain representations (e.g. the STAR model of Rothman *et al,* 2001), even though a consistent nonlinear out-of-sample test of non-causality at a particular horizon is available (Corradi and Swanson, 2002). Apparently there does not exist (to our knowledge) a published work which details testable parametric causality chains for inherently non-linear models (e.g. SETAR). I would like to thank an anonymous referee for pointing out the important issue of nonlinearity in the present context.

Consider the $(X', Y', Z')'$-conformable partition of the coefficient sequence

$$\pi_j^{(h)} = \begin{bmatrix} \pi_{XX,j}^{(h)} & \pi_{XY,j}^{(h)} & \pi_{XZ,j}^{(h)} \\ \pi_{YX,j}^{(h)} & \pi_{YY,j}^{(h)} & \pi_{YZ,j}^{(h)} \\ \pi_{ZX,j}^{(h)} & \pi_{ZY,j}^{(h)} & \pi_{ZZ,j}^{(h)} \end{bmatrix}. \tag{2.4}$$

For example, for every $j \geq 1$, $\pi_{XY,j}^{(h)}$ denotes an $m_x \times m_y$ matrix of constant real numbers.

The following theorem, due to Dufour and Renault (1998: Theorem 3.1), provides a fundamental nonlinear basis for parametric tests of non-causality h-steps ahead.

Theorem 2.2 *Consider any m-vector process $W_t = (X_t', Y_t', Z_t')'$ that satisfies (2.1). $Y \overset{h}{\nrightarrow} X|I_{XZ}$ if and only if $\pi_{XY,j}^{(h)} = 0$, $\forall j = 1, 2, ...$*

3. Causality Chains

Because $Y \overset{1}{\nrightarrow} X$ and $Y \overset{1}{\nrightarrow} Z$ will imply non-causation at all horizons, $Y \overset{(\infty)}{\nrightarrow} X$ (cf. Theorem 2.1), we assume causation $Y \overset{1}{\rightarrow} Z$ throughout the remainder of the paper, unless otherwise noted. Without loss of generality, assume X and Y are univariate processes ($m_x = m_y = 1$)[8].

In order to understand what is required for non-causation to occur through some arbitrary horizon $h \geq 2$, consider $h = 2$. If $Y \overset{1}{\nrightarrow} X|I_{XZ}$ and (2.1) hold, then the orthogonal 1-step ahead projection of X_{t+2} is exactly

$$\hat{X}_{t+2}|I_W(t+1) = \sum\nolimits_{i=1}^{\infty} \pi_{XX,i} X_{t+2-i} + \sum\nolimits_{i=1}^{\infty} \pi_{XZ,i} Z_{t+2-i}. \tag{3.1}$$

Whether Y causes X at any other horizon $h \geq 2$ depends on a causal chain through Z (Theorem 2.1.*iii*,*iv*), and therefore on the coefficients $\pi_{XZ,i}$. Projecting both sides onto $I_{XZ}(t) + Y(-\infty, t]$, we obtain the best 2-step ahead forecast of X_{t+2} by iterated projections and $Y \overset{1}{\nrightarrow} X|I_{XZ}$

$$\begin{aligned} \hat{X}_{t+2}|I_W(t) \quad = \quad & \pi_{XX,1}\hat{X}_{t+1}|I_{XZ}(t) + \pi_{XZ,1}\hat{Z}_{t+1}|I_W(t) \\ & + \sum\nolimits_{i=2}^{\infty} \pi_{XX,i} X_{t+2-i} + \sum\nolimits_{i=2}^{\infty} \pi_{XZ,i} Z_{t+2-i}. \end{aligned} \tag{3.2}$$

Clearly $\hat{X}_{t+2}|I_W(t) \in I_{XZ}(t)$ such that $Y \overset{2}{\nrightarrow} X|I_{XZ}$ *if and only if* $\pi_{XZ,1}\hat{Z}_{t+1}|I_W(t) \in I_{XZ}(t)$ with probability one for all t. If Z is vector-valued, then $\pi_{XZ,1}\hat{Z}_{t+1}|I_W(t) \in I_{XZ}(t)$ is feasible simply via nonlinear row-column combinations and the cancellation of Y-components.

Because the span $I_{XZ}(t) + Y(-\infty, t]$ can be written as $X(-\infty, t] + Y(-\infty, t] + Z(-\infty, t]$, we may write $\hat{Z}_{t+1}|I_W(t) = a_{zx}(t) + a_{zy}(t) + a_{zz}(t)$ for some elements $a_{zx}(t) \in X(-\infty, t]$, $a_{zy}(t) \in Y(-\infty, t]$ and $a_{zz}(t) \in Z(-\infty, t]$. Hence, $\hat{X}_{t+2}|I_W(t) \in I_{XZ}(t)$ *if and only if*

$$\pi_{XZ,1}\hat{Z}_{t+1}|I_W(t) \in I_{XZ}(t) \Rightarrow \pi_{XZ,1} a_{zy}(t) \in I_{XZ}(t). \tag{3.3}$$

[8]Dufour and Renault (1998) prove that noncausation from vector process Y to vector process X is equivelant to noncausation from each scalar component Y_i to each scalar component X_j. Thus, it suffices to consider the causal structure from Y to X by considering the scalar components individually.

If the element $a_{zy}(t) \in I_{XZ}(t)$ for all t[9], then $\hat{Z}_{t+1}|I_W(t) \in I_{XZ}(t)$ and $Y \overset{1}{\nrightarrow} Z|I_{XZ}$. Conversely, if $Y \overset{1}{\rightarrow} Z$ then $a_{zy}(t) \neq 0$ with probability one for some t, hence $\pi_{XZ,1}a_{zy}(t) \in I_{XZ}(t)$ for all t *if and only if* $\pi_{XZ,1}a_{zy}(t) = 0$ with probability one for all t. If Z is scalar-valued then $\pi_{XZ,1}a_{zy}(t) = 0$ *if and only if* $\pi_{XZ,1} = 0$[10].

3.1. **Recursive Representations for $m_z = 1$.** The above example through 2-steps can be easily replicated through h-horizons ahead. The coefficient recursion (2.3) renders the XY^{th}-block of π_j as

$$\pi^{(h+1)}_{XY,j} = \pi^{(h)}_{XY,j+1} + \pi^{(h)}_{XX,1}\pi_{XY,j} + \pi^{(h)}_{XY,1}\pi_{YY,j} + \pi^{(h)}_{XZ,1}\pi_{ZY,j}. \tag{3.4}$$

It follows that if non-causality up to horizon h is true, $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$, then, cf. Theorem 2, $\pi^{(k)}_{XY,j} = 0$ for each $k = 1...h$, and subsequently, cf. (3.1) and Theorem 2, $Y \overset{h+1}{\nrightarrow} X$ *if and only if*

$$\pi^{(h+1)}_{XY,j} = \pi^{(h)}_{XZ,1}\pi_{ZY,j} = 0, \forall j \geq 1. \tag{3.5}$$

See Corollary 3.1 of Dufour and Renault(1998). Thus, non-causality up to horizon $h \geq 1$ and causality at $h+1$ can only occur if a causality chain exists such that $\pi^{(h)}_{XZ,1}\pi_{ZY,j} \neq 0$, for some $j \geq 1$. Provided $Y \overset{1}{\rightarrow} Z|I_{XZ}$, then some scalar component of $\pi_{ZY,j}$ is non-zero for some $j \geq 1$.

If the auxiliary variable Z is scalar-valued ($m_z = 1$) and if for any j we have $\pi_{ZY,j} \neq 0$, then $Y \overset{1}{\rightarrow} Z$, cf. Theorem 2.1, and we conclude from (3.5) that $Y \overset{h+1}{\nrightarrow} X$ follows *if and only if* $\pi^{(h)}_{XZ,1} = 0$. Simply by using recursion (2.3) in a sequential manner we may deduce a simple characterization of the parameter $\pi^{(h)}_{XZ,1}$.

Lemma 3.1 *Consider the VAR process* (2)*, and let* $m_z = 1$*. Assume non-causation* $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ *for any* $h \geq 2$*, and causation* $Y \overset{1}{\rightarrow} Z|I_{XZ}$ *are true. Then* $\pi^{(2)}_{XZ,1} = \pi_{XZ,2}$ *for* $h = 2$*, and for any other* $h > 2$*,*

$$\pi^{(h)}_{XZ,1} = \pi_{XZ,h} + \sum\nolimits_{i=1}^{h-1} \left(\pi^{(h-i)}_{XX,1}\pi_{XZ,i}\right). \tag{3.6}$$

The following theorem delivers a simple linear necessary and sufficient condition for non-causality up to horizon $h \geq 1$.

Theorem 3.2 *Consider the VAR process* (2)*, and assume* $m_z = 1$*. Assume causation* $Y \overset{1}{\rightarrow} Z|I_{XZ}$ *is true.*

i. For all $h \geq 2$, $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ *if and only if* $Y \overset{1}{\nrightarrow} X|I_{XZ}$ *and* $\pi_{XZ,k} = 0$, $k = 1...h-1$;

ii.. For all $h \geq 2$*, if* $Y \overset{(h-1)}{\nrightarrow} X|I_{XZ}$*, then* $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ *if and only if* $\pi_{XZ,h-1} = 0$.

Remark 1: For any $h \geq 1$, non-causation $Y \overset{(h)}{\nrightarrow} X$ followed by causation $Y \overset{h+1}{\rightarrow} X$ is feasible only if a causal chain $Y \overset{1}{\rightarrow} Z \overset{1}{\rightarrow} X$ exists (cf. Theorem 2.1) and *if and only if* $\pi_{XZ,i} = 0$, $i = 1...h-1$, and $\pi_{XZ,h} \neq 0$. Conversely, if a causal chain

[9]Under the maintained assumptions this is possible only if $a_{zy}(t) = 0$ with probability one for all t.

[10]If $a_{zy}(t) = 0$ with probability one for all t, then $\hat{Z}_{t+1}|I_W(t) \in I_{XZ}(t)$, a contradiction of the assumption $Y \overset{1}{\rightarrow} Z|I_{XZ}$.

$Y \overset{1}{\rightarrow} Z \overset{1}{\rightarrow} X$ exists and Z is univariate, then $\pi_{XZ,h} \neq 0$ for *some* $h \geq 1$ and either $Y \overset{1}{\rightarrow} X$, or $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ followed by $Y \overset{h+1}{\rightarrow} X|I_{XZ}$, occurs.

The result that $Y \overset{1}{\nrightarrow} X|I_{XZ}$ and $\pi_{XZ,i} = 0$, $i = 1...h$, sequentially imply non-causation $Y \overset{(h+1)}{\nrightarrow} X|I_{XZ}$ when Z is univariate suggests a simple graph-theoretic representation of causality chains. See, e.g., Geiger and Pearl (1990) and Studený and Bouckaert (1998) for details on causal chain graph theory, and see Swanson and Granger (1997) for an application of the graph-theoretic approach to Wold-form innovations decompositions in a macroeconomic context. Because the chain representation $Y \overset{1}{\rightarrow} Z \overset{1}{\rightarrow} X$ neither suffices to suggest causation will even occur when Z is vector-valued, nor provides enough information concerning when causation takes place if Z is univariate, we adopt a more concise notation. Write $Y \overset{1:h_{ZY}}{\rightarrow} Z$ to imply Y causes Z one-step ahead, and $Y_{t-h_{ZY}}$ is the most recent component of $Y(-\infty, t\ ]$ to enter into $\hat{Z}_{t+t}|I_W(t)$. For any $h \geq 1$, if $Y \overset{1}{\nrightarrow} X|I_{XZ}$ and if Z is univariate, then by Theorem 3.2 the chain graph[11]

$$Y \overset{1:h_{ZY}}{\rightarrow} Z \overset{1:h}{\rightarrow} X \tag{3.7}$$

provides the unambiguous interpretation that $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ and $Y \overset{h+1}{\rightarrow} X|I_{XZ}$. Indeed, the notation $Y \overset{1:h_{ZY}}{\rightarrow} Z \overset{1:h}{\rightarrow} X$ unambiguously implies $\hat{X}_{t+h+1}|I_W(t+1)$ is a simple linear function of Z_{t+1} and $\hat{Z}_{t+1}|I_W(t)$ is a simple linear function of some component of $Y(-\infty, t]$, hence, by iterated projections and the fact that Z is univariate, $\hat{X}_{t+h+1}|I_W(t)$ is a linear function of some component of $Y(-\infty, t]$.

3.2. **Multivariate vs. Univariate Z.** The sequential conditions of Theorem 3.2 are necessary and sufficient for non-causation if Z is univariate. For multivariate auxiliary variable models ($m_z > 1$), the conditions are not necessary although they remain sufficient. The non-necessity follows from the possibility that $\pi^{(h)}_{XZ,1}\pi_{ZY,j} = 0$ may be true when $\pi^{(h)}_{XZ,1} \neq 0$ and $Y \overset{1}{\rightarrow} Z$.

The chain notation $Y \overset{1:h_{ZY}}{\rightarrow} Z \overset{1:h}{\rightarrow} X$ in the multivariate Z-case, however, does not contain sufficient information to describe whether, when and how causation takes place. Consider a simple example: if Z is a 2-vector and $Y \overset{1}{\nrightarrow} X|I_{XZ}$, then $Y \overset{1:1}{\rightarrow} Z \overset{1:1}{\rightarrow} X$ need only imply $Y \overset{1:1}{\rightarrow} Z_1$ and $Z_2 \overset{1:1}{\rightarrow} X$, in which case non-causation $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$ occurs: a direct path from Y to X does not exist at $h = 2$.

In general, such causal chain graphs may not have much use in multivariate auxiliary variable models, unless strong assumptions are imposed. For example, if $Z = (Z_1, Z_2')'$ where Z_1 is a scalar and Z_2 has an arbitrary dimension, then $(Y, Z_2) \overset{1}{\nrightarrow} X|I_{XZ_1}$ implies that the use of $(X, Y, Z_1)'$ and the conditions of Theorem 3.2 suffice to characterize non-causation from Y to X within the augmented system $(X, Y, Z_1, Z_2')'$.

Let $\delta^{(h)}_j$ denote the VAR coefficients in the projection of X_{t+h} onto the linear sub-space $I_{XZ_1}(t) + Y(-\infty, t]$, and consider a preliminary result.

Lemma 3.3 *Let $W = (X, Y, Z_1', Z_2')'$ where each Z_i has arbitrary dimension*

[11]The chain $Y \overset{1:h_{ZY}}{\rightarrow} Z \overset{1:h-1}{\rightarrow} X$ depicts a *directed, acyclic* chain: the arrows depict the direction of influence, and the chains are inherently acyclic because causation occurs over time and time is unidirectional. See, e.g., Geiger and Pearl (1990) and Studený and Bouckaert (1998).

$m_{z_i} \geq 0$. *Non-causation* $Y \overset{(h)}{\nrightarrow} X|I_{XZ_1}$ *occurs if and only if* $\delta^{(k)}_{XY,j} = 0$, $k = 1...h$, $j \geq 1$, *where*

$$\delta^{(k)}_{XY,j} \equiv \pi^{(k)}_{XY,j} + \sum\nolimits_{i=1}^{\infty} \pi^{(k)}_{XZ_2,i} \beta^{1-i}_{Z_2Y,j}, \tag{3.8}$$

and $\beta^{1-i}_{Z_2Y,j}$ *denotes the* Y*-specific coefficients in the projection of each vector* $Z_{2,t+1-i}$ *onto* $I_{XZ_1} + Y(-\infty, t]$, $i \geq 1$.

Remark 1: Formula (3.8) implies $Y \overset{1}{\nrightarrow} X|I_{XZ_1}$ ($\delta_{XY,k} = 0$, $\forall k$) and $Y \overset{1}{\rightarrow} X|I_{XZ}$ ($\pi_{XY,k} \neq 0$, for some k) may simultaneously be true due to possible neutralization effects through the multiple causal routes from Y to Z_2 to X ($\pi_{XZ_2,i}\beta^{1-i}_{Z_2Y,k}$) and Y to X ($\pi_{XY,k}$). Notice that the coefficients $\beta^{1-i}_{Z_2Y,k}$ represent the marginal "impact" components of $Y(-\infty, t]$ have on contemporary and past Z_{t+1-i}, thus the chain "Y to Z_2 to X" covers only one period.

If $\pi_{XZ_2,i} = 0$ for all i, then $Z_2 \overset{1}{\nrightarrow} X|I_{XZ_1}$ and (3.8) dictates $Y \overset{1}{\nrightarrow} X|I_{XZ}$ *if and only if* $Y \overset{1}{\nrightarrow} X|I_{XZ_1}$. In general, we have the following result.

Theorem 3.4 *Let* $Z = (Z_1, Z_2')'$ *for some scalar* Z_1 *and vector* Z_2 *of arbitrary dimension* $m_{z_2} \geq 0$.

i. *If* $Z_2 \overset{1}{\nrightarrow} X|I_{XZ_1}$, *then* $Y \overset{1}{\nrightarrow} X|I_{XZ}$ *if and only if* $Y \overset{1}{\nrightarrow} X|I_{XZ_1}$.

ii. *If* $(Y, Z_2) \overset{1}{\nrightarrow} X|I_{XZ_1}$, *then for any* $h \geq 1$, $Y \overset{(h)}{\nrightarrow} X|I_{XZ_1}$ *implies* $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$, *and* $Y \overset{h+1}{\rightarrow} X|I_{XZ_1}$ *implies* $Y \overset{h+1}{\rightarrow} X|I_{XZ}$.

iii. *If* $(Y, Z_2) \overset{1}{\nrightarrow} X|I_{XZ_1}$ *and* $Y \overset{(h)}{\nrightarrow} X|I_{XZ_1}$, *then* $Y \overset{(h+1)}{\nrightarrow} X|I_{XZ}$ *if and only if* $\pi_{XZ_1,h} = 0$.

Remark 1: The implication that $(Y, Z_2) \overset{1}{\nrightarrow} X|I_{XZ_1} \Longrightarrow Y \overset{1}{\nrightarrow} X|I_{XZ}$ is parallel to Dufour and Renault's (1998) Proposition 2.4 (see also Theorem 2.1, above). They prove a more restricted implication that if Z satisfies the "separation" condition $I_{XZ} = I_{XZ_1} + Z_2(-\infty, t]$, where $I_{XZ_1} = H + X(-\infty, t] + Z_1(-\infty, t]$, then $(Y, Z_2) \overset{1}{\nrightarrow} (X, Z_1)|I_{XZ_1}$ is sufficient for $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$. Similarly, see Corollary 3.6 of Dufour and Renault (1998) in which the conditions $Y \overset{1}{\nrightarrow} (X, Z)|I_{XZ}$ or $(Y, Z) \overset{1}{\nrightarrow} X|I_X$ are shown to be necessary *and* sufficient for non-causation at all horizons, $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$, when Z is univariate.

Remark 2: If $(Y, Z_2) \overset{1}{\nrightarrow} X|I_{XZ_1}$ then the chain graph $Y \overset{1:h_{Z_1Y}}{\rightarrow} Z_1 \overset{1:h}{\rightarrow} X$ has an unambiguous interpretation for either the reduced system $(X, Y, Z_1)'$, $Y \overset{(h)}{\nrightarrow} X|I_{XZ_1}$ and $Y \overset{h+1}{\rightarrow} X|I_{XZ_1}$, or the augmented system $(X, Y, Z_1, Z_2')'$, $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ and $Y \overset{h+1}{\rightarrow} X|I_{XZ}$.

4. Tests for Causation at Arbitrary Time Horizons

We now construct a strategy for testing non-causality up to arbitrary time horizons, and analyze test size bounds.

4.1. **Sequential Test**

. Step 1: **Test** $Y \overset{(\infty)}{\nrightarrow} X$

We test both

$$Y \overset{1}{\nrightarrow} (X, Z) \tag{Test 0.1}$$

$$(Y, Z) \overset{1}{\nrightarrow} X. \tag{Test 0.2}$$

Evidence in favor of either hypothesis provides evidence in favor of $Y \overset{(\infty)}{\nrightarrow} X$, cf. Theorem 2.1. We proceed to test $Y \overset{1}{\nrightarrow} X$ only if we reject both tests.

Step 2: **Test individually $Y \overset{1}{\nrightarrow} X$, $Y \overset{1}{\nrightarrow} Z$, and $Z \overset{1}{\nrightarrow} X$**

If the hypothesis

$$Y \overset{1}{\nrightarrow} X \tag{Test 1.0}$$

is rejected, the test procedure is stopped. If we find evidence in favor of a broken causal chain by failing to reject either

$$Y \overset{1}{\nrightarrow} Z \tag{Test 1.1}$$

$$Z \overset{1}{\nrightarrow} X \tag{Test 1.2}$$

then evidence suggests non-causation at all horizons, $Y \overset{(\infty)}{\nrightarrow} X$. If we reject both Tests 1.1 and 1.2, we proceed to Step 3.

Step 3: **Test $Y \overset{(h)}{\nrightarrow} X$, $h \geq 2$**

By Theorem 3.2, provided $Y \overset{1}{\rightarrow} Z$ sequential evidence in favor of $\pi_{XZ,h-1} = 0$ is evidence in favor of non-causation up to horizon h. Thus, we test the linear compound hypothesis

$$Y \overset{1}{\nrightarrow} X, \pi_{XZ,i} = 0, i = 1...h-1 \tag{Test h.0}$$

for each $h \geq 2$. Failure to reject provides evidence in favor of $Y \overset{(h)}{\nrightarrow} X$.

4.2. **Size Bounds.** Due to the sequential nature of the tests of $Y \overset{(h)}{\nrightarrow} X | I_{XZ}$, $h = 1, 2, ...$, we require an upper bound on the test size. We reject $Y \overset{1}{\nrightarrow} X$ only if we first reject both Tests 0.1 and 0.2 ($Y \overset{(\infty)}{\nrightarrow} X$) and then reject Test 1.0 ($Y \overset{1}{\nrightarrow} X$); we reject $Y \overset{(2)}{\nrightarrow} X$ if we reject $Y \overset{1}{\nrightarrow} X$, or fail to reject $Y \overset{1}{\nrightarrow} X$, reject $Y \overset{1}{\nrightarrow} Z$ and reject $Z \overset{1}{\nrightarrow} X$, and reject the compound hypothesis $Y \overset{1}{\nrightarrow} X$, $\pi_{XZ,1} = 0$; and so on. Let $\alpha_{\#.\#}$ denote the nominal size of Test #.#.

Lemma 4.1 *Let $H_0^{(h)}$ denote the hypothesis $H_0 : Y \overset{(h)}{\nrightarrow} X$, for any $h \geq 1$. Let $p^{(h)} \equiv P(rej.\ H_0^{(h)} | H_0^{(h)}$ is true$)$.*

i. If $Y \overset{1}{\nrightarrow} Z \overset{1}{\rightarrow} X$ then $p^{(h)} \leq p_1 = \min\{\alpha_{0.1}, \alpha_{1.0} + (h-1) \times \alpha_{1.1}\}$;

ii. If $Y \overset{1}{\rightarrow} Z \overset{1}{\nrightarrow} X$ then $p^{(h)} \leq p_2 = \min[\alpha_{0.2}, \alpha_{1.0} + \sum_{i=2}^{h} \min\{\alpha_{1.2}, \alpha_{i.0}\}]$;

iii. If $Y \overset{1}{\nrightarrow} Z \overset{1}{\nrightarrow} X$ then $p^{(h)} \leq p_3 = \min[\alpha_{0.1}, \alpha_{0.2}, \alpha_{1.0} + (h-1) \times \min\{\alpha_{1.1}, \alpha_{1.2}\}]$;

iv. If $Y \overset{1}{\rightarrow} Z \overset{1}{\rightarrow} X$ then $p^{(h)} \leq p_4 = \sum_{i=1}^{h} \alpha_{i.0}$.

In general,

$$P\left(rej.\ H_0^{(h)} | H_0^{(h)} \text{ is true}\right) \leq \max_{1 \leq i \leq 4}\{p_i\}. \tag{4.1}$$

Bound (4.1) generalizes every possibility for a false rejection of $H_0^{(h)}$. Let $h \geq 2$. If $Y \overset{1}{\nrightarrow} Z \overset{1}{\rightarrow} X$ or $Y \overset{1}{\nrightarrow} Z \overset{1}{\nrightarrow} X$, for example, then the conditions outlined in Theorem 3.2 are only sufficient for non-causation, but not necessary. From (3.2), we may have $Y \overset{1}{\nrightarrow} X$, $\pi_{XZ,1} \neq 0$ and $Y \overset{(2)}{\nrightarrow} X$: in such a case if a consistent test

statistic is used, then there is a probability one asymptotically that we reject $Y \overset{1}{\nrightarrow} X$, $\pi_{XZ,1} \neq 0$ and falsely deduce $Y \overset{(2)}{\nrightarrow} X$. In cases (i) and (iii), the upper bound of the sequential test size embodies the probabilities of erroneous rejections of Tests 0.1 and 0.2 ($Y \overset{(\infty)}{\nrightarrow} X$) and Tests 1.1 and 1.2 ($Y \overset{1}{\nrightarrow} Z$ and $Z \overset{1}{\nrightarrow} X$). Neither bound depends on the nominal horizon-specific sizes $\alpha_{h.0}$ because the parametric conditions of Tests $h.0$ are not necessary for non-causation in these cases. The probability bound of a Type I error in these cases can be controlled simply by setting the nominal size $\alpha_{1.1}$ of the test $Y \overset{1}{\nrightarrow} Z$ to a small value (e.g. .01).

In practice, a simple rule will likely be applied. For example, put $\alpha_{0.1} = \alpha_{0.2} = \alpha$, $\alpha_{1.1} = \alpha_{1.2} = \alpha_{i.0} = \beta$ for each $i = 1...h$. Then (4.1) reduces to

$$P\left(rej.\ H_0^{(h)} | H_0^{(h)} \text{ is true}\right) \leq \max[\min\{\alpha, h \times \beta\}, h \times \beta] = h \times \beta, \tag{4.2}$$

the standard Bonferonni bounds, depending only on the common β.

5. U.S. Income and Monetary Aggregates

We now investigate the causal relationships within aggregate income, money, oil prices, unemployment and interest rates. For the period January 1959 - December 2002, we use the logarithm of monthly, seasonally adjusted, nominal M1 and M2 ($m1$, $m2$), the logarithm of seasonally adjusted disposable income (y), the logarithm of the spot oil price (o), the civilian unemployment rate (u), the 90-day Treasury bill rate (r_b), the 90-day commercial paper rate (r_p) and the spread between the two rates ($rr = r_b - r_p$).

Except for the commercial paper rate, all data are taken from the databases made publicly available by the Federal Reserve Bank of Saint Louis. The commercial paper rate was taken from the NBER data archive for the period 1959:01-1971:12, and from the Federal Reserve Bank of Saint Louis for the period 1972:01-2002:12. Seasonal adjustment, where applicable, was performed at the source. In order to control for any apparent trend, we pass all final (e.g. post-differenced) processes though linear trend filters.

Significant evidence suggests one positive unit root exists in each series, except for the rate spread r_b-r_p: the rate spread is likely $I(0)$, implying the process may represent one possible error correction term within a system of y, $m1$, $m2$, r_b, and r_p, with an error correction vector $(0, 0, 0, 1, -1)$[12]. Using industrial output y, aggregate money m, prices p, and the Treasury and commercial paper rates, Swanson (1998) finds in a rolling window framework the rate spread $r_b - r_p$ and the velocity of money $y + m - p$ are likely the only two error correction terms.

Considering the amassed, yet uneven, evidence in support of integration within the individual processes and cointegration between money, real income and interest rates, we implement two widely practiced VAR methods. We construct VAR models of de-trended first differences (except for the rate spread) in order to control for integration of order one: the processes are Δy, $\Delta m1$, $\Delta m2$, Δo, Δu, and rr. Second, we employ the excess-lag technique of Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) for VAR models of de-trended level processes in order to control

[12]Stock and Watson (1993) similarly find evidence of cointegration among M1, industrial output, and the Treasury bill rate. Hafer and Jensen (1991) find evidence for cointegration within M2, real income and a short-term interest rate at quarterly increments, and conclude all evidence for cointegration vanishes once M2 is replace by M1.

for cointegration of unknown form. For this procedure, we specify a VAR(p) model in levels adding lags equal to the maximum order of suspected integration d (in this case, $d = 1$), and test only the first $p - d$ coefficient matrices[13].

There is ample evidence in the literature that standard Wald tests in multivariate models tend to lead to over-rejection of null hypotheses: see, e.g., Dufour (2002) and Dufour *et al* (2003). A parametric bootstrap method for simulating small sample p-values has been shown to provide sharp approximations to the chosen significance level, although over-rejections may persist if the test statistic asymptotic distribution involves nuisance parameters (see, e.g., Andrews, 2000; and Dufour and Jouini, 2003). We perform standard and parametric bootstrap tests for each VAR method separately[14]. Consult Hill (2004a) for a simulation study demonstrating the merits of the sequential test strategy using standard and bootstrap test methods on VARMA processes.

We perform sequential tests on 3-vector systems with real disposable income y, money $m1$, and one auxiliary variable chosen from the set $\{m2,\ u,\ o,\ rr\}$ for the period 1959:01-2002:12, and for rolling sample periods of increasing and fixed width. VAR model orders are selected by minimizing the AIC over possible orders $p = 1...18$, subject to reasonably noisy residuals.

5.1. **Sample Period 1959-2002.** Extended test results for all auxiliary variables can be found in Table 1. For brevity, however, in the following we only discuss results based on the parametric bootstrap for models with either the unemployment rate or M2. Consult Hill (2004a) for complete discussions and test results using standard p-values.

5.1.1. ***Unemployment.*** For the process (Δy,$\Delta m1$,Δu), the optimally selected VAR order is 8, however portmanteau tests suggest lags may have been omitted. The minimum order for which we fail to reject the white noise hypothesis for the residual series is $p = 12$ (.131), however test results based on either level are qualitatively similar. We opt to use the parsimonious specification VAR(8). In order to control for cointegration of unknown form, the optimal order in levels is 9, hence we use a VAR(10) model.

We reject both hypotheses for $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ at the 10%-level (Test 0.1: .080, and Test 0.2: .040)[15]. We fail to reject $\Delta m1 \overset{1}{\nrightarrow} \Delta y$ (Test 1.0: .626), and reject at the

[13]Swanson *et al* (1996) demonstrate in a monte carlos study of tests of one-step ahead non-causation that the excess-lag method provides excellent empirical sizes, but tends to generate low power.

[14]The parametric bootstrap (i.e. asymptotic Monte Carlo test based on a consistent point estimate) is performed as follows for an arbitrary hypothesis: *i.* obtain estimated VAR coefficients, $\hat{\pi} = (\hat{\pi}_1, ..., \hat{\pi}_p)$, where p minimizes the AIC; *ii.* derive the test statistic, denoted T_n; *iii.* simulate J series $W_{t,j}$, $j = 1...J$, $t = 1...n$, based on the the estimated parameters $\hat{\pi}$ with the null hypothesis restrictions imposed (for example, a test of $Y \overset{1}{\nrightarrow} X$ imposes $\hat{\pi}_{XY,i} = 0$, $i = 1...p$): the process $W_{t,j}$ is simulated as $W_{t,j} = \sum_{i=1}^{p} \hat{\pi}_i W_{t-1} + \epsilon_t$ where ϵ_t are 3-vector *iid* draws from a standard normal distribution; *iv.* use the double-array $\{W_{t,j}\}_{t,j=1}^{n,J}$ to estimate J separate VAR(p) models, and generate J test statistics $T_{n,j}$ for the hypothesis in question; *v.* the approximate p-value is simply the percent frequency of the event $T_{n,j} > T_n$. For all tests, we set $J = 1000$. For first order asymptotic validity of the above parametric bootstrap, see Proposition 6.1 of Dufour (2005).

[15]Parenthetical values denote p-values derived from a parametric bootstrap.

nominal 5%-level sequentially only the compound Test 4.0, $\Delta m1 \overset{(4)}{\nrightarrow} \Delta y$[16]. If we perform each sequential test at the 1%-level, then we fail to reject $\Delta m1 \overset{(5)}{\nrightarrow} \Delta y$ at a bounded 5%-level; if we perform each test at the level of the smallest compound test p-value (i.e. .032), then we reject $\Delta m1 \overset{(4)}{\nrightarrow} \Delta y$ at a bounded 13%-level. We have, therefore, conflicting (and weak) evidence in support of either $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ or $\Delta m1 \overset{4}{\rightarrow} \Delta y$.

For the excess-lag VAR(10) model in levels, we fail to reject $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ (Test 0.2: .216) and $\Delta m1 \overset{1}{\nrightarrow} \Delta y$ (Test 1.0: .148). Evidence suggests a broken chain, $\Delta m1 \overset{1}{\rightarrow} \Delta u \overset{1}{\nrightarrow} \Delta y$ (Test 1.1: .000; Test 1.2: .376). If we pursue tests at subsequent horizons, we reject $\Delta m1 \overset{(2)}{\nrightarrow} \Delta y$ at a nominal 5%-level (Test 2.0: .024). If we perform each sequential test at the 1%-level, then we only reject $\Delta m1 \overset{(5)}{\nrightarrow} \Delta y$ at a bounded 5%-level, suggesting $\Delta m1 \overset{5}{\rightarrow} \Delta y$ and strengthening the above evidence for a causal delay when cointegration is ignored. Use of either method suggests at least three months pass before growth of the money supply anticipates growth of real income, through the intermediary impact fluctuations in the money supply have on the unemployment rate.

5.1.2. ***M2***. For a VAR model with M2, the minimum AIC order is $p = 6$ for first differences. The lowest order at which we fail to reject the white-noise hypothesis for the residual series is 10 (.370), hence we opt for the VAR(10) model. Similarly, the optimal order for levels is $p = 7$ while white-noise is detected in the VAR residuals at a lowest order of 11. We therefore opt to use a VAR(12) model of excess lags in levels.

In the VAR(10) case bootstrap tests fail to reject $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ (Test 0.2: .198), suggesting the growth of M1 never causes income growth. If we proceed to check individual horizons, we fail to reject $\Delta m1 \overset{1}{\nrightarrow} \Delta y$ (Test 1.0: .466), we find highly significant evidence for a causal chain, $\Delta m1 \overset{1}{\rightarrow} \Delta m2 \overset{1}{\rightarrow} \Delta y$, and reject the compound hypothesis $\Delta m1 \overset{(h)}{\nrightarrow} \Delta y$ at the nominal 1%-level only for $h = 11$, and therefore at a bounded 11%-level (Test 11.0: .006). This suggests either $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$, or nearly one year passes before fluctuations in the money supply will have an impact on real income.

For the excess-lag VAR(12) model in levels, we reject every null hypothesis considered at below the 1%-level: we immediately deduce $\Delta m1 \overset{1}{\rightarrow} \Delta y$. Similar to the model with the unemployment rate, once cointegration is controlled for significant evidence for causation expands sharply, supporting the major findings of Swanson (1998)[17].

---

[16]We find significant evidence of a causal chain $\Delta m1 \overset{1}{\rightarrow} \Delta u \overset{1}{\rightarrow} \Delta y$ (Test 1.1: .000, Test 1.2: .026). Indeed, for each auxililary variable Z in models of either levels or differences, we find evidence in favor of $\Delta m \rightarrow Z$, with the level of significance below .1%. Thus, evidence strongly suggests the non-causality conditions of Theorem 4 are necessary and sufficient: we will not comment on the issue below.

[17]It should be pointed out that Swanson (1998) uses an industrial production index as "real income", aggregate prices and several measures of supply of money (M1, M2 and the Divisia measure of money) in a multivariate model, and control for cointegration of unknown form by use

5.1.3. ***Oil Price.*** Using first differences, the VAR order $p = 4$ both minimizes the AIC and maximizes the portmanteau test p-value, however we still strongly reject the white noise hypothesis (.004). This suggests a severe form of model mis-specification may exist, possibly with respect to unmodeled cointegration[18]. We consider, therefore, both the parsimonious VAR(4) model in differences, and an optimally selected VAR(3.5) excess-lag model in levels.

In the VAR(4) case, we fail to reject both tests of $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ (Test 0.1: .856, Test 0.1: .560). If we decide to pursue subsequent tests, we fail to reject $\Delta m1 \overset{1}{\nrightarrow} \Delta y$ (Test 1.0: .9540), and find evidence for a broken causal chain, $\Delta m1 \overset{1}{\rightarrow} \Delta o \overset{1}{\nrightarrow} \Delta y$ (Test 1.1: .000, Test 1.2: .194). Compound tests do not reveal a causal delay: we fail to reject $\Delta m1 \overset{(h)}{\nrightarrow} \Delta y$ for each $h = 2...5$. This provides support for the prior evidence of non-causation at all horizons.

In the VAR(3.5) excess-lag model of levels, by comparison, we find weak evidence that non-causation at all horizons fails to hold, and we reject $\Delta m1 \overset{1}{\nrightarrow} \Delta y$ (Test 1.0: .026) at the 3%-level. If we decide to perform each test of $\Delta m1 \overset{(h)}{\nrightarrow} \Delta y$ at the 1%-level, then we fail to reject every compound test $\Delta m1 \overset{(h)}{\nrightarrow} \Delta y$.

5.1.4. ***Rate Spread.*** The optimal VAR order for first differences is 6, however we reject the white noise hypothesis for all orders considered, with the largest p-value (.020) occurring at $p = 13$. We consider, therefore, a VAR(3.5) model in differences, and a VAR(8) model in levels. For the VAR(3.5) model we fail to reject $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ (Test 0.1: .718) and fail to reject $\Delta m1 \overset{1}{\nrightarrow} \Delta y$ (Test 1.0: .964). If we pursue subsequent tests, we find evidence for $\Delta m1 \overset{1}{\rightarrow} rr$ (Test 1.1: .000) and fail to reject every subsequent hypothesis $\Delta m1 \overset{(h)}{\nrightarrow} \Delta y$.

For the excess-lag VAR(8) model in levels we still reject the white noise hypothesis for all orders considered. We now reject both tests of $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ and reject $\Delta m1 \overset{1}{\nrightarrow} \Delta y$ at the 5%-level (Test 1.0: .032). If we perform each test at the 1%-level, we fail to reject each test $\Delta m1 \overset{(h)}{\nrightarrow} \Delta y$, $h = 1...5$. Thus, while previous studies find the rate spread may be a statistically significant error correction term in a cointegrated VAR system of money, real income and interest rates, controlling for common stochastic trends within the present trivariate model here does not significantly alter the fundamental conclusion that a lengthy delay exists before causation occurs, if at all.

5.1.5. ***All Auxiliary Variables.*** Notice that we fail to reject the hypothesis $Z \overset{1}{\nrightarrow} \Delta y$ (Test 1.2) for each scalar $Z = \Delta u$, Δo, or rr. Moreover, we reject 1-month ahead noncausation from M1 to income in the truncated system $(\Delta y, \Delta m1, \Delta m2)$.

---

of the excess lag technique. We use real disposable income in a trivariate model (e.g. income, M1 and M2) similar in spirit to Boudjellaba *et al* (1992, 1994).

[18]Of course, numerous other types of mis-specification may be in play, including unmodeled conditional variance and nonlinearity. We do not pursue these topics in the present setting of tests of multiple horizon noncausation. See Rothmant et al (2001) for a vector smooth transition autoregression (STAR) study of one-step ahead noncausation between money and income, with auxiliary variables similar to those used in the present paper. See, also, Corradi and Swanson (2002) who develop an out-of-sample Bierens-type test of functional specification for VAR models of causality.

Based on the ideas presented in Section 3.2, it is worthwhile, therefore, to check if the causality properties in the augmented system $(\Delta y, \Delta m1, \Delta u, \Delta m2, \Delta o, rr)$ are the same as in the system $(\Delta y, \Delta m1, \Delta m2)$.

We estimated a VAR(12) excess lag model in levels[19], tested the joint hypothesis $(\Delta u, \Delta o, rr) \overset{1}{\nrightarrow} \Delta y$, and obtained a bootstrapped p-value of .229. Thus, evidence supports $Z_2 \overset{1}{\nrightarrow} X$ where $Z_2 = (\Delta u, \Delta o, rr)'$. By Theorem 3.4.$i$ we may infer that causation in the complete vector system matches what was obtained for the truncated system with $M2$, hence $\Delta m1 \overset{1}{\rightarrow} \Delta y | I_{\Delta y, \Delta m2}$ and $\Delta m1 \overset{1}{\rightarrow} \Delta y | I_{\Delta y, (\Delta u, \Delta m2, \Delta o, rr)}$. Of course, a classic 1-step ahead noncausation test can be performed directly: a test of $\Delta m1 \overset{1}{\nrightarrow} \Delta y | I_{\Delta y, (\Delta u, \Delta m2, \Delta o, rr)}$ produces a bootstrapped p-value of .022.

Evidence in the truncated system $(\Delta y, \Delta m1, \Delta u)$ suggests 1-month ahead noncausation from M1 to income, $\Delta m1 \overset{1}{\nrightarrow} \Delta y | I_{\Delta y, \Delta u}$, and we conclude 1-month ahead causation in the complete system, $\Delta m1 \overset{1}{\rightarrow} \Delta y | I_{\Delta y, (\Delta u, \Delta m2, \Delta o, rr)}$. Using the notation of Section 3.2, for the 3-vector $(\Delta y, \Delta m1, \Delta u)$ evidence therefore suggests $\delta_{XY,j} = 0$ for all j, for the complete system $(\Delta y, \Delta m1, \Delta u, \Delta m2, \Delta o, rr)$ evidence suggests $\pi_{XY,j} \neq 0$ for some j, and therefore neutralization $\pi_{XY,j} + \sum_{i=1}^{p} \pi_{XZ_2,i} \beta_{Z_2 Y, j}^{1-i} = 0$ is evidently occurring for some j, where $X = \Delta y$, $Y = \Delta m1$ and $Z_2 = (\Delta m2, \Delta o, rr)'$: the association between $\Delta m1$ and $(\Delta m2, \Delta o, rr)$, and the 1-month ahead causal impact $(\Delta m2, \Delta o, rr)$ has on Δy, evidently exactly offsets the causal influence $\Delta m1$ has on Δy when all auxiliary variables are present.

5.2. **Rolling Windows.** Finally, we study trivariate causal patterns in money and income over rolling sample periods of increasing and fixed length. Increasing windows begin and end with the sample periods 1959:01 - 1985:12 and 1959:01-2002:12, hence the initial window contains $n = 324$ months (before truncation due to lagging), and ends with $n = 528$ months for a total of 204 windows. We then fix the window to 324 months, a sample size that corresponds to Stock and Watson's (1989) seminal study. In this case, the initial sample period is 1959:01 - 1985:12 and the final period is 1971:11-2002:12, generating 205 windows.

Due to the large volume of tests required, we perform tests rather mechanically. VAR models of differences and levels (with excess lags) are employed, and VAR orders are selected by minimizing the AIC over $p = 1...18$. For the excess-lag models we add one lag to the optimally selected order in lieu of evidence that the largest order of integration is one in any window. Although we collect residual white-noise test p-values for each window, the information is not used for model selection. We perform both standard and bootstrap tests of non-causality for each window for each VAR model in differences and levels, and keep a running count of rejections of the various non-causality hypotheses. Tests of $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ are performed at the 5%-level; all other tests are performed at the nominal 1%-level.

The criterion for detection of non-causation at all horizons ($\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$) is a failure to reject either of Tests 0.1 or 0.2. We reject at $h = 1$ if we reject $\Delta m1 \overset{1}{\nrightarrow}$

[19]Based on the AIC and Ljung-Box tests, the optimal VAR order for the compete vector process $(\Delta y, \Delta m1, \Delta u, \Delta m2, \Delta o, rr)$ is $p = 8$. In order to improve comparability with the above tests on the truncated system $(\Delta y, \Delta m1, \Delta m2)$, we opted for a VAR(12) excess-lag model in levels.

Δy; we reject $\Delta m1 \overset{(2)}{\nrightarrow} \Delta y$ if we fail to reject $\Delta m1 \overset{1}{\nrightarrow} \Delta y$, reject both Tests 1.1 ($\Delta m1 \overset{1}{\nrightarrow} Z$) and 1.2 ($Z \overset{1}{\nrightarrow} \Delta y$), and reject Test 2.0 ($\Delta m1 \overset{1}{\nrightarrow} \Delta y$, $\pi_{\Delta y,Z,1} = 0$); and so on.

For a particular window we do not allow for rejection at multiple horizons: if we reject $\Delta m1 \overset{(h)}{\nrightarrow} \Delta y$ we stop the test procedure. In this sense, our analysis concerns the earliest horizon at which causation takes place. We do, however, allow for simultaneous detection of non-causation at all horizons $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ and causation at some horizon, $\Delta m1 \overset{h}{\rightarrow} \Delta y$: we employ the tests of non-causation at all horizons $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ separately from the remaining horizon-specific tests, and do not force the tests of non-causation at $h \geq 1$ to be contingent on the results of tests of $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$. We present window frequencies in which the two sets of tests contradict each other (i.e. detect $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ and $\Delta m1 \overset{h}{\rightarrow} \Delta y$). Horizon specific causality frequencies can be found in Table 2 for both increasing and fixed window length, and models of differences and levels[20].

First Differences, Increasing Windows

For VAR systems with the unemployment rate, sequential tests based on the parametric bootstrap detect non-causation $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ in only 2.94% of all windows; causation 1-month or 2-months ahead is never detected; and causation 3 and 4 months ahead are detected in roughly 45% and 13% of all sample periods. In under 1% of all sample windows do we detect both non-causation at all horizons $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ and causation at some horizon $\Delta m1 \overset{h}{\rightarrow} \Delta y$. Thus, there exists an unambiguous tendency for growth in the money supply to anticipate real income growth after a discreet delay of 2-3 months as the unemployment rate adjusts. This both corroborates and strengthens evidence for causation at $h = 4$ within the complete sample period 1959-2002, cf. Section 5.1.1.

For VAR systems with M2 test evidence suggests both non-causation in all periods $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ (over 96% of all periods), or causation 1-2 months ahead (23%-37% of all periods). Use of M2 generates extensive conflicting evidence of non-causation and causation: in roughly 58% of all sample periods evidence exists both in favor of non-causation at all horizons $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ and causation at some horizon tested $\Delta m1 \overset{h}{\rightarrow} \Delta y$. Considering that nearly all evidence for causation takes place at horizons 1 or 2 for a combined 59.8% of all windows, we can infer that whenever $\Delta m1 \overset{1}{\rightarrow} \Delta y$ or $\Delta m1 \overset{2}{\rightarrow} \Delta y$ was detected, so was $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$. Such highly ambiguous evidence suggests extreme caution should be applied when interpreting tests of 1-step ahead non-causation in related money-income models with M2 (e.g. Boudjellaba *et al*, 1992,1994; Amato and Swanson, 2001).

First Differences, Fixed Windows

In this context, evidence for causal delays expands in several notable directions. A system with the unemployment rate provides evidence of causation 3-4 months ahead, with a substantial increase in the number of windows suggesting causation exactly 4-months ahead $\Delta m1 \overset{4}{\rightarrow} \Delta y$: allowing the sample period to increase (and thereby allowing the system to evolve toward a steady-state) suggests causation

[20]Consult Hill (2004a: Figures 1-12) for graphic representations of window-specific causality counts, as well as model selection accuracy.

4-months ahead occurs in only 12.75% of all windows; fixing sample periods to 324 months (and thereby allowing period-specific non-stationarity) generates evidence for causation at the same horizon in 55.88% of all sample periods. This pattern extends to the price of oil and the rate spread.

Except for the system with M2, the most prominent characteristic is the significant increase in the number of windows providing any evidence of causation. Causation takes place between 1-5 months ahead through the unemployment rate in 93.63% of all sample periods, compared to 57.85% when sample periods increase in length. Similar to the increasing window case, VAR models with the unemployment rate lead to a negligible frequency of contradictory test results.

Levels with Excess Lags, Increasing Windows

Once cointegration is controlled for, a vastly different causation picture emerges. In over 53% and 40% of all sample periods for models with the unemployment rate and the interest rate spread, respectively, money causes income 1-month ahead. Indeed, when M2 is included, in 100% of all windows direct causation from money to income is detected (i.e. 1-month ahead), strongly supporting Swanson (1998). Notice, however, that except for the model with M2, tests of non-causation at all horizons and at specific horizons are in substantial conflict.

Levels with Excess Lags, Fixed Windows

Using rolling windows of fixed 324-month length and controlling for cointegration, evidence now strongly points toward causality exactly one-month ahead (unemployment, M2), or causality 1-2 months ahead (oil). Similar to increasing windows with levels, inclusion of M2 (unemployment) points to causation 1-month ahead in 95.61% (90%) of all windows. In this case, only the model with M2 leads to a negligible frequency of contradictory test results (under 10% of all windows).

Final Remarks

In general, VAR systems of first differences suggest a clear evolutionary trend exists over the last five decades. For fixed window models with unemployment or M2, evidence for causation at any horizon begins roughly in the periods ending in mid-1992 (window 80), the recovery period of the recession of early 1990's: see Hill (2005a: Figures 1 and 4). The most prominent characteristic for all models of increasing or fixed windows as the sample extends into the 1990's is evidence for white-noise in the residuals series: a pronounced spike across all models occurs in the period ending in Sept.-Nov. 2001 (roughly windows 190-192), suggesting samples containing the social events surrounding the third and fourth quarters of 2001 may contain a substantial outlier. It is interesting to note that models with either the unemployment rate or M2 generate essentially bimodal Q-test p-value spikes, one occurring near the end of 1992 and the other in the end of 2001, both coinciding with periods of recovery following economic down turns and periods surrounding events associated with war.

VAR specification significance reaches extreme levels of significance in the latter sample periods (beginning in the 1990's) of the fixed 324-month rolling window framework: this can be clearly seen in Hill (2005: Figures 9-12). This suggests the excess-lag model in levels for these latter sample periods reasonably captures the dynamic traits of the included macroeconomic processes. Moreover, this points out the complexity of the model specification and estimation issues surrounding money-income systems, and surrounding Stock and Watson's (1989) seminal investigation and Friedman and Kuttner's (1993) influential follow-up. The results of those

popular studies should be seen as initial steps toward a complete investigation of sample period dynamics and common stochastic trending prevalent in many macroeconomic systems.

6. Conclusion

In the present study we develop a simple VAR parametric recursion that, for trivariate processes with one scalar auxiliary variable, always allows for sequential linear parametric conditions for non-causality up to horizon $h \geq 1$. We prove causation must eventually occur from Y to X when the auxiliary variable Z is univariate, and provide a special case when a simple linear parametric condition for noncausation up to arbitrary horizons is identical in VARs with univariate or multivariate auxiliary variables.

An empirical analysis of the money-income relationship reveals significant evidence in favor of linear causation from money to income, either directly when we control for cointegration, or indirectly in models of first differences. Multiple horizon tests of non-causation over rolling windows provides evidence of either causal pattern evolution or substantial sample period non-stationarity. Sample windows incorporating recent information (starting from about 1990), in particular for fixed windows which remove information from the 1960's, produce extreme spikes in model significance. This suggests causal evidence from these periods is arguably the most reliable, although there is not a pronounced pattern of causality or non-causality across all models (i.e. differences, levels, Z) in these latter periods.

Appendix 1

Table 1

Auxiliary: Unemployment Rate					
		Differences		Levels	
Test #	Hypothesis	p-value[a]	p-boot[b]	p-value	p-boot
Test 0.1	$\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$	.0373	.080	.0000	.000
Test 0.2	$\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$	.0086	.040	.0012	.216
Test 1.0	$\Delta m1 \overset{1}{\nrightarrow} \Delta y$	.5693	.626	.0070	.148
Test 1.1	$\Delta m1 \overset{1}{\nrightarrow} \Delta u$	.0096	.000	.0000	.000
Test 1.2	$\Delta u \overset{1}{\nrightarrow} \Delta y$	.0055	.026	.4264	.376
Test 2.0	$\Delta m1 \overset{(2)}{\nrightarrow} \Delta y$	.3182	.450	.0000	.024
Test 3.0	$\Delta m1 \overset{(3)}{\nrightarrow} \Delta y$	.0194	.092	.0000	.012
Test 4.0	$\Delta m1 \overset{(4)}{\nrightarrow} \Delta y$	.0030	.032	.0000	.020
Test 5.0	$\Delta m1 \overset{(5)}{\nrightarrow} \Delta y$	.0024	.032[c]	.0000	.008[d]
Min. AIC VAR Order p		8		10	
Ljung-Box p-value		.045		.0089	

Auxiliary: M2					
		Differences		Levels	
Test #	Hypothesis	p-value	p-boot	p-value	p-boot
Test 0.1	$\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$	.0000	.000	.0000	.000
Test 0.2	$\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$	.0093	.198	.0000	.002
Test 1.0	$\Delta m1 \overset{1}{\nrightarrow} \Delta y$	.4938	.466	.0000	.008
Test 1.1	$\Delta m1 \overset{1}{\nrightarrow} \Delta o$	.0000	.000	.0000	.000
Test 1.2	$\Delta o \overset{1}{\nrightarrow} \Delta y$	.0383	.244	.0000	.000
Test 2.0	$\Delta m1 \overset{(2)}{\nrightarrow} \Delta y$	.0092	.082	.0000	.000
Test 3.0	$\Delta m1 \overset{(3)}{\nrightarrow} \Delta y$	.0170	.126	.0000	.000
Test 4.0	$\Delta m1 \overset{(4)}{\nrightarrow} \Delta y$	.0088	.128	.0000	.000
Test 5.0	$\Delta m1 \overset{(5)}{\nrightarrow} \Delta y$	.0022	.066	.0000	.000
...		....	....	....	....
Test 11.0	$\Delta m1 \overset{(11)}{\nrightarrow} \Delta y$	.0000	.006[e]	.0000	.000[f]
Min. AIC VAR Order p		10		12	
Ljung-Box p-value		.370		.183	

Notes: a. p-values based on the chi-squared distribution; b. p-values based on parametric bootstrap.

c. Reject $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ at 10%-level, and reject $\Delta m1 \overset{(4)}{\nrightarrow} \Delta y$ at bounded 13%-level.

d. Fail to reject $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ at 10%-level; or reject $\Delta m1 \overset{(2)}{\nrightarrow} \Delta y$ at bounded 5%-level.

e. Fail to reject $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ at 10%-level; or reject $\Delta m1 \overset{(11)}{\nrightarrow} \Delta y$ at bounded 11%-level

f. Reject $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ at 1%-level, and $\Delta m1 \overset{1}{\nrightarrow} \Delta y$ at 1%-level.

Table 1 - Cont.

Auxiliary: Oil Price					
		Differences		Levels	
Test #	Hypothesis	p-value	p-boot	p-value	p-boot
Test 0.1	$\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$	.7900	.856	.0021	.020
Test 0.2	$\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$	.0524	.560	.0000	.106
Test 1.0	$\Delta m1 \overset{1}{\nrightarrow} \Delta y$	.9590	.954	.0038	.026
Test 1.1	$\Delta m1 \overset{1}{\nrightarrow} \Delta o$	.3780	.000	.0379	.000
Test 1.2	$\Delta o \overset{1}{\nrightarrow} \Delta y$	.1925	.194	.3686	.516
Test 2.0	$\Delta m1 \overset{(2)}{\nrightarrow} \Delta y$	.9343	.890	.6230	.670
...		...	...	...	...
Test 5.0	$\Delta m1 \overset{(5)}{\nrightarrow} \Delta y$	.5664	.734[a]	.8361	.850[b]
Min. AIC VAR Order p		4		6	
Ljung-Box p-value		.004		.0014	
Auxiliary: Rate Spread					
		Differences		Levels	
Test #	Hypothesis	p-value	p-boot	p-value	p-boot
Test 0.1	$\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$	.7715	.718	.0000	.000
Test 0.2	$\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$	.0081	.166	.0000	.020
Test 1.0	$\Delta m1 \overset{1}{\nrightarrow} \Delta y$	.9769	.964	.0000	.032
Test 1.1	$\Delta m1 \overset{1}{\nrightarrow} rr$	.3240	.000	.0000	.000
Test 1.2	$rr \overset{1}{\nrightarrow} \Delta y$	.0009	.034	.8579	.526
Test 2.0	$\Delta m1 \overset{(2)}{\nrightarrow} \Delta y$	.9043	.912	.0847	.402
Test 3.0	$\Delta m1 \overset{(3)}{\nrightarrow} \Delta y$	.7937	.822	.1264	.396
Test 4.0	$\Delta m1 \overset{(4)}{\nrightarrow} \Delta y$	.8208	.892	.1012	.508
Test 5.0	$\Delta m1 \overset{(5)}{\nrightarrow} \Delta y$	.6289	.710[c]	.1145	.526[d]
Min. AIC VAR Order p		6		8	
Ljung-Box p-value		.009		.002	

Notes: a. Fail to reject $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ at 10%-level; or fail to reject $\Delta m1 \overset{(5)}{\nrightarrow} \Delta y$.

b. Reject $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ at 10%-level, and reject $\Delta m1 \overset{1}{\nrightarrow} \Delta y$ at 5%-level, or fail to reject $\Delta m1 \overset{(5)}{\nrightarrow} \Delta y$ at bounded 5%-level

c. Fail to reject $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ at 10%-level; or fail to reject $\Delta m1 \overset{(5)}{\nrightarrow} \Delta y$.

d. Reject $\Delta m1 \overset{(\infty)}{\nrightarrow} \Delta y$ at 5%-level, and reject $\Delta m1 \overset{1}{\nrightarrow} \Delta y$ at 5%-level or fail to reject $\Delta m1 \overset{(5)}{\nrightarrow} \Delta y$ at bounded 5%-level.

Table 2

Horizon Rejection Frequencies: First Differences									
Increasing Width Rolling Windows					Fixed Width Rolling Windows				
Horizon	u	o	rr	$m2$	Horizon	u	o	rr	$m2$
0^a	.2647	.6667	.7281	.8775	0	.5245	.9510	.8137	.9069
	[.0294]^b	[.6373]	[.7647]	[.9608]		[.0294]	[.2402]	[.8725]	[.9706]
1	.0000	.0000	.0000	.2255	1	.3872	.1176	.1520	.1324
	[.0000]	[.0000]	[.0000]	[.2304]		[.1716]	[.0490]	[.1569]	[.1275]
2	.2990	.2353	.0147	.3725	2	.0000	.3775	.1521	.0931
	[.0000]	[.1569]	[.0000]	[.3676]		[.0049]	[.3824]	[.1078]	[.1029]
3	.3137	.0000	.0733	.0000	3	.2108	.0000	.0343	.0196
	[.4510]	[.0147]	[.0392]	[.0000]		[.2010]	[.0000]	[.0343]	[.0049]
4	.1422	.0000	.0000	.0000	4	.3725	.00000	.0000	.0245
	[.1275]	[.0000]	[.0000]	[.0000]		[.5588]	[.0000]	[.0000]	[.0049]
5	.0000	.0000	.1618	.0049	5	.0000	.0000	.0000	.0833
	[.0049]	[.0000]	[.0000]	[.0049]		[.0049]	[.0000]	[.0000]	[.0000]
$\geq 1^c$	.7549	.2353	.2498	.6029	≥ 1	.9705	.4951	.3384	.2696
	[.5785]	[.1716]	[.0392]	[.5980]		[.9363]	[.4314]	[.2990]	[.2402]
$0, \geq 1^d$	.1324	.0392	.2500	.4804	$0, \geq 1$	.5245	.4559	.1666	.2598
	[.0098]	[.0294]	[.0392]	[.5735]		[.0294]	[.1078]	[.2304]	[.2157]

Notes: a. h = 0 denotes noncausation at all horizons: values are window frequencies for which we fail to reject H_0^∞;

b. Bracketed values denote window frequencies based on bootstrapped p-values;

c. Window frequencies for causation at any horizon $h \geq 1$;

d. Window frequencies for noncausation at all horizons, h = 0, and causation at some horizon $h \geq 1$.

Table 2 - Cont.

Horizon Rejection Frequencies: Levels with Excess Lags									
Increasing Width Rolling Windows					Fixed Width Rolling Windows				
Horizon	u	o	rr	$m2$	Horizon	u	o	rr	$m2$
0	.0000	1.000	1.000	.00000	0	.2634	.2488	1.000	.0439
	[.0000]	[1.000]	[1.000]	[.0000]		[.4098]	[.3658]	[1.000]	[.0976]
1	.6049	.0293	.4634	1.000	1	.9122	.4049	.4195	.9561
	[.5317]	[.0585]	[.4049]	[1.000]		[.9024]	[.6146]	[.4585]	[.9561]
2	.3171	.6341	.0195	.0000	2	.0146	.4537	.1902	.0439
	[.2341]	[.6683]	[.0488]	[.0000]		[.0244]	[.1707]	[.0976]	[.0390]
3	.0000	.0244	.0000	.0000	3	.0049	.0732	.0000	.0000
	[.0098]	[.0000]	[.0000]	[.0000]		[.0439]	[.0000]	[.0049]	[.0000]
4	.0000	.0000	.0000	.0000	4	.0000	.0000	.0244	.0000
	[.0000]	[.0000]	[.0000]	[.0000]		[.0000]	[.0000]	[.0098]	[.0000]
5	.0000	.0000	.0000	.0000	5	.0390	.0000	.0000	.0000
	[.1756]	[.0000]	[.0000]	[.0000]		[.0000]	[.0000]	[.0000]	[.0000]
≥ 1	.9220	.6878	.4829	1.000	≥ 1	.9707	.9318	.6341	1.000
	[.9512]	[.7268]	[.4537]	[1.000]		[.9707]	[.7853]	[.5708]	[.9951]
$0, \geq 1$	.9220	.6878	.4878	.0000	$0, \geq 1$	.2341	.2488	.6341	.0439
	[.9512]	[.7268]	[.4537]	[.0000]		[.3805]	[.2927]	[.5707]	[.0927]

Figure 1: Z = ue (diff., inc.)

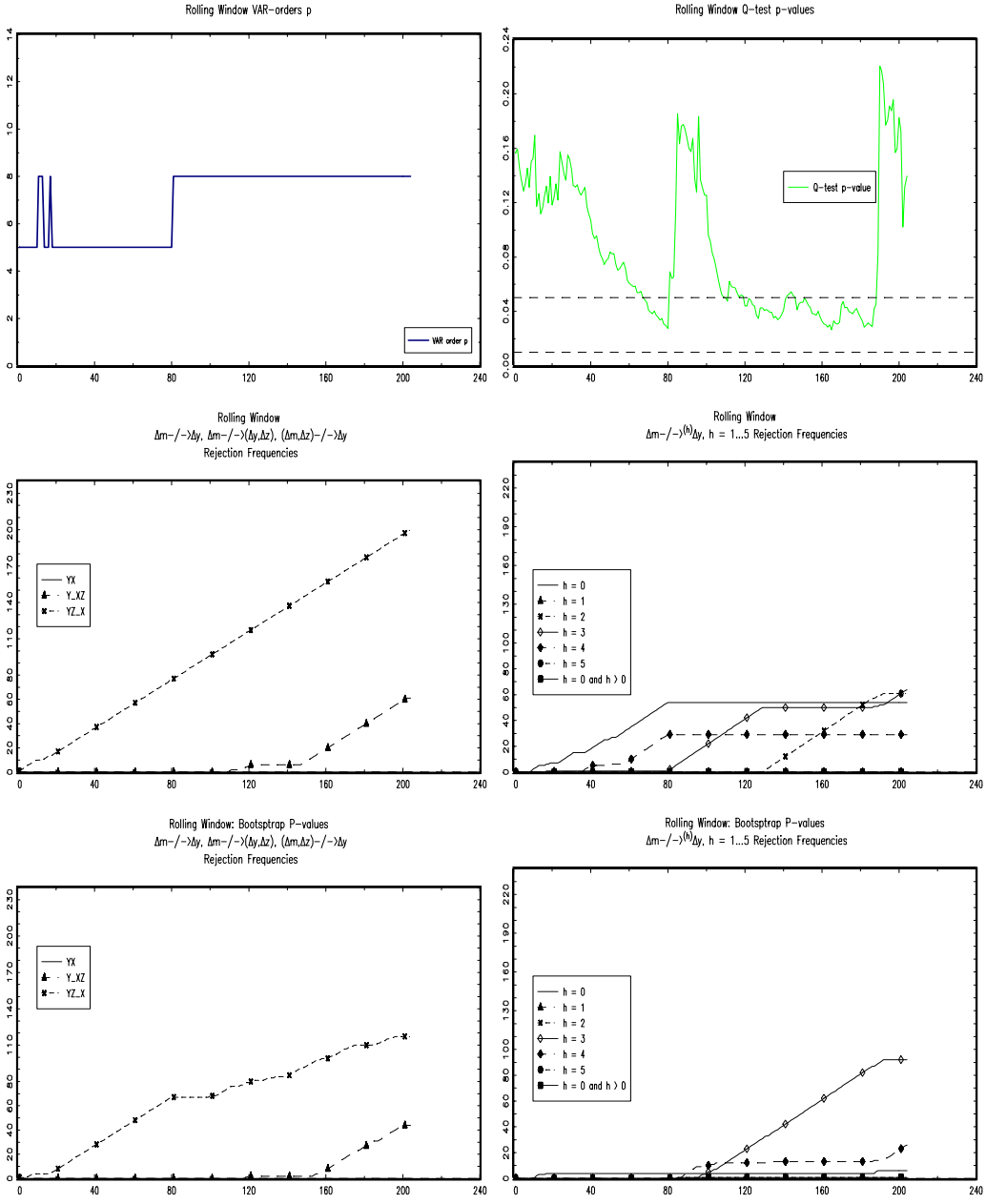


Figure 2: Z = rr (diff., inc.)

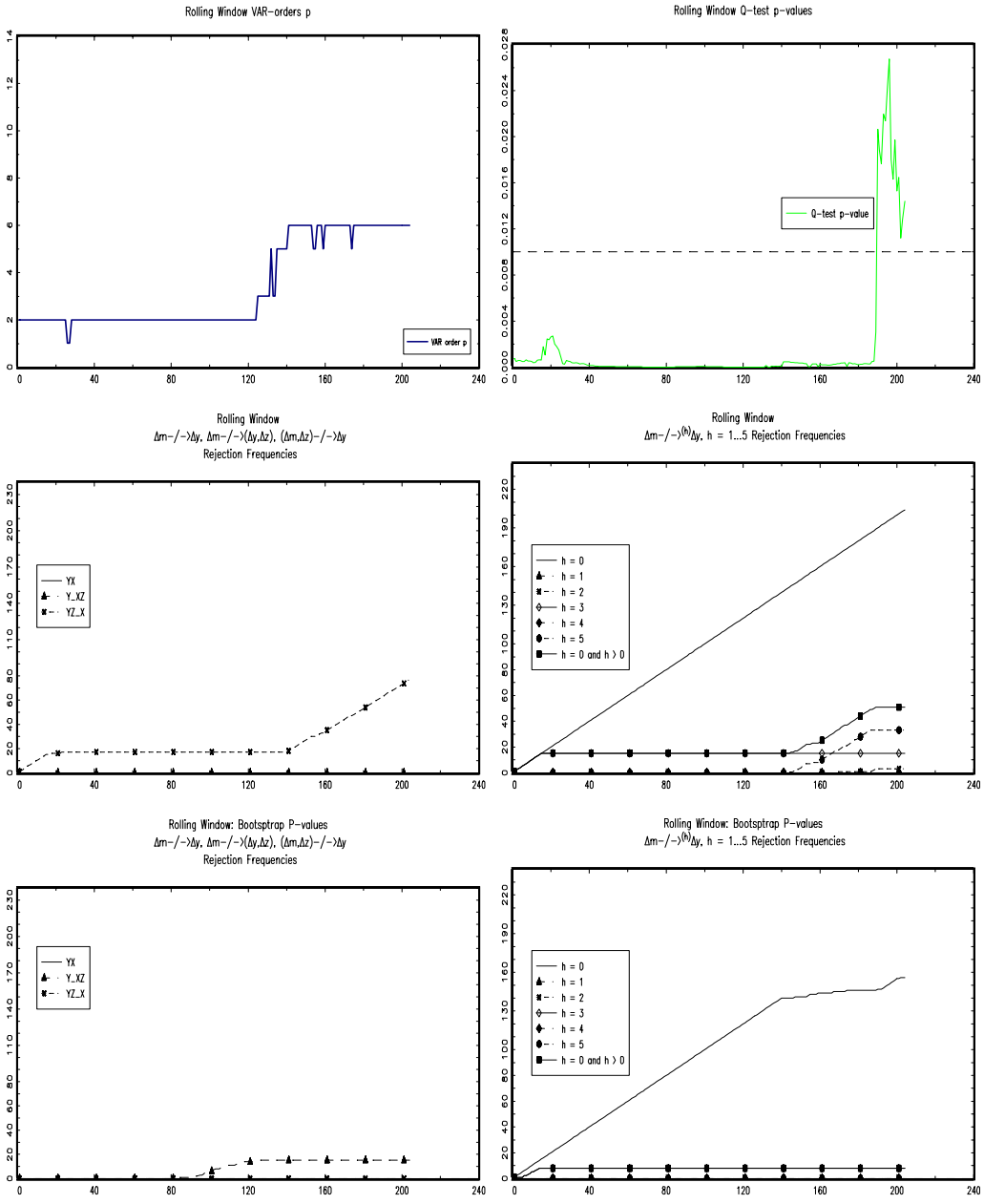


Figure 3: Z = o (diff., inc.)

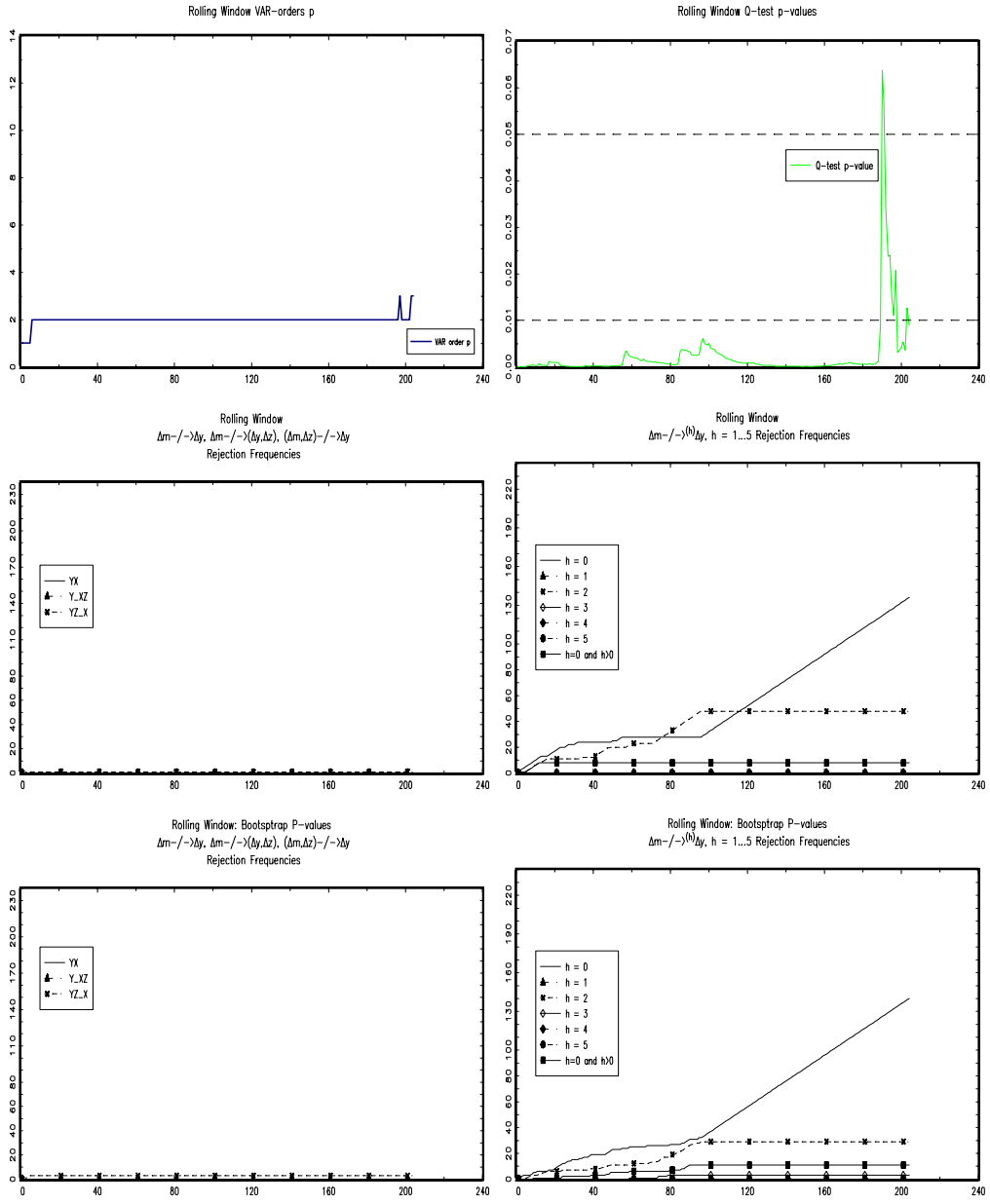


Figure 4: Z = m2 (diff., inc.)

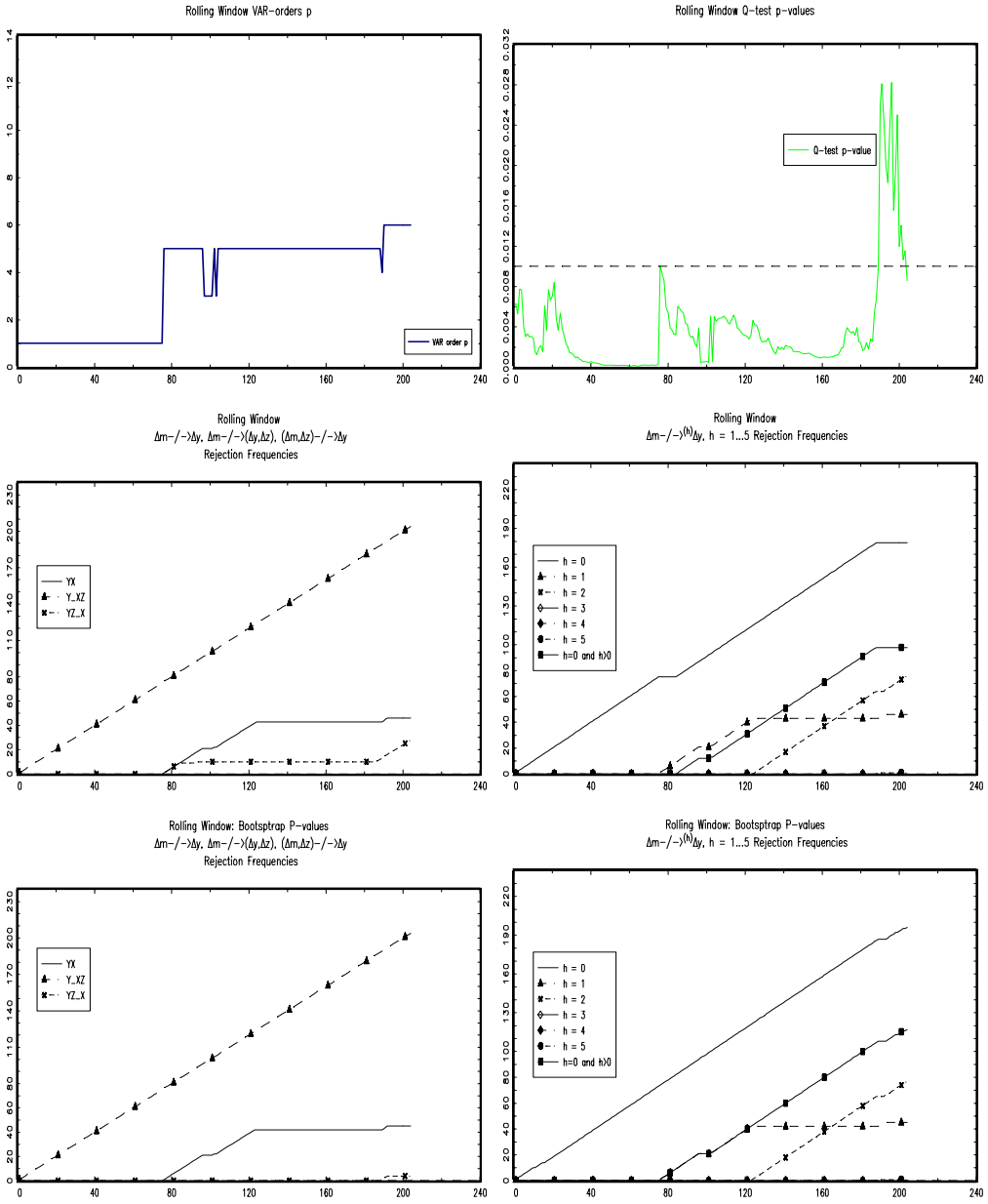


Figure 5: Z = ue (levels, inc.)

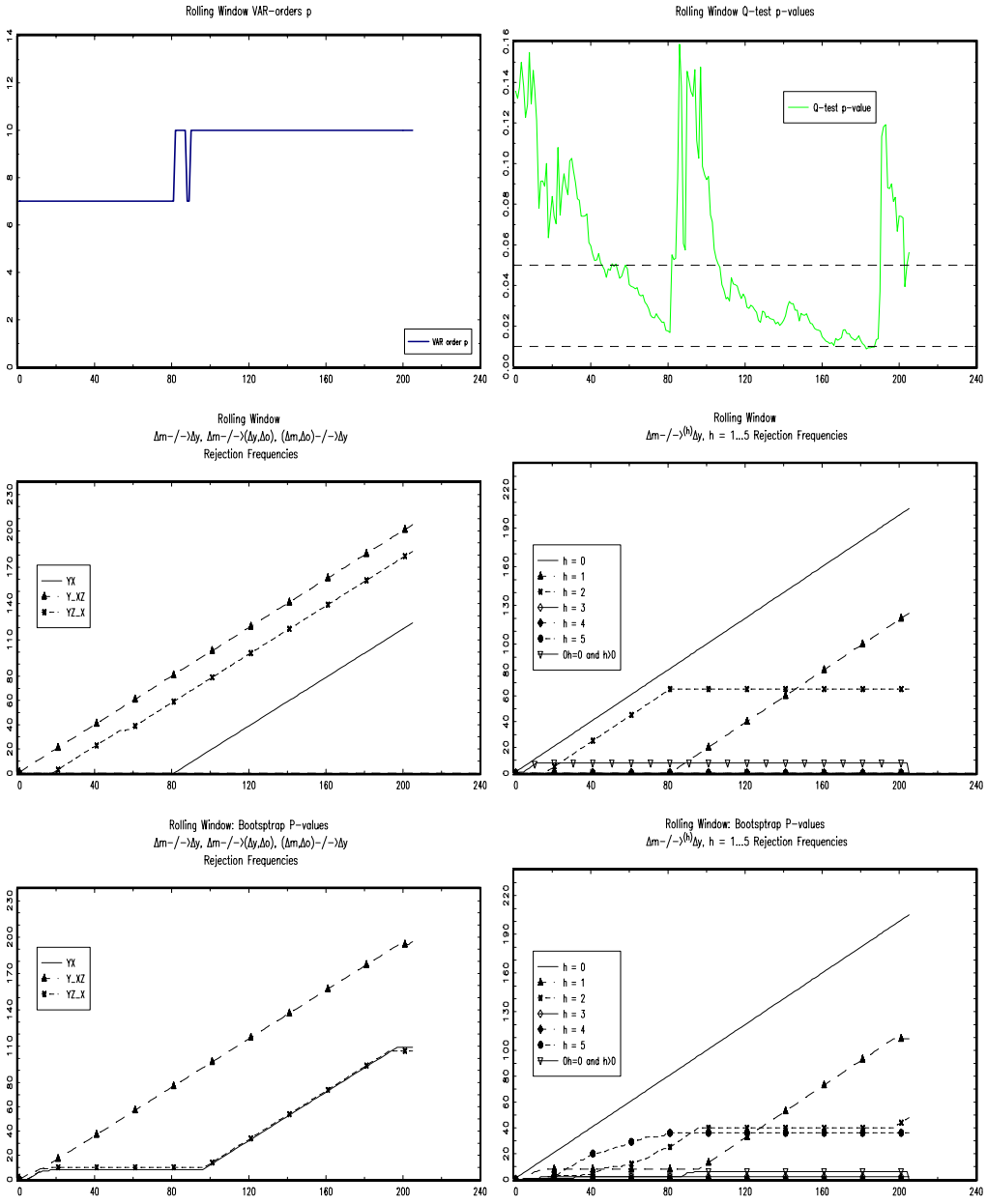


Figure 6: Z = rr (levels, inc.)

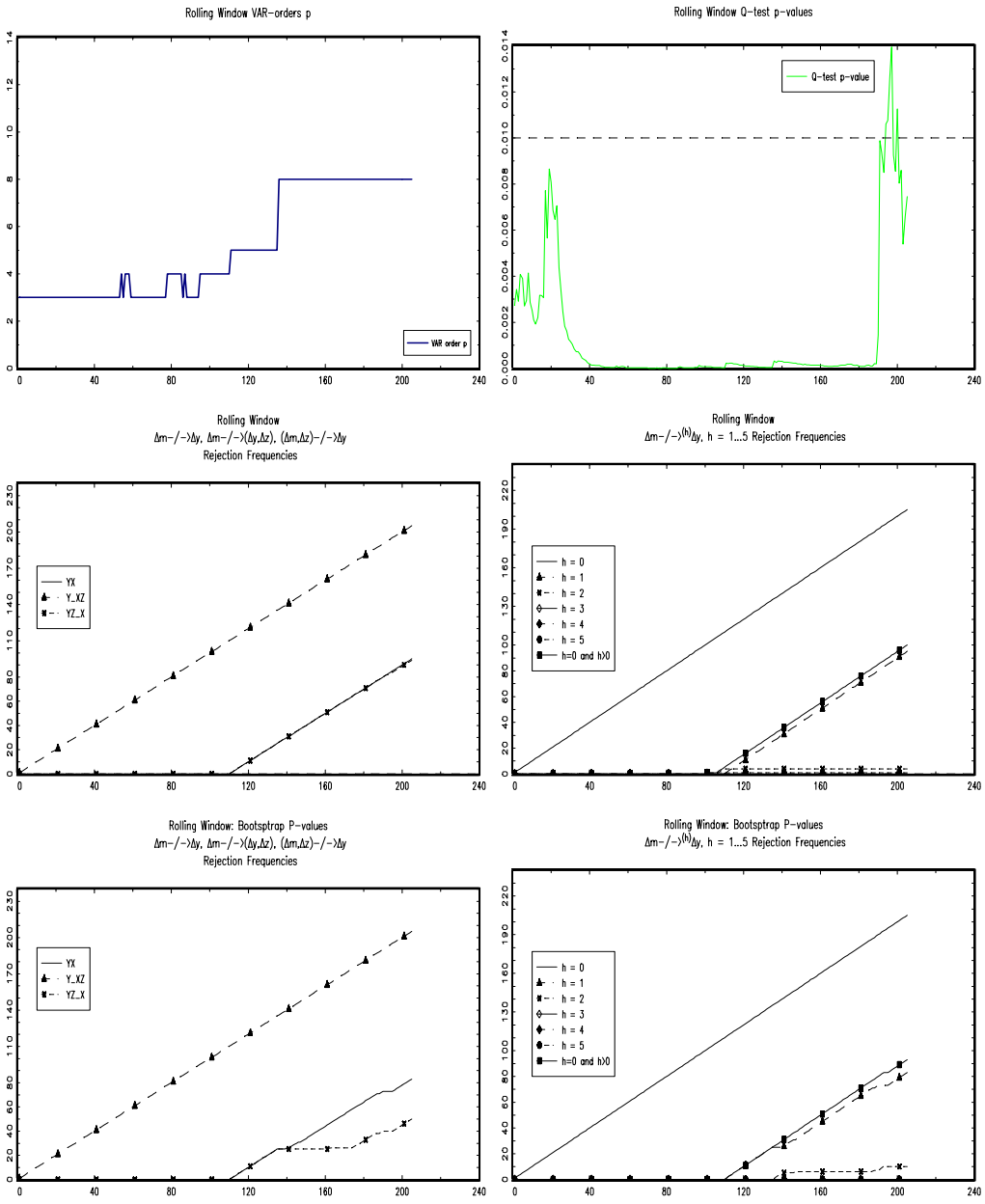


Figure 7: Z = o (levels, inc.)

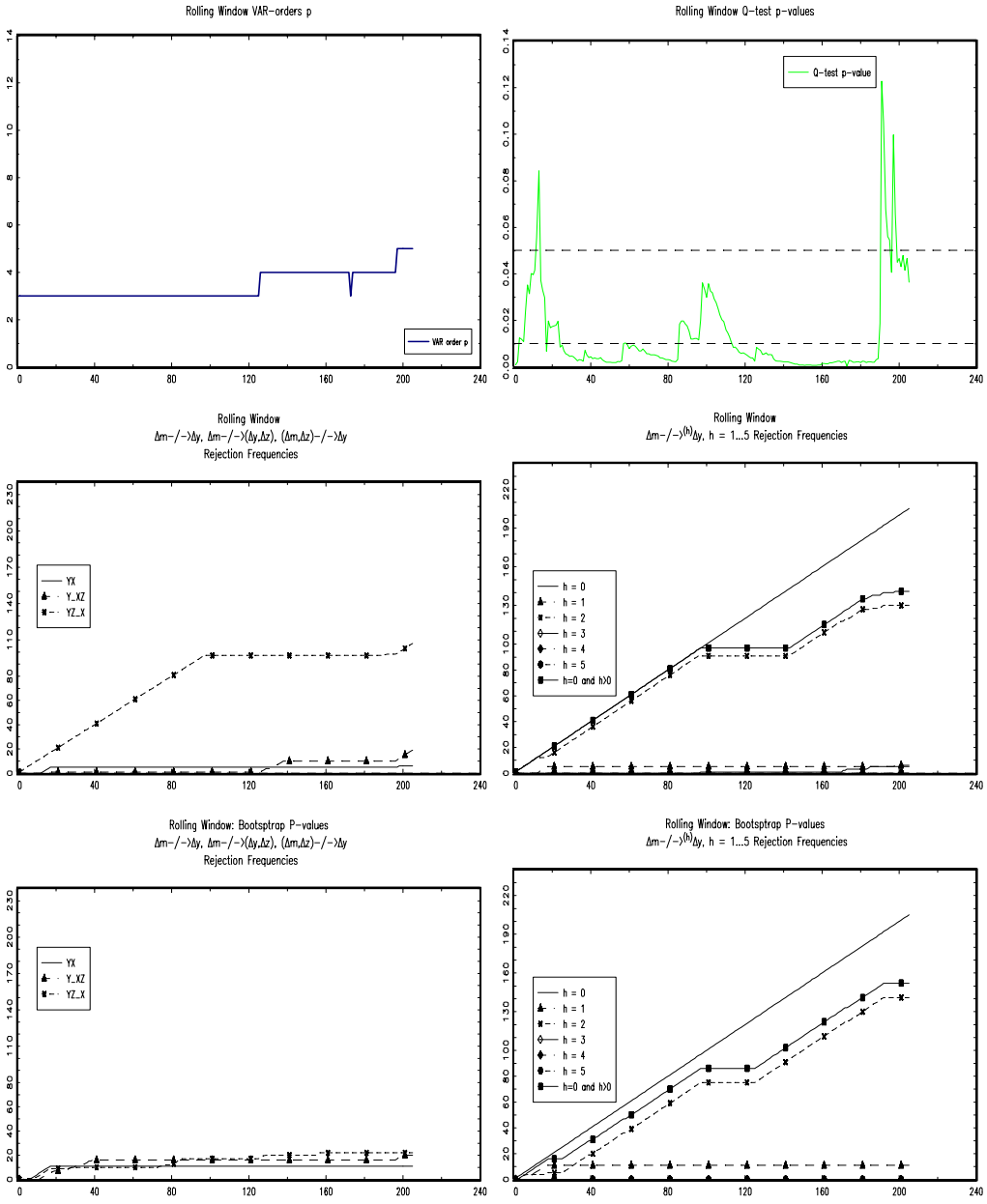


Figure 8: Z = m2 (levels, inc.)

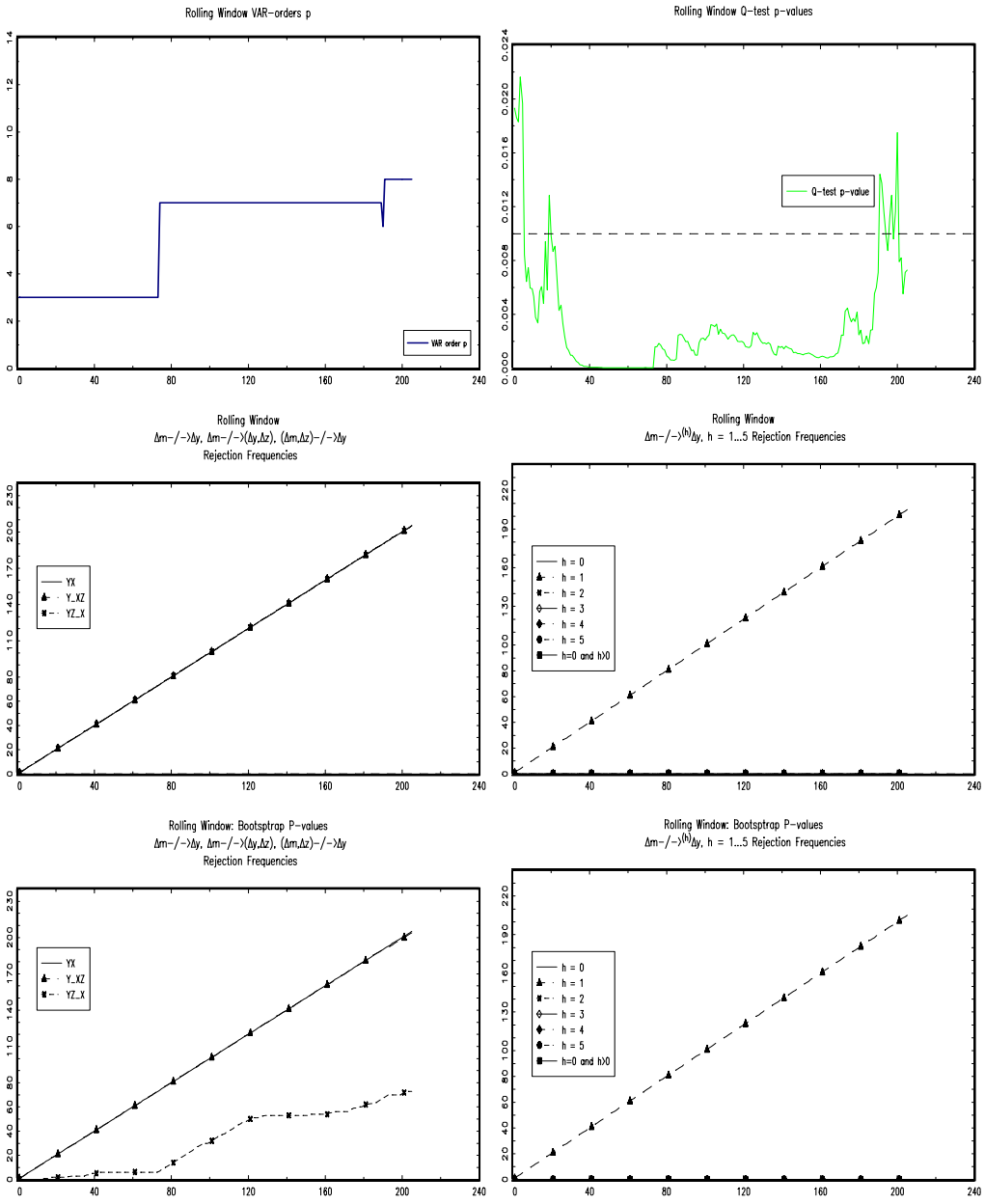


Figure 9: Z = ue (levels, fix)

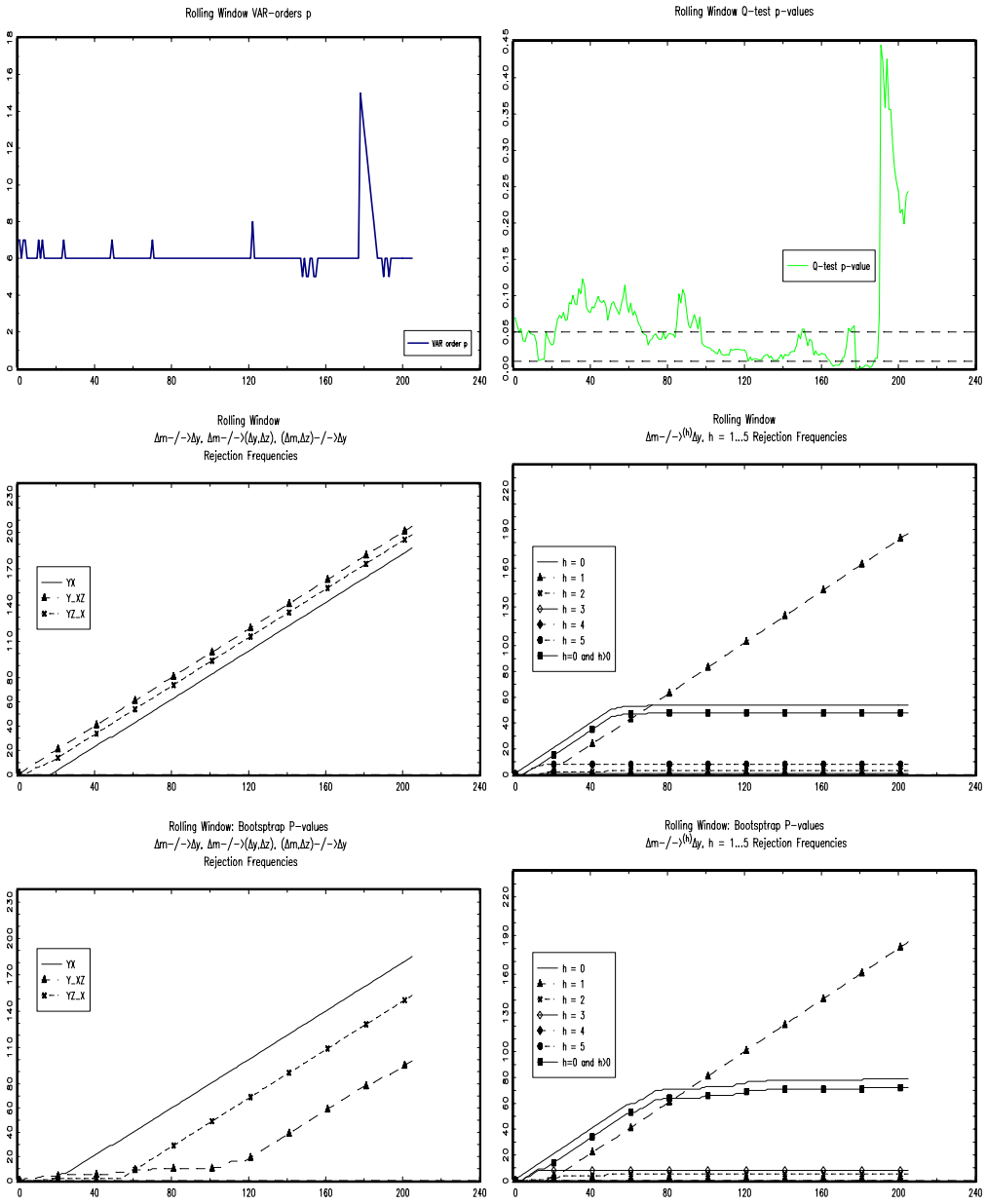


Figure 10: Z = rr (levels, fix)

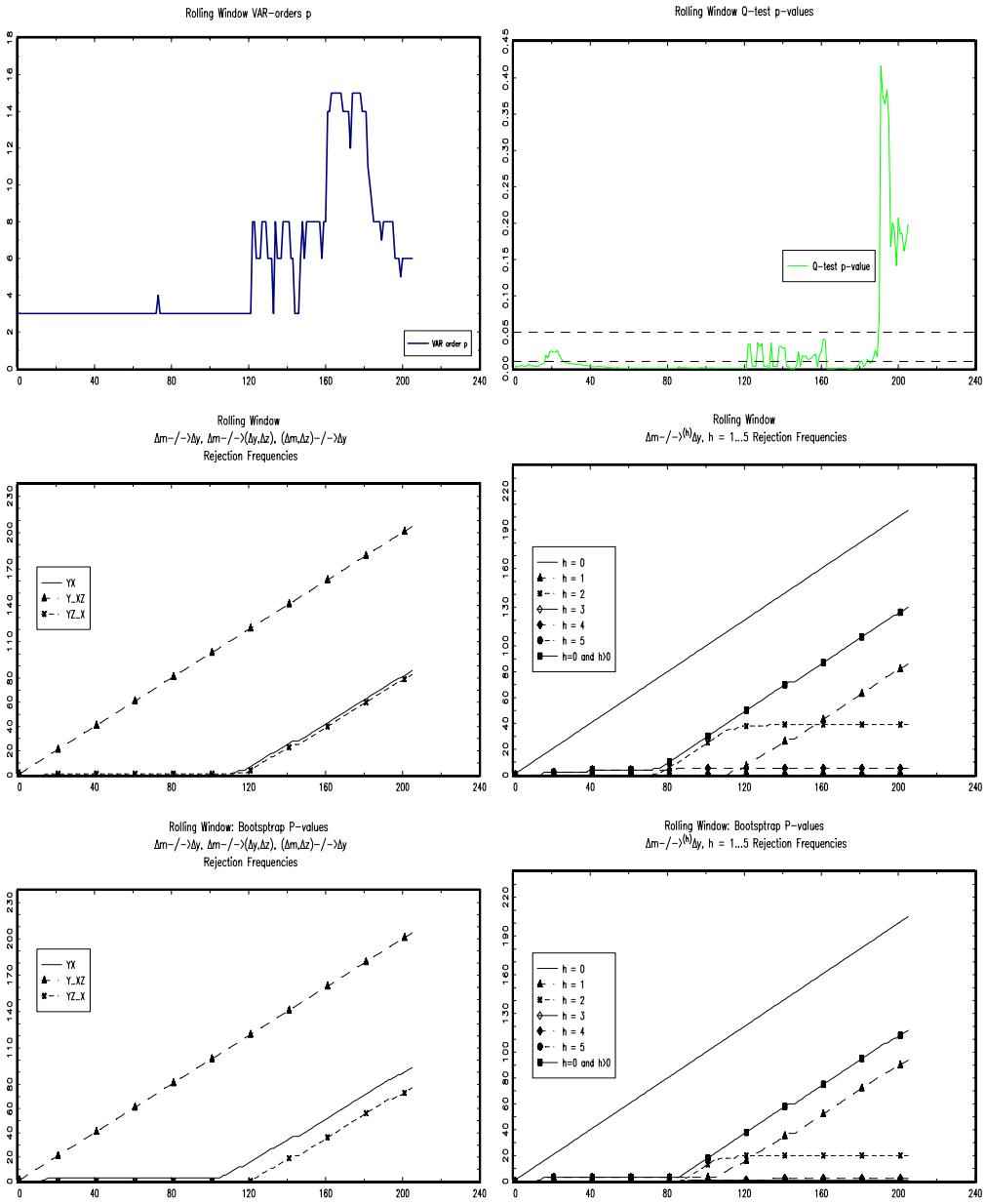


Figure 11: Z = o (levels, fix)

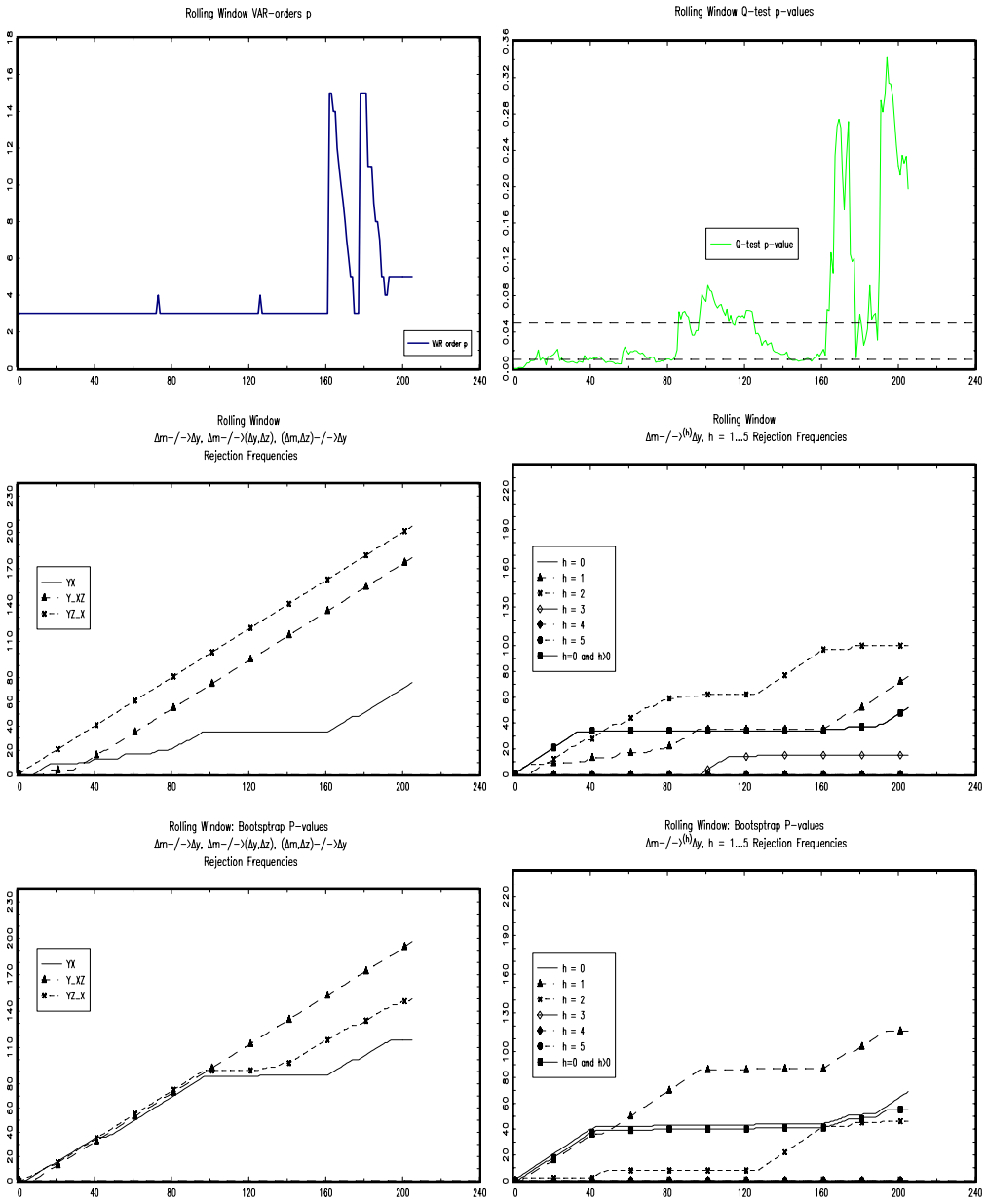
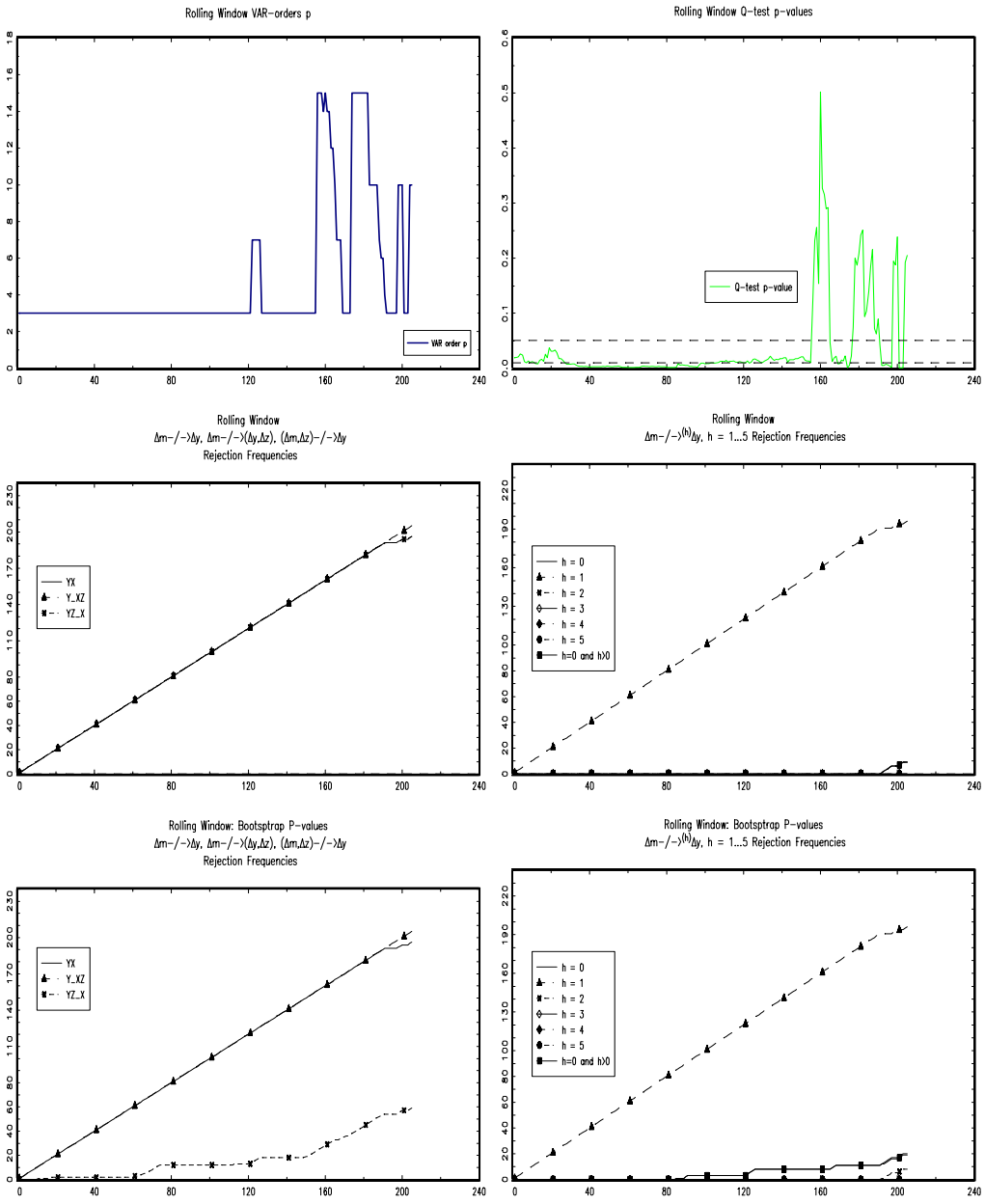


Figure 12: Z = m2 (levels, fix)



Appendix 2: Simulation Study

In order to study the performance of the above test procedure, we employ a controlled experiment for derivation of empirical test size and power for various VAR and VARMA processes. We employ Wald tests analyzed by p-values derived both by the asymptotic distribution and by a parametric bootstrap method.

5.1 Set Up

For our study, we generated VAR(6) and vector MA(1) processes under the null of non-causation at all horizons, and under alternatives of causation at horizons $h = 1, 2$, and 3. In all cases, $m_x = m_y = m_z = 1$ such that $m = 3$, sample sizes are restricted to $T \in \{100,\ 200, 300, 400, 500\}$ and 1000 series are generated for each test.

VAR(6) Construction and Hypotheses

For the VAR(6) process we simulate $W_t = \sum_{i=1}^{6} \pi_i W_{t-i} + \epsilon_t$, where ϵ_t denotes an *iid* 3-vector with mutually independent components $\epsilon_t = (\epsilon_{x,t}, \epsilon_{y,t}, \epsilon_{zt,})'$ drawn from a standard normal distribution. The matrix coefficients π_i are generated as uniform *iid* random numbers from the cube $[-.5, .5]^3$: we use $\pi = (\pi_1, ..., \pi_6)$ only if the resulting characteristic polynomial $I_3 - \pi_1 z - ... - \pi_6 z^6$ has all roots outside the unit circle, ensuring stability.

During the simulation process we impose the following restrictions (or lack, thereof), depending upon the hypothesis to be tested:

$$H_0^{\infty} : \pi_{XY,i} = \pi_{XZ,i} = 0,\ i = 1...6$$

$$H_1^1 : \pi_{XY,i} \neq 0,\ i = 1...6$$

$$H_1^2 : \pi_{XY,i} = 0,\ i = 1...6,\ \pi_{ZY,i} \neq 0,\ \pi_{XZ,i} \neq 0,\ i = 1...6$$

$$H_1^3 : \pi_{XY,i} = 0,\ i = 1...6,\ \pi_{ZY,i} \neq 0,\ i = 1...6,\ \pi_{XZ,i} \neq 0,\ i = 2...6$$

Under H_0^{∞}, we deduce $Y \overset{1}{\nrightarrow} (X, Z)|I_{XZ}$, cf. Theorem 2.2, and therefore Y never causes X, $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$, cf. Theorem 2.1. Under H_1^1, Y causes X at horizon $h = 1$. Under H_1^2, non-causation $Y \overset{1}{\nrightarrow} X|I_{XZ}$, and causation $Y \overset{1}{\rightarrow} Z \overset{1}{\rightarrow} X$ are true, with $\pi_{XZ,1} \neq 0$, thus $Y \overset{2}{\rightarrow} X|I_{XZ}$ is true, cf. Theorem 3.2. Finally, under H_1^3, $Y \overset{1}{\nrightarrow} X|I_{XZ}$, $Y \overset{1}{\rightarrow} Z \overset{1}{\rightarrow} X$, $\pi_{XZ,1} = 0$ and $\pi_{XZ,2} \neq 0$, thus $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$ and $Y \overset{3}{\rightarrow} X|I_{XZ}$ are true, cf. Theorem 3.2.

VMA(1) Construction and Hypotheses

For the VMA(1) processes, we simulate $W_t = \theta\epsilon_{t-1} + \epsilon_t$ by drawing *iid* uniform numbers θ from the cube $[-.9, .9]^3$, retaining only those matrices θ with characteristic roots outside the unit circle, ensuring invertibility. We employ the following restrictions:

$$\begin{aligned} H_0 &: \theta_{XY} = \theta_{XZ} = 0 \\ H_1 &: \theta_{XY} = 0. \end{aligned} \tag{6.1}$$

In order to deduce that nature of multiple horizon non-causation for invertible VMA processes in VAR form, we require necessary and sufficient conditions for

VAR noncausality in terms of VARMA coefficients. Boudjellaba *et al* (1992) derive reasonably simple necessary and sufficient conditions for non-causality at horizon $h = 1$ for such processes. Consider the general VARMA(p, q) process in lag form

$$\Phi(L)W(t) = \Theta(L)\epsilon(1), \tag{6.2}$$

where $\Phi(L)$ and $\Theta(L)$ denote the associated p^{th} and q^{th}-order lag $m \times m$ matrix-polynomials

$$\Phi(L) = I_m - \sum_{i=1}^{p} \phi_i L^i, \quad \Theta(L) = I_m + \sum_{i=1}^{q} \theta_i L^i \tag{6.3}$$

It is assumed that the polynomials do not have common roots, and all roots lie outside the unit circle. By Theorem 1 of Boudjellaba *et al* (1992, 1994), non-causality from scalar W_i to scalar W_j exists *if and only if*

$$\det\left(\Phi_i(z), \Theta_{(j)}(z)\right) = 0, \ |z| < \delta, \tag{6.4}$$

for some $\delta > 0$, where $\Phi_i(z)$ denotes the i^{th} column of $\Phi(z)$ and $\Theta_{(j)}(z)$ denotes the matrix $\Theta(z)$ with the j^{th} column removed. In the 3-vector MA(1) case with $W = (W_1, W_2, W_3)' = (X, Y, Z)'$, it follows that $\Phi(z) = I_m$, and $Y \overset{1}{\nrightarrow} X$ holds *if and only if*

$$\begin{aligned} & \det\left(\Phi_2(z), \Theta_{(1)}(z)\right) \\ = \ & \det\begin{pmatrix} 0 & \theta_{12}z & \theta_{13}z \\ 1 & 1+\theta_{22}z & \theta_{23}z \\ 0 & \theta_{32}z & 1+\theta_{33}z \end{pmatrix} \\ = \ & \theta_{13}\theta_{32}z^2 - \theta_{12}z - \theta_{12}\theta_{33}z^2 \\ = \ & 0, \ |z| < \delta. \end{aligned} \tag{6.5}$$

This occurs for *every* complex $|z| < \delta$, $\delta > 0$, *if and only if* $\theta_{12} = \theta_{13}\theta_{32} = 0$. Similarly, $Y \overset{1}{\nrightarrow} Z$ *if and only if* $\theta_{32} = \theta_{31}\theta_{12} = 0$, and $Z \overset{1}{\nrightarrow} X$ *if and only if* $\theta_{13} = \theta_{12}\theta_{23} = 0$. Consult Boudjellaba *et al* (1992, 1994), Dufour and Tessier (1993), and Dufour and Renault (1998) for further details on parametric conditions of non-causation at $h = 1$ for VARMA processes.

From the above details and Theorem 2.1, we deduce the hypothesis of non-causation at all horizons $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$ is true *if and only if* $\theta_{12} = \theta_{13}\theta_{32} = 0$ and either $\theta_{32} = \theta_{31}\theta_{12} = 0$ and/or $\theta_{13} = \theta_{12}\theta_{23} = 0$. Therefore, the VMA(1) coefficients in (14) under H_0 in fact satisfy $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$: the identity $\theta_{12} = \theta_{13} = 0$ (i.e. $\theta_{XY} = \theta_{XZ} = 0$) implies $Y \overset{1}{\nrightarrow} X$ and $Z \overset{1}{\nrightarrow} X$, therefore $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$.

It is interesting to point out in the 3-vector MA(1) case that either non-causation at all horizons $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$ or standard causation $Y \overset{1}{\rightarrow} X|I_{XZ}$ must be true, similar to the bivariate VAR case. Consider if non-causation is true $Y \overset{1}{\nrightarrow} X|I_{XZ}$, then either $\theta_{12} = \theta_{32} = 0$ and/or $\theta_{12} = \theta_{13} = 0$ must be true: in the former case $Y \overset{1}{\nrightarrow} Z$ follows, and in the latter case $Z \overset{1}{\nrightarrow} X$ follows. In either case, a causal chain does not exist, and Theorem 2.1 implies $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$. Therefore, for 3-vector MA(1) vector-processes, $Y \overset{1}{\nrightarrow} X|I_{XZ}$ *if and only if* $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$, which implies causation lags and causal neutralization are impossible. Thus, we deduce under H_1 in (6.1) that causation $Y \overset{1}{\rightarrow} X|I_{XZ}$ is true.

For each simulated series $\{W_t\}_{t=1}^n$ a minimum AIC method is employed for VAR order p selection, the VAR coefficients are estimated, and standard Wald tests are implemented for the linear compound hypotheses. All tests are performed at the 5%-level.

5.2 Parametric Bootstrap

There is ample evidence in the literature, however, that standard Wald tests in multivariate models tend to lead to over-rejection of null hypotheses. In order to analyze the problem, we employ a parametric bootstrap method for simulating the asymptotic p-value of each test statistic. In brief, the parametric bootstrap is performed as follows for an arbitrary hypothesis:

i. Obtain estimated VAR coefficients, $\hat{\pi} = (\hat{\pi}_1, ..., \hat{\pi}_p)$, where p minimizes the AIC;

ii. Derive the test statistic, denoted T_n;

iii. Simulate J series $W_{t,j}$, $j = 1...J$, $t = 1...n$, based on the the estimated parameters $\hat{\pi}$ with the null hypothesis restrictions imposed. For example, a test of $Y \overset{1}{\nrightarrow} X|I_{XZ}$ imposes $\pi_{XY,i} = 0$, hence the X,Y-block $\hat{\pi}_{XY}$ is replaced by zeros. The series are simulated as

$$W_{t,j} = \sum_{i=1}^{p} \hat{\pi}_i W_{t-1} + \epsilon_t$$

where ϵ_t is an *iid* 3-vector draw from a standard normal distribution.

iv. Use the double-array $\{W_{t,j}\}_{t,j=1}^{n,J}$ to generate J test statistics $T_{n,j}$ for the hypothesis in question;

v. The approximate p-value is simply the percent frequency of the event $T_{n,j} > T_n$.

For all tests, we set $J = 1000$. See Dufour (2002) for a proof of the asymptotic validity of the parametric bootstrap.

5.3 Simulation Results

Tables 3 and 4 below contain all simulation results. Columns in each table contain empirical rejection frequencies based on p-values derived from the asymptotic chi-squared distribution, and the empirical bootstrap method [in brackets]. Tests at horizon $h = 0$ are tests of noncausation at all horizons: we fail to reject $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$ for some series $\{W_t\}_{t=1}^n$ when we fail to reject $Y \overset{1}{\nrightarrow} X$, and fail to reject either $Y \overset{1}{\nrightarrow} Z$ and/or $Z \overset{1}{\nrightarrow} X$. For tests at individual horizons $h \geq 2$ we detect causation $Y \overset{h}{\rightarrow} X|I_{XZ}$ when we reject the compound hypothesis $Y \overset{1}{\nrightarrow} X$, $\pi_{XZ,i} = 0, i = 1...h-1$.

VAR Simulations

Consider the results for VAR processes based on p-values derived from the asymptotic distribution. For processes that satisfy $Y \overset{(\infty)}{\nrightarrow} X|I_{XZ}$ and for the sample size $n = 500$, rejection frequencies at horizons $h \geq 1$ are not far from the nominal level of 5% for tests of noncausation at all horizons: empirical sizes at $h \geq 1$ ranged from .066 to .076.

When causation occurs at horizons $h \geq 1$, tests rarely suggest noncausation at all horizons occur: evidence for noncausation at all horizons in such cases occurred in 7.2% or fewer of simulated series for $n \geq 300$, and for $n = 500$ in 2% or fewer of such series.

Moreover, when causation occurs exactly one-step ahead, rejection frequencies at $h \geq 1$ reach above 90% for sample sizes $n \geq 300$. For the same sample size range noncausation in all horizons is detected in fewer than 5% of all such series.

When a one-period causal delay exists such that $Y \overset{1}{\nrightarrow} X$ and $Y \overset{2}{\rightarrow} X$, again standard tests work reasonably well, generating empirical sizes at $h = 1$ near the 5%-level (.064 with $n = 500$), and producing reasonable empirical powers at subsequent horizons $h \geq 2$ (.812 with $n = 500$ at $h = 3$).

However, when causation occurs at horizon $h = 3$ (i.e. $Y \overset{(2)}{\nrightarrow} X$ and $Y \overset{3}{\rightarrow} X$), noticeable size distortions occur for tests at lower horizons 1 and 2. For such tests, empirical sizes approach .20 for nominal levels of 5% and $n \geq 300$. This implies we are more likely to detect causation at low horizons when in fact true causal delays are longer.

Bootstrapped p-values clearly provide better size approximations to the null distribution than standard p-values. However, even the bootstrapped p-values lead to over-rejections of the fundamental null of noncausality when non-causality occurs at all horizons: for sample sizes under 400, rejection rates reached 10% for tests at the 5%-level. Encouragingly, however, for sample sizes $n \geq 400$, rejection rates were very close to the nominal level.

When a causal lag exists, empirical sizes are again near the nominal level, however empirical powers are noticeably low. For example, with a sample size of 500 and a true causal lag of 2 periods such that $Y \overset{(2)}{\nrightarrow} X$ and $Y \overset{3}{\rightarrow} X$ are true, the bootstrap test detected causation at $h = 3$ in under 48% of simulated series.

For both asymptotic and bootstrap tests, however, empirical power diminishes severely as the horizon of causation increases. When causation occurs at $h = 1$, powers reach above 90% even for small n. However, when causation occurs at $h = 3$, powers drop to under 70% for the standard tests, and below 50% for bootstrap tests.

We argue that this evidence alone portrays a far more complicated picture of the relative merits of standard and bootstrap tests than typically argued in the literature. Neither method generates both competitive empirical sizes and powers in a benchmark Gaussian VAR environment in which model coefficients are randomly generated. Which method we favor in practice depends on whether we favor a conservative test with low power (bootstrap test), or a liberal test with excessive probability of a Type I error for some hypotheses (conventional test).

VMA Simulations

Next, consider test results for VMA(1) processes. For series in which $Y \overset{(\infty)}{\nrightarrow} X$ and for small sample sizes, standard asymptotic tests produce large empirical sizes for the fundamental tests of non-causation, in particular for tests at horizons $h \geq 2$. For $n \geq 400$, however, erroneous detection of causality dropped to frequencies of 5.1%-8.9% for tests of noncausation at horizons $h = 1...3$.

It is important to point out that for tests of noncausation at all horizons, in 95.8% (95.9%) of all series with $n = 400$ (500) did tests *correctly* conclude noncausation occurred at all horizons, which implies an the effective empirical size is 4.2% (4.1%) based on this fundamental hypothesis. Bootstrap tests, by comparison, generated empirical sizes near the nominal 5%-level for tests at $n \geq 300$, with extreme accuracy at $n = 500$.

When causation occurs one-period ahead ($Y \overset{1}{\rightarrow} X$), both tests work well when judged by whether they detect causation at all, although standard tests uniformly perform better. However, both tests struggle with tests of noncausation at all horizons: for a large sample size $n = 500$, standard (bootstrap) tests incorrectly detect $Y \overset{(\infty)}{\nrightarrow} X$ in 36.7% (32.8%) of all series

Noticeable lags exist before either test method leads to the correct detection of causation. In general, tests are sensitive to causation at $h \geq 2$, and comparatively weak at $h = 1$. For example, with $n = 400$ (500) standard tests correctly detect causation one-step ahead in only 18.3% (31.5%) of all series, however causation is detected at $h = 2$ in 70.1% (81.9%) of all series. It seems that a relatively large sample size would be required in order for empirical rejection rates at $h = 1$ to reach a reasonable level, in particular for the bootstrap tests.

The characteristic that causal relationships in VMA processes may not be sufficiently detected using VAR models presents a clear case for the need to implement multiple-horizon causality tests. Beyond the obvious necessity for such tests when true causal lags exists, even when a causal lag is impossible Wald tests may not be able to detect causation in the classic sense of a one-step ahead VAR forecast improvement. If classic tests at $h = 1$ were the only tests performed in the present setting, empirical power would be a dismal .315 (.328) for a sample size $n = 500$ based on the asymptotic (bootstrap) distribution. However, if we generalize the concept of power to engross the probability of detecting causation at *any horizon* at or beyond the true horizon of causality (in this case, $h = 1$), power reaches 82.3% (78.9%).

Table 3

Empirical Size and Power:VAR(6)

$H_0^\infty : Y \overset{(\infty)}{\nrightarrow} X \mid I_{XZ}$					$H_1^1 : Y \overset{1}{\rightarrow} X \mid I_{XZ}$				
n	h=0[a]	h=1	h=2	h=3	n	h=0	h=1	h=2	h=3
100	.955	.092[b]	.136	.137	100	.596	.623	.833	.886
	[.946]	[.055][c]	[.063]	[.047]		[.535]	[.377]	[.395]	[.465]
200	.909	.149	.201	.209	200	.133	.945	.973	.992
	[.937]	[.104]	[.094]	[.086]		[.102]	[.906]	[.945]	[.945]
300	.913	.081	.145	.162	300	.050	.967	.983	.989
	[.939]	[.070]	[.064]	[.058]		[.028]	[.961]	[.978]	[.983]
400	.885	.076	.111	.124	400	.026	.992	.994	.998
	[.899]	[.065]	[.059]	[.061]		[.022]	[.986]	[.996]	[.998]
500	.946	.066	.070	.076	500	.003	.993	.993	.999
	[.941]	[.061]	[.061]	[.054]		[.001]	[.989]	[.992]	[.994]

$H_1^2 : Y \overset{1}{\nrightarrow} X \mid I_{XZ}, Y \overset{2}{\rightarrow} X \mid I_{XZ}$					$H_1^3 : Y \overset{(2)}{\nrightarrow} X \mid I_{XZ}, Y \overset{3}{\rightarrow} X \mid I_{XZ}$				
n	h=0	h=1	h=2	h=3	n	h=0	h=1	h=2	h=3
100	.569	.119	.390	.517	100	.612	.150	.284	.400
	[.468]	[.037]	[.093]	[.167]		[.598]	[.058]	[.052]	[.120]
200	.136	.098	.432	.604	200	.186	.124	.198	.466
	[.090]	[.062]	[.244]	[.364]		[.188]	[.066]	[.058]	[.220]
300	.030	.112	.540	.658	300	.071	.071	.176	.547
	[.021]	[.040]	[.402]	[.552]		[.065]	[.053]	[.041]	[.317]
400	.015	.087	.551	.777	400	.022	.108	.186	.592
	[.012]	[.053]	[.418]	[.671]		[.018]	[.051]	[.049]	[.401]
500	.003	.064	.582	.812	500	.020	.150	.177	.642
	[.004]	[.051]	[.487]	[.735]		[.015]	[.049]	[.048]	[.473]

Notes: a. Values below "h = 0" denote series frequencies for which we fail to reject H_0^∞.
b. Rejection rates based on p-values derived from the chi-squared distribution.
c. Rejection rates based on p-values derived from the parametric bootstrap.

Table 4

Empirical Size and Power: VMA(1)

$H_0 : Y \overset{(\infty)}{\nrightarrow} X \mid I_{XZ}$					$H_1 : Y \overset{1}{\rightarrow} X \mid I_{XZ}$				
n	h=0	h=1	h=2	h=3	n	h=0	h=1	h=2	h=3
100	.964	.074	.172	.197	100	.850	.033	.392	.408
	[.947]	[.039]	[.066]	[.053]		[.683]	[.025]	[.242]	[.242]
200	.900	.108	.152	.183	200	.525	.117	.675	.652
	[.892]	[.097]	[.108]	[.083]		[.375]	[.118]	[.609]	[.567]
300	.917	.034	.076	.108	300	.467	.219	.758	.759
	[.932]	[.057]	[.059]	[.068]		[.264]	[.225]	[.717]	[.729]
400	.958	.076	.118	.132	400	.442	.183	.701	.708
	[.936]	[.058]	[.057]	[.066]		[.274]	[.207]	[.669]	[.643]
500	.959	.058	.051	.068	500	.367	.315	.819	.823
	[.942]	[.050]	[.025]	[.042]		[.199]	[.328]	[.820]	[.789]

Appendix 3: Formal Proofs

Proof of Theorem 2.1. Consider (iv) and assume $Y \xrightarrow{1} Z \xrightarrow{1} X$ where Z is univariate. Then either $Y \xrightarrow{1} X|I_{XZ}$ or $Y \overset{1}{\nrightarrow} X|I_{XZ}$. Suppose $Y \overset{1}{\nrightarrow} X|I_{XZ}$: we will prove $Y \overset{(h-1)}{\nrightarrow} X|I_{XZ}$ and $Y \xrightarrow{h} X|I_{XZ}$ for some $h \geq 2$.

From Lemma 2.1.2, below, for any $h \geq 2$ given $Y \xrightarrow{1} Z$, if $Y \overset{(h-1)}{\nrightarrow} X|I_{XZ}$ then $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ *if and only if* $c^{t,1}_{z,t-h+1,t-h+1} = 0$ with probability one for every t, where $c^{t,1}_{z,t-h+1,t-h+1}$ denotes that unique component of the subspace $Z(t-h+1, t-h+1]$ that enters into the orthogonal projection of X_{t+1} onto $I_{XZ}(t) + Y(-\infty, t]$. Thus, no component of $Z(t-h+1, t-h+1]$ enters into the 1-step ahead metric projection of X_t. However, because $Z \xrightarrow{1} X$, it must be the case that $c^{t,1}_{z,t-h+1,t-h+1} \neq 0$ for *some* $h \geq 2$: some component of the sub-space $Z(t-h+1, t-h+1]$ must enter into the 1-step ahead metric projection of X_t for some $h \geq 2$. Therefore, by Lemma 2.1.2, $Y \overset{(h-1)}{\nrightarrow} X|I_{XZ}$ and $c^{t,1}_{z,t-h+1,t-h+1} \neq 0$ for some $h \geq 2$ implies $Y \xrightarrow{h} X|I_{XZ}$. □

Lemma 2.1.1 *Let Z_t be scalar-valued. Denote by $c^{t-k_3,h}_{z,t-k_1,t-k_2}$ that unique element of the span $Z(t-k_1, t-k_2]$ that enters into the projection of X_{t+h} onto $I_{XZ}(t-k_3) + Z(-\infty, t-k_3]$, $k_3 \leq k_2 \leq k_1$. For any $h \geq 2$, if $Y \overset{(h-1)}{\nrightarrow} X|I_{XZ}$ and $Y \xrightarrow{1} Z$, then $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ if and only if $c^{t,h-1}_{z,t,t} = 0$ with probability one for every t.*

Lemma 2.1.2 *Let Z_t be scalar-valued. For any $h \geq 1$, if $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ and $Y \xrightarrow{1} Z$ then $Y \overset{(h+1)}{\nrightarrow} X|I_{XZ}$ if and only if $c^{t,1}_{z,t-h+1,t-h+1} = 0$ with probability one for every t.*

Proof of Lemma 2.1.1. Assume $Y \overset{(h-1)}{\nrightarrow} X|I_{XZ}$ for some $h \geq 2$. By iterated projections for Hilbert space projection operators

$$\begin{aligned} & P\left(X_{t+h}|I_{XZ}(t) + Y(-\infty, t]\right) \\ = \; & P\left(P\left(X_{t+h}|I_{XZ}(t+1) + Y(-\infty, t+1]\right)|I_{XZ}(t) + Y(-\infty, t]\right) \\ = \; & P\left(P\left(X_{t+h}|I_{XZ}(t+1)\right)|I_{XZ}(t) + Y(-\infty, t]\right). \end{aligned} \tag{6.6}$$

Notice $I_{XZ}(t+1)$ decomposes into

$$\begin{aligned} I_{XZ}(t+1) &= H + X(-\infty, t+1] + Z(-\infty, t+1] \\ &= H + X(-\infty, t] + X(t+1, t+1] + Z(-\infty, t] + Z(t+1, t+1] \\ &= I_{XZ}(t) + X(t+1, t+1] + Z(t+1, t+1]. \end{aligned} \tag{6.7}$$

Hence, we may write $P(X_{t+h}|I_{XZ}(t+1))$ as

$$P\left(X_{t+h}|I_{XZ}(t+1)\right) = a^{t+1,h}_{xz,t} + b^{t+1,h}_{x,t+1,t+1} + c^{t+1,h}_{z,t+1,t+1}, \tag{6.8}$$

where

$$a^{t+1,h}_{xz,t} \in I_{XZ}(t), \;\; b^{t+1,h}_{x,t+1,t+1} \in X(t+1, t+1], \;\; c^{t+1,h}_{z,t+1,t+1} \in Z(t+1, t+1]. \tag{6.9}$$

We obtain from projection operator linearity, the assumption $Y \overset{1}{\nrightarrow} X|I_{XZ}$ and $b^{t+1,h}_{x,t+1,t+1} \in X(t+1,t+1]$,

$$\begin{aligned} & P\left(X_{t+h}|I_{XZ}(t)+Y(-\infty,t]\right) \qquad (6.10)\\ = \;& P\left(P\left(X_{t+h}|I_{XZ}(t+1)\right)|I_{XZ}(t)+Y(-\infty,t]\right)\\ = \;& P\left(a^{t+1,h}_{xz,t}+b^{t+1,h}_{x,t+1,t+1}+c^{t+1,h}_{z,t+1,t+1}|I_{XZ}(t)+Y(-\infty,t]\right)\\ = \;& a^{t+1,h}_{xz,t}+P\left(b^{t+1,h}_{x,t+1,t+1}|I_{XZ}(t)+Y(-\infty,t]\right)\\ & +P\left(c^{t+1,h}_{z,t+1,t+1}|I_{XZ}(t)+Y(-\infty,t]\right)\\ = \;& a^{t+1,h}_{xz,t}+P\left(b^{t+1,h}_{x,t+1,t+1}|I_{XZ}(t)\right)+P\left(c^{t+1,h}_{z,t+1,t+1}|I_{XZ}(t)+Y(-\infty,t]\right). \end{aligned}$$

Because $Y \overset{1}{\to} Z$, $c^{t+1,h}_{z,t+1,t+1} \in Z(t+1,t+1]$, and $Z(t+1,t+1]$ is a scalar-valued Hilbert space, we deduce $P(c^{t+1,h}_{z,t+1,t+1}|I_{XZ}(t)+Y(-\infty,t]) = P(c^{t+1,h}_{z,t+1,t+1}|I_{XZ}(t))$ with probability one *if and only if* $c^{t+1,h}_{z,t+1,t+1}=0$ with probability one for all t. If Z_t were multivariate, then elements of the single-period span $Z(t+1,t+1]$ contain linear combinations of the multiple $Z_{t,i}$-components, hence $c^{t+1,h}_{z,t+1,t+1} \neq 0$ would be possible while also $P(c^{t+1,h}_{z,t+1,t+1}|I_{XZ}(t)+Y(-\infty,t]) = P(c^{t+1,h}_{z,t+1,t+1}|I_{XZ}(t)=0$ due to linearity of the projection operator and causal neutralization. Therefore if $Y \overset{(h-1)}{\nrightarrow} X|I_{XZ}$, $Y \overset{1}{\to} Z$, and Z is scalar-valued, then $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ *if and only if* $c^{t+1,h}_{z,t+1,t+1}=0$ with probability one for every t. Because t and h are arbitrary, we conclude $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ *if and only if* $c^{t,h-1}_{z,t,t}=0$ with probability one for every t. □

Proof of Lemma 2.1.2. We prove the claim by induction. Let $Y \overset{1}{\nrightarrow} X|I_{XZ}$. By Lemma 2.1.1, $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$ *if and only if* $c^{t,1}_{z,t,t}=0$ with probability one for every t. This proves the claim for $h=1$.

Now, for any $h \geq 2$ assume $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ *if and only if* $c^{t,1}_{z,t-k+2,t-k+2}=0$ with probability one for every t and each $k=2...h$. We will prove if $Y \overset{(h+1)}{\nrightarrow} X|I_{XZ}$ *if and only if* $c^{t,1}_{z,t-h+1,t-h+1}=0$.

By iterated projections, the assumption $Y \overset{1}{\nrightarrow} X|I_{XZ}$, and the decomposition

$$I_{XZ}(t+h) = I_{XZ}(t)+X(t+1,t+h]+Z(t+1,t+h], \qquad (6.11)$$

we obtain

$$\begin{aligned} & P\left(X_{t+h+1}|I_{XZ}(t)+Y(-\infty,t]\right) \qquad (6.12)\\ & P\left(P\left(X_{t+h+1}|I_{XZ}(t+h)+Y(-\infty,t+h]\right)|I_{XZ}(t)+Y(-\infty,t]\right)\\ = \;& P\left(P\left(X_{t+h+1}|I_{XZ}(t+h)\right)|I_{XZ}(t)+Y(-\infty,t]\right)\\ = \;& P\left(a^{t+h,h+1}_{xz,t}+b^{t+h,h+1}_{x,t+1,t+h}+c^{t+h,h+1}_{z,t+1,t+h}|I_{XZ}(t)+Y(-\infty,t]\right)\\ = \;& a^{t+h,h+1}_{xz,t}+P\left(b^{t+h,h+1}_{x,t+1,t+h}|I_{XZ}(t)+Y(-\infty,t]\right)\\ & +P\left(c^{t+h,h+1}_{z,t+1,t+h}|I_{XZ}(t)+Y(-\infty,t]\right). \end{aligned}$$

By $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$, $b^{t+h,h+1}_{x,t+1,t+h} \in X(t+1,t+h]$ and projection operator linearity, we deduce

$$P\left(b^{t+h,h+1}_{x,t+1,t+h}|I_{XZ}(t)+Y(-\infty,t]\right) = P\left(b^{t+h,h+1}_{x,t+1,t+h}|I_{XZ}(t)\right). \tag{6.13}$$

Moreover, the element $c^{t+h,h+1}_{z,t+1,t+h}$ denotes that unique component of the subspace $Z(t+1,t+h]$ that enters into the orthogonal projection of X_{t+h+1} onto $I_{XZ}(t+h)+Y(-\infty,t+h]$. Because $Z(t+1,t+h]$ decomposes into

$$Z(t+1,t+h] = Z(t+1,t+1]+...+Z(t+h,t+h] \tag{6.14}$$

we deduce $c^{t+h,h+1}_{z,t+1,t+h}$ satisfies for every t

$$c^{t+h,h+1}_{z,t+1,t+h} = c^{t+h,h+1}_{z,t+1,t+1}+...+c^{t+h,h+1}_{z,t+h,t+h}, \tag{6.15}$$

hence, because t and h are arbitrary,

$$c^{t,1}_{z,t-h+1,t} = c^{t,1}_{z,t-h+1,t-h+1}+...+c^{t,1}_{z,t,t}. \tag{6.16}$$

By the induction assumption $c^{t,1}_{z,t-k+2,t-k+2}=0$ (hence $c^{t+h,1}_{z,t+h-k+2,t+h-k+2}=0$) with probability one for every t and each $k=2...h$, thus

$$c^{t+h,h+1}_{z,t+1,t+h} = c^{t+h,h+1}_{z,t+1,t+1}. \tag{6.17}$$

This implies

$$\begin{aligned} & P\left(X_{t+h+1}|I_{XZ}(t)+Y(-\infty,t]\right) \qquad (6.18)\\ = \;& a^{t+h,h+1}_{xz,t}+P\left(b^{t+h,h+1}_{x,t+1,t+h}|I_{XZ}(t)+Y(-\infty,t]\right)+P\left(c^{t+h,h+1}_{z,t,t}|I_{XZ}(t)+Y(-\infty,t]\right)\\ = \;& a^{t+h,h+1}_{xz,t}+P\left(b^{t+h,h+1}_{x,t+1,t+h}|I_{XZ}(t)\right)+P\left(c^{t+h,h+1}_{z,t+1,t+1}|I_{XZ}(t)+Y(-\infty,t]\right). \end{aligned}$$

We deduce if $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ then $Y \overset{(h+1)}{\nrightarrow} X|I_{XZ}$ *if and only if* $P(c^{t+h,h+1}_{z,t+1,t+1}|I_{XZ}(t)+Y(-\infty,t]) = P(c^{t+h,h+1}_{z,t+1,t+1}|I_{XZ}(t)])$ for all t. Using the logic from the line of proof of Lemma 2.1.1, because $Z(t+1,t+1]$ is a scalar-valued Hilbert space, $c^{t+h,h+1}_{z,t+1,t+1} \in Z(t+1,t+1]$, and $Y \overset{1}{\rightarrow} Z$, we deduce $Y \overset{(h+1)}{\nrightarrow} X|I_{XZ}$ *if and only if* $c^{t+h,h+1}_{z,t+1,t+1}=0$ with probability one for all t, or $c^{t,1}_{z,t-h+1,t-h+1}=0$ with probability one for all t. □

Proof of Lemma 3.1. Recall we assume $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ and $Y \overset{1}{\rightarrow} Z|I_{XZ}$. From (3.5), and the assumption $m_z=1$, we know $Y \overset{(h+1)}{\nrightarrow} X|I_{XZ}$ *if and only if* $\pi^{(h)}_{XZ,1}=0$. Therefore, if $Y \overset{1}{\nrightarrow} X|I_{XZ}$, then $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$ *if and only if* $\pi_{XZ,1}=0$.

Now, from (2.3)) we deduce

$$\pi^{(h)}_{XZ,1} = \pi^{(h-1)}_{XZ,2}+\pi^{(h-1)}_{XX,1}\pi_{XZ,1}+\pi^{(h-1)}_{XY,1}\pi_{YZ,1}+\pi^{(h-1)}_{XZ,1}\pi_{ZZ,1}. \tag{6.19}$$

For $h=2$, $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$ implies $\pi^{(1)}_{XZ,1}=0$ from above, hence

$$\begin{aligned} \pi^{(2)}_{XZ,1} &= \pi_{XZ,2}+\pi_{XX,1}\pi_{XZ,1}+\pi_{XY,1}\pi_{YZ,1}+\pi_{XZ,1}\pi_{ZZ,1} \qquad (6.20)\\ &= \pi_{XZ,2}+\pi_{XX,1}\times 0+0\times\pi_{YZ,1}+0\times\pi_{ZZ,1}\\ &= \pi_{XZ,2}. \end{aligned}$$

For $h=3$, $Y \overset{(3)}{\nrightarrow} X|I_{XZ}$ implies $\pi^{(2)}_{XZ,1} = \pi^{(1)}_{XZ,1} = 0$ from above, hence

$$\begin{aligned} \pi^{(3)}_{XZ,1} &= \pi^{(2)}_{XZ,2} + \pi^{(2)}_{XX,1}\pi_{XZ,1} + \pi^{(2)}_{XY,1}\pi_{YZ,1} + \pi^{(2)}_{XZ,1}\pi_{ZZ,1} \\ &= \pi^{(2)}_{XZ,2} + \pi^{(2)}_{XX,1} \times 0 + 0 \times \pi_{YZ,1} + 0 \times \pi_{ZZ,1} \\ &= \pi^{(2)}_{XZ,2}, \end{aligned} \tag{6.21}$$

where

$$\begin{aligned} \pi^{(2)}_{XZ,2} &= \pi_{XZ,3} + \pi_{XX,1}\pi_{XZ,2} + \pi_{XY,1}\pi_{YZ,2} + \pi_{XZ,1}\pi_{ZZ,2} \\ &= \pi_{XZ,3} + \pi_{XX,1}\pi_{XZ,2} + 0 \times \pi_{YZ,2} + 0 \times \pi_{ZZ,2} \\ &= \pi_{XZ,3} + \pi_{XX,1}\pi_{XZ,2}. \end{aligned} \tag{6.22}$$

Thus,

$$\pi^{(3)}_{XZ,1} = \pi^{(2)}_{XZ,2} = \pi_{XZ,3} + \pi_{XX,1}\pi_{XZ,2}. \tag{6.23}$$

Recursively we deduce for $h \geq 2$, $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ implies

$$\pi^{(h)}_{XZ,1} = \pi_{XZ,h} + \sum\nolimits_{i=1}^{h-1} \pi^{(h-i)}_{XX,1}\pi_{XZ,i}. \tag{6.24}$$

Notice in (6.24) we include the term $\pi^{(h-1)}_{XX,1}\pi_{XZ,1} = 0$ which follows from above: $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ for $h \geq 2$ implies $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$, hence $\pi_{XZ,1} = 0$. □

Proof of Theorem 3.2*. i.* Consider $h=2$ and assume $Y \overset{1}{\nrightarrow} X|I_{XZ}$. By the assumptions $Y \overset{1}{\nrightarrow} X|I_{XZ}$ and $Y \overset{1}{\rightarrow} Z|I_{XZ}$, and $m_z = 1$, from (3.5) we deduce $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$ *if and only if* $\pi_{XZ,1} = 0$.

For the remainder of the proof, assume $h > 2$. Let $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$. From Lemma 3.1, it follows that $\pi^{(k)}_{XZ,1}$ has the representation $\pi^{(2)}_{XZ,1} = \pi_{XZ,2}$, and for $k > 2$

$$\pi^{(k)}_{XZ,1} = \pi_{XZ,k} + \sum\nolimits_{i=1}^{k-1} \left(\pi^{(k-i)}_{XX,1}\pi_{XZ,i} \right). \tag{6.25}$$

Moreover, $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ implies by definition $Y \overset{k}{\nrightarrow} X|I_{XZ}$, $k = 1...h$, and we deduce from (3.5) that $\pi^{(k)}_{XZ,1} = 0$, $k = 1...h-1$. Using the zero identities $\pi^{(k)}_{XZ,1} = 0$, $k = 1...h-1$, we deduce from Lemma 3.1, cf. formula (3.6),

$$\begin{aligned} 0 &= \pi^{(1)}_{XZ,1} = \pi_{XZ,1} \\ 0 &= \pi^{(2)}_{XZ,1} = \pi_{XZ,2} + \left(\pi^{(1)}_{XX,1}\pi_{XZ,1} \right) = \pi_{XZ,2} \\ 0 &= \pi^{(3)}_{XZ,1} = \pi_{XZ,3} + \left(\pi^{(2)}_{XX,1}\pi_{XZ,1} \right) + \left(\pi^{(1)}_{XX,1}\pi_{XZ,2} \right) \\ &= \pi_{XZ,3} + \left(\pi^{(2)}_{XX,1} \times 0 \right) + \left(\pi^{(1)}_{XX,1} \times 0 \right) = \pi_{XZ,3} \\ &\ldots \\ 0 &= \pi^{(h-1)}_{XZ,1} \\ &= \pi_{XZ,h-1} + \sum\nolimits_{i=1}^{h-2} \left(\pi^{(h-1-i)}_{XX,1} \times 0 \right) = \pi_{XZ,h-1}, \end{aligned} \tag{6.26}$$

which gives $\pi_{XZ,k} = 0$, $k = 1...h-1$. This proves the first direction.

Conversely, let $Y \overset{1}{\nrightarrow} X|I_{XZ}$ and $\pi_{XZ,i} = 0$, $i = 1...h-1$. From equation (6.19) in the line of proof of Lemma 3.1, we have

$$\pi^{(k)}_{XZ,1} = \pi^{(k-1)}_{XZ,2} + \left(\pi^{(k-1)}_{XX,1}\pi_{XZ,1} + \pi^{(k-1)}_{XY,1}\pi_{YZ,1} + \pi^{(k-1)}_{XZ,1}\pi_{ZZ,1}\right). \tag{6.27}$$

If $k = 1$, then trivially $\pi^{(1)}_{XZ,1} = \pi_{XZ,1} = 0$, and by (3.5) $Y \overset{2}{\nrightarrow} X|I_{XZ}$ follows. Thus, along with $Y \overset{1}{\nrightarrow} X|I_{XZ}$ by assumption, we obtain $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$. However, by Theorem 1 and Lemma 3.1, $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$ implies $\pi^{(k)}_{XY,j} = 0$, $k = 1, 2$, $j \geq 1$ and

$$\pi^{(2)}_{XZ,1} = \pi_{XZ,2} + \sum\nolimits_{i=1}^{2-1}\left(\pi^{(2-i)}_{XX,1}\pi_{XZ,i}\right) = \pi_{XZ,2} = 0. \tag{6.28}$$

Thus, by (3.5), $Y \overset{3}{\nrightarrow} X|I_{XZ}$ follows immediately, and in conjunction with $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$, we deduce $Y \overset{(3)}{\nrightarrow} X|I_{XZ}$. Recursively, for each $k = 1...h-1$ that we have $Y \overset{(k)}{\nrightarrow} X|I_{XZ}$, from Lemma 3.1 we deduce

$$\pi^{(k)}_{XZ,1} = \pi_{XZ,k} + \sum\nolimits_{i=1}^{k-1}\left(\pi^{(k-i)}_{XX,1}\pi_{XZ,i}\right), \tag{6.29}$$

which reduces to

$$\begin{aligned} \pi^{(k)}_{XZ,1} &= \pi_{XZ,k} + \sum\nolimits_{i=2}^{k-1}\left(\pi^{(k-i)}_{XX,1}\pi_{XZ,i}\right) \\ &= 0 + \sum\nolimits_{i=1}^{k-1}\left(\pi^{(k-i)}_{XX,1} \times 0\right) = 0. \end{aligned} \tag{6.30}$$

By (3.5) and $\pi^{(k)}_{XZ,1} = 0$, we deduce $Y \overset{k+1}{\nrightarrow} X|I_{XZ}$ for $k = 1...h-1$. Combined with the assumption $Y \overset{1}{\nrightarrow} X|I_{XZ}$, it follows that $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$. This proves claim (i).

ii. The result is a direct consequence of claim (i): $Y \overset{(h+1)}{\nrightarrow} X|I_{XZ}$ *if and only if* $Y \overset{1}{\nrightarrow} X|I_{XZ}$ and $\pi_{XZ,i} = 0$, $i = 1...h$. Thus, given $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$, we have $\pi_{XZ,i} = 0$, $i = 1...h-1$, hence $Y \overset{(h+1)}{\nrightarrow} X|I_{XZ}$ follows *if and only if* $\pi_{XZ,h} = 0$. □

Proof of Lemma 3.3. From (2.1), due to error orthogonality $\epsilon_t \perp W(-\infty, t-1]$ we have

$$\begin{aligned} P\left(X_{t+h}|I_{XZ} + Y(-\infty, t]\right) &= \sum\nolimits_{k=1}^{\infty}\pi^{(h)}_{XX,k}X_{t+1-k} \\ &\quad + \sum\nolimits_{k=1}^{\infty}\pi^{(h)}_{XY,k}Y_{t+1-k} + \sum\nolimits_{k=1}^{\infty}\pi^{(h)}_{XZ,k}Z_{t+1-k}. \end{aligned} \tag{6.31}$$

Projecting both sides onto the subspace $I_{XZ_1} + Y(-\infty,t] \subseteq I_{XZ} + Y(-\infty,t]$, and invoking iterated projections and projection operator linearity, we obtain

$$\begin{aligned}
& P\left(P\left(X_{t+h}|I_{XZ}+Y(-\infty,t]\right)|I_{XZ_1}+Y(-\infty,t]\right) \qquad (6.32)\\
= \;& P\left(X_{t+h}|I_{XZ_1}+Y(-\infty,t]\right)\\
= \;& \sum_{i=1}^{\infty}\pi_{XX,i}^{(h)}X_{t+1-i}+\sum_{i=1}^{\infty}\pi_{XY,i}^{(h)}Y_{t+1-i}\\
& +\sum_{i=1}^{\infty}\pi_{XZ,i}^{(h)}P\left(Z_{t+1-i}|I_{XZ_1}+Y(-\infty,t]\right)\\
= \;& \sum_{i=1}^{\infty}\pi_{XX,i}^{(h)}X_{t+1-i}+\sum_{i=1}^{\infty}\pi_{XY,i}^{(h)}Y_{t+1-i}+\sum_{i=1}^{\infty}\pi_{XZ_1,i}^{(h)}Z_{1,t+1-i}\\
& +\sum_{i=1}^{\infty}\pi_{XZ_2,i}^{(h)}P\left(Z_{2,t+1-i}|I_{XZ_1}+Y(-\infty,t]\right)\\
= \;& \sum_{i=1}^{\infty}\pi_{XX,i}^{(h)}X_{t+1-i}+\sum_{i=1}^{\infty}\pi_{XY,i}^{(h)}Y_{t+1-i}+\sum_{i=1}^{\infty}\pi_{XZ_1,i}^{(h)}Z_{1,t+1-i}\\
& +\sum_{i=1}^{\infty}\pi_{XZ_2,i}^{(h)}\left[\sum_{k=1}^{\infty}\beta_{Z_2X,k}^{j,t+1-i}X_{t+1-k}\right]\\
& +\sum_{i=1}^{\infty}\pi_{XZ_2,i}^{(h)}\left[\sum_{k=1}^{\infty}\beta_{Z_2Y,k}^{j,t+1-i}Y_{t+1-k}+\sum_{k=1}^{\infty}\beta_{Z_2Z_1,k}^{j,t+1-i}Z_{1,t+1-k}\right]\\
= \;& \sum_{i=1}^{\infty}\pi_{XX,i}^{(h)}X_{t+1-i}+\sum_{i=1}^{\infty}\pi_{XY,i}^{(h)}Y_{t+1-i}+\sum_{i=1}^{\infty}\pi_{XZ_1,i}^{(h)}Z_{1,t+1-i}\\
& +\sum_{k=1}^{\infty}X_{t+1-k}\left(\sum_{i=1}^{\infty}\pi_{XZ_j,i}^{(h)}\beta_{Z_2X,k}^{j,1-i}\right)\\
& +\sum_{k=1}^{\infty}Y_{t+1-k}\left(\sum_{i=1}^{\infty}\pi_{XZ_j,i}^{(h)}\beta_{Z_2Y,k}^{j,1-i}\right)\\
& +\sum_{k=1}^{\infty}Z_{1,t+1-k}\left(\sum_{i=1}^{\infty}\pi_{XZ_j,i}^{(h)}\beta_{Z_2Z_1,k}^{j,1-i}\right)\\
= \;& \sum_{k=1}^{\infty}X_{t+1-k}\left(\pi_{XX,k}^{(h)}+\sum_{i=1}^{\infty}\pi_{XZ_j,i}^{(h)}\beta_{Z_2X,k}^{j,1-i}\right)\\
& +\sum_{i=k}^{\infty}Y_{t+1-k}\left(\pi_{XY,k}^{(h)}+\sum_{i=1}^{\infty}\pi_{XZ_j,i}^{(h)}\beta_{Z_2Y,k}^{j,1-i}\right)\\
& +\sum_{i=k}^{\infty}Z_{1,t+1-k}\left(\pi_{XZ_1,k}^{(h)}+\sum_{i=1}^{\infty}\pi_{XZ_j,i}^{(h)}\beta_{Z_2Z_1,k}^{j,1-i}\right)\\
= \;& \sum_{i=k}^{\infty}\delta_{XX,k}^{(h)}X_{t+1-k}+\sum_{k=1}^{\infty}\delta_{XY,k}^{(h)}Y_{t+1-k}+\sum_{k=1}^{\infty}\delta_{XZ_1,k}^{(h)}Z_{1,t+1-k}
\end{aligned}$$

hence

$$\begin{aligned}
P\left(X_{t+h}|I_{XZ_1}+Y(-\infty,t]\right) \;=\; & \sum_{k=1}^{\infty}\delta_{XX,k}^{(h)}X_{t+1-k} \qquad (6.33)\\
& +\sum_{k=1}^{\infty}\delta_{XY,k}^{(h)}Y_{t+1-k}+\sum_{k=1}^{\infty}\delta_{XZ_1,k}^{(h)}Z_{1,t+1-k}
\end{aligned}$$

where

$$\delta_{XY,k}^{(h)} \equiv \pi_{XY,k}^{(h)}+\sum_{i=1}^{\infty}\pi_{XZ_2,i}^{(h)}\beta_{Z_2Y,k}^{1-i}, \qquad (6.34)$$

and $\beta_{Z_2Y,k}^{1-i}$ denotes the Y-specific coefficients in the projection of each vector $Z_{2,t+1-i}$ onto $I_{XZ_1}+Y(-\infty,t]$, $i \geq 1$. □

Proof of Theorem 3.4*. i.* Let $Z_2 \overset{1}{\nrightarrow} X|I_{XZ_1}$. Then $\pi_{XZ_2,j}=0$, $\forall j \geq 1$, hence from Lemma 3.3, cf. (3.8), we obtain $\delta_{XY,j}=\pi_{XY,j}$. Therefore $\delta_{XY,j}=0$ *if and only if* $\pi_{XY,j}=0$, which implies $Y \overset{1}{\nrightarrow} X|I_{XZ_1}$ *if and only if* $Y \overset{1}{\nrightarrow} X|I_{XZ}$.

ii. Let $(Y,Z_2) \overset{1}{\nrightarrow} X|I_{XZ_1}$ and $Y \overset{(h)}{\nrightarrow} X|I_{XZ_1}$ for any $h \geq 1$, and recall $\delta_j^{(h)}$ denote the VAR coefficients in the projection of W_{t+h} onto $I_{XZ_1}(t)+Y(-\infty,t]$. By the

assumption $Z_2 \overset{1}{\nrightarrow} X|I_{XZ_1}$, we just proved $Y \overset{(h)}{\nrightarrow} X|I_{XZ_1}$ *if and only if* $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ for $h = 1$. Therefore let $h \geq 2$. By Lemma 3.1 and Theorem 3.2, $Y \overset{(h)}{\nrightarrow} X|I_{XZ_1}$ implies $\delta^{(k)}_{XZ_1,1} = \delta_{XZ_1,k} = 0$, $k = 1...h-1$, and using Lemma 3.3 for $\delta^{(h)}_{XZ_1,j}$ and $\delta^{(k)}_{XZ_1,1}$ we deduce

$$\begin{aligned} \delta^{(h)}_{XY,j} &\equiv \pi^{(h)}_{XY,j} + \sum\nolimits_{i=1}^{\infty} \pi^{(h)}_{XZ_2,i}\beta^{1-i}_{Z_2Y,j} \\ \delta^{(k)}_{XZ_1,1} &= \delta_{XZ_1,k} = \pi_{XZ_1,k} + \sum\nolimits_{i=1}^{\infty} \pi_{XZ_2,i}\beta^{1-i}_{Z_2Z_1,k} = 0. \end{aligned} \tag{6.35}$$

The assumption $Z_2 \overset{1}{\nrightarrow} X|I_{XZ_1}$ implies $\pi_{XZ_2,i} = 0$, $i \geq 1$, hence $\pi_{XY,j} = \delta_{XY,j} = 0\ \forall j \geq 1$, giving $Y \overset{1}{\nrightarrow} X|I_{XZ}$, cf. Theorem 2.2. Additionally, the zeros $\delta_{XZ_1,k} = 0$, $k = 1...h-1$, imply $\delta_{XZ_1,k} = \pi_{XZ_1,k} = 0$, $k = 1...h-1$, hence $\pi_{XZ,1} = [\pi_{XZ_1,1}, \pi_{XZ_2,1}] = 0$. From (3.5) we deduce $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$ given $\pi_{XZ,1}\pi_{ZY,j} = 0$.

Similarly, using (2.3) non-causation $Y \overset{(2)}{\nrightarrow} X|I_{XZ}$ and $\pi_{XZ,1} = 0$ imply

$$\begin{aligned} \pi^{(2)}_{XZ,1} &= \pi_{XZ,2} + \pi_{XX,1}\pi_{XZ,1} + \pi_{XY,1}\pi_{YZ,1} + \pi_{XZ,1}\pi_{ZZ,1} \\ &= \pi_{XZ,2} + \pi_{XX,1} \times 0 + 0 \times \pi_{YZ,1} + 0 \times \pi_{ZZ,1} \\ &= \pi_{XZ,2}. \end{aligned} \tag{6.36}$$

If $h \geq 3$, then non-causation $Z_2 \overset{1}{\nrightarrow} X|I_{XZ_1}$ and the identity $\delta_{XZ_1,k} = \pi_{XZ_1,k} = 0$, $k = 1...h-1$ again imply $\pi_{XZ,2} = 0$, thus $\pi^{(2)}_{XZ,1}\pi_{ZY,j} = 0$ for all j, giving $Y \overset{(3)}{\nrightarrow} X|I_{XZ}$, cf. (3.5). Repeating this logic,

$$\begin{aligned} \pi^{(3)}_{XZ,1} &= \pi^{(2)}_{XZ,2} + \pi^{(2)}_{XX,1}\pi_{XZ,1} + \pi^{(2)}_{XY,1}\pi_{YZ,1} + \pi^{(2)}_{XZ,1}\pi_{ZZ,1} \\ &= \pi^{(2)}_{XZ,2} + \pi^{(2)}_{XX,1} \times 0 + 0 \times \pi_{YZ,1} + 0 \times \pi_{ZZ,1} \\ &= \pi^{(2)}_{XZ,2} \\ &= \pi_{XZ,3} + \pi_{XX,1}\pi_{XZ,2} + \pi_{XY,1}\pi_{YZ,2} + \pi_{XZ,1}\pi_{ZZ,2} \\ &= \pi_{XZ,3} + \pi_{XX,1} \times 0 + 0 \times \pi_{YZ,2} + 0 \times \pi_{ZZ,2} \\ &= \pi_{XZ,3}, \end{aligned} \tag{6.37}$$

and so on. Recursively we deduce $(Y, Z_2) \overset{1}{\nrightarrow} X|I_{XZ_1}$ and $Y \overset{(h)}{\nrightarrow} X|I_{XZ_1}$ imply $\delta_{XZ_1,k} = \pi_{XZ_1,k} = 0$, $k = 1...h-1$, $\pi_{XZ_2,i} = 0$, $\forall i \geq 1$, and $\pi_{XZ,k} = \pi^{(k)}_{XZ,1}$, $k = 1...h-1$, hence $\pi^{(k)}_{XZ,1} = 0$, $k = 1...h-1$, giving $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$.

iii. and *iv.* For any $h \geq 1$, if $(Y, Z_2) \overset{1}{\nrightarrow} X|I_{XZ_1}$, $Y \overset{(h)}{\nrightarrow} X|I_{XZ_1}$ and $Y \overset{h+1}{\rightarrow} X|I_{XZ_1}$, then $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$ also holds by (*ii*). Moreover, by Theorem 2.1 causation $Y \overset{h+1}{\rightarrow} X|I_{XZ_1}$ implies a causal chain $Y \overset{1}{\rightarrow} Z_1 \overset{1}{\rightarrow} X$ must exist. Using the above logic, we recursively deduce $\pi^{(h)}_{XZ,1} = \pi_{XZ,h}$. Because $Z_2 \overset{1}{\nrightarrow} X|I_{XZ_1}$, $Y \overset{(h)}{\nrightarrow} X|I_{XZ_1}$, and $Y \overset{h+1}{\rightarrow} X|I_{XZ_1}$, by Theorems 2.2 and 3.2 it must be the case that $\pi_{XZ_2,i} = 0, \forall i \geq 1$, $_{XZ_1,h} \neq 0$, hence

$$\pi^{(h)}_{XZ,1} = \pi_{XZ,h} = [\pi_{XZ_1,h}, 0]. \tag{6.38}$$

Therefore, by (3.5), $Y \overset{(h+1)}{\nrightarrow} X|I_{XZ}$ *if and only if*

$$\pi^{(h)}_{XZ,1}\pi_{ZY,j} = \pi_{XZ,h}\pi_{ZY,j} = \pi_{XZ_1,h}\pi_{Z_1Y,j} = 0 \tag{6.39}$$

for all $j \geq 1$. Because $Y \overset{1}{\to} Z_1|I_{XZ_1}$ it must be the case that $\pi_{Z_1Y,j} \neq 0$ for at least one j, therefore $Y \overset{(h+1)}{\nrightarrow} X|I_{XZ}$ *if and only if* $\pi_{XZ_1,h} = 0$. Because $\pi_{XZ_1,h} \neq 0$ and $Y \overset{(h)}{\nrightarrow} X|I_{XZ}$, we conclude $Y \overset{h+1}{\to} X|I_{XZ}$. □

Proof of Lemma 4.1. We reject $Y \overset{1}{\nrightarrow} X$ if we reject Tests 0.1- 0.2 ($Y \overset{(\infty)}{\nrightarrow} X$) and Test 1.0 ($Y \overset{1}{\nrightarrow} X$) :

$$\begin{aligned} & P\left(rej.\ H_0^{(1)}|H_0^{(1)} \text{ is true}\right) \\ = \ & P\left(rej.\ 0.1 \cap rej.\ 0.1 \cap rej.\ 1.0|H_0^{(1)} \text{ is true}\right) \\ \leq \ & P\left(rej.\ 1.0|H_0^{(1)} \text{ is true}\right) = \alpha_{1.0}. \end{aligned} \tag{6.40}$$

We reject $Y \overset{(2)}{\nrightarrow} X$ if we reject Tests 0.1- 0.2 ($Y \overset{(\infty)}{\nrightarrow} X$), and either reject Test 1.0 ($Y \overset{1}{\nrightarrow} X$), or fail to reject Test 1.0 and reject Tests 1.1 and 1.2 ($Y \overset{1}{\nrightarrow} Z$ and $Z \overset{1}{\nrightarrow} X$) and Test 2.0 ($Y \overset{1}{\nrightarrow} X$, $\pi_{XZ,1} = 0$). We have

$$\begin{aligned} & P\left(rej.\ H_0^{(2)}|H_0^{(2)} \text{ is true}\right) \\ = \ & P\left((rej0.1 \cap rej0.2) \cap (rej1.0 \cup [fail1.0 \cap rej1.1 \cap rej1.2 \cap rej2.0])\,|H_0^{(1)} \text{ is true}\right). \end{aligned} \tag{6.41}$$

There are four cases to consider: $Y \overset{1}{\nrightarrow} Z \overset{1}{\to} X$, $Y \overset{1}{\to} Z \overset{1}{\nrightarrow} X$, $Y \overset{1}{\nrightarrow} Z \overset{1}{\nrightarrow} X$ and $Y \overset{1}{\to} Z \overset{1}{\to} X$. Under $H_0^{(2)}$, if $Y \overset{1}{\nrightarrow} Z \overset{1}{\to} X$ then $Y \overset{1}{\nrightarrow} (X,Z)$ is true, $(Y,Z) \overset{1}{\nrightarrow} X$ is false, and Test 2.0 represents only a sufficient condition for non-causation $Y \overset{(2)}{\nrightarrow} X$. Hence

$$\begin{aligned} & P\left(rej.\ H_0^{(2)}|H_0^{(2)} \text{ is true}\right) \\ = \ & P\left((rej0.1 \cap rej0.2) \cap (rej1.0 \cup [fail1.0 \cap rej1.1 \cap rej1.2 \cap rej2.0])\,|H_0^{(1)} \text{ is true}\right) \\ \leq \ & P\left(rej0.1 \cap (rej1.0 \cup [fail1.0 \cap rej1.1 \cap rej1.2 \cap rej2.0])\,|H_0^{(1)} \text{ is true}\right) \\ \leq \ & \min\left\{\alpha_{0.1}, P\left(rej1.0 \cup [fail1.0 \cap rej1.1 \cap rej1.2 \cap rej2.0]\,|H_0^{(1)} \text{ is true}\right)\right\} \\ \leq \ & \min\left\{\alpha_{0.1}, \alpha_{1.0} + P\left((rej1.1 \cap rej1.2 \cap rej2.0)\,|H_0^{(1)} \text{ is true}\right)\right\} \\ \leq \ & \min\left\{\alpha_{0.1}, \alpha_{1.0} + \alpha_{1.1}\right\} \end{aligned} \tag{6.42}$$

If $Y \overset{1}{\to} Z \overset{1}{\nrightarrow} X$, then $Y \overset{1}{\nrightarrow} (X,Z)$ is false, $(Y,Z) \overset{1}{\nrightarrow} X$ is true, and Test 2.0 represents a valid necessary and sufficient condition for non-causation $Y \overset{(2)}{\nrightarrow} X$,

hence

$$
\begin{aligned}
& P\left(rej.\ H_0^{(2)} | H_0^{(2)} \text{ is true}\right) && (6.43)\\
= \;& P\left((rej0.1 \cap rej0.2) \cap (rej1.0 \cup [fail1.0 \cap rej1.1 \cap rej1.2 \cap rej2.0]) | H_0^{(1)} \text{ is true}\right)\\
\leq \;& P\left(rej0.2 \cap (rej1.0 \cup [fail1.0 \cap rej1.1 \cap rej1.2 \cap rej2.0]) | H_0^{(1)} \text{ is true}\right)\\
\leq \;& \min\left\{\alpha_{0.2}, \alpha_{1.0} + P\left(rej1.1 \cap rej1.2 \cap rej2.0 | H_0^{(1)} \text{ is true}\right)\right\}\\
\leq \;& \min[\alpha_{0.2}, \alpha_{1.0} + \min\left\{\alpha_{1.2}, \alpha_{2.0}\right\}]
\end{aligned}
$$

If both $Y \overset{1}{\nrightarrow} Z \overset{1}{\nrightarrow} X$, then both $Y \overset{1}{\nrightarrow} (X, Z)$ and $(Y, Z) \overset{1}{\nrightarrow} X$ are true, and Test 2.0 does not present a necessary condition, and

$$
\begin{aligned}
& P\left(rej.\ H_0^{(2)} | H_0^{(2)} \text{ is true}\right) && (6.44)\\
= \;& P\left((rej0.1 \cap rej0.2) \cap (rej1.0 \cup [fail1.0 \cap rej1.1 \cap rej1.2 \cap rej2.0]) | H_0^{(1)} \text{ is true}\right)\\
\leq \;& \min[\alpha_{0.1}, \alpha_{0.2}, \alpha_{1.0} + \min\{\alpha_{1.1}, \alpha_{1.2}\}]
\end{aligned}
$$

Finally, if $Y \overset{1}{\rightarrow} Z \overset{1}{\rightarrow} X$ then both $Y \overset{1}{\nrightarrow} (X, Z)$ and $(Y, Z) \overset{1}{\nrightarrow} X$ are false, and

$$
\begin{aligned}
& P\left(rej.\ H_0^{(2)} | H_0^{(2)} \text{ is true}\right) && (6.45)\\
= \;& P\left((rej0.1 \cap rej0.2) \cap (rej1.0 \cup [fail1.0 \cap rej1.1 \cap rej1.2 \cap rej2.0]) | H_0^{(1)} \text{ is true}\right)\\
\leq \;& \alpha_{1.0} + \alpha_{2.0}.
\end{aligned}
$$

Repeating the above for each $H_0^{(h)}$, $h \geq 2$, gives the case-specific size bounds. □

References

[1] Amato, J.D. and N.R. Swanson, 2001, The Real Time Predictive Content of Money for Output, Journal of Monetary Economics 48, 3-24.

[2] Andrews, D.W.K., 2000, Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space, Econometrica 68, 399-405.

[3] Boudjellaba, H., J. Dufour and R. Roy, 1992, Testing Causality Between Two Vectors in Multivariate Autoregressive Moving Average Models, Journal of the American Statistical Society 87, 1082-1090.

[4] Boudjellaba, H., J. Dufour and R. Roy, 1994, Simplified Conditions for Non-causality Between Vectors in Multivariate ARMA Models, Journal of Econometrics 63, 271-287.

[5] Brooks, C., 1998, Predicting Stock Index Volatility: Can Market Volume Help? Journal of Forecasting 17, 59-80.

[6] Christiano, L.J., and L. Ljungqvist, 1988, Money does Granger-cause Output in the Bivariate Money-Output Relation, Journal of Monetary Economics 22, 217-235.

[7] Chao, J., Corradi, V. and N. Swanson, 2001, An Out of Sample Test for Granger Causality, Macroeconomic Dynamics 5, 598-620.

[8] Coe, P. J. and J. M. Nason, 2004, Long-Run Monetary Neutrality and Long-Horizon Regressions, Journal of Applied Econometrics 19, 355-373.

[9] Comte, F. and O. Lieberman, 2000, Second-Order Non-causality in Multivariate GARCH Processes, Journal of Time Series Analysis, 21, 535–557.

[10] Comte, F. and E. Renault, 1996, Noncausality in Continuous Time Models, Econometric Theory 12, 215-256.

[11] Corradi, V. and N. Swanson, 2002, A Consistent Test for Nonlinear Out of Sample Predictive Accuracy, Journal of Econometrics 110, 353-381.

[12] Dolado, J. and H. Lütkepohl, 1996, Making Wald Tests Work for Cointegrated VAR Systems, Econometric Reviews 15, 369-386.

[13] Dufour, J., 2002, Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics, Journal of Econometrics (forthcoming).

[14] Dufour, J. and T. Jouini, 2003, Finite-Sample Simulation-Based Tests in VAR Models with Applications to Order Selection and Causality Testing, Discussion paper, Centre Interuniversitaire de Recherche en Économie Quantitative (CIREQ), Université de Montréal.

[15] Dufour, J. and E. Renault, 1998, Short Run and Long Run Causality in Time Series: Theory, Econometrica 66, 1099-1125.

[16] Dufour, J. and D. Tessier, 1993, On the Relationship Between Impulse Response Analysis: Innovation Accounting and Granger Causality, Economics Letters 42, 327-333.

[17] Dufour, J., D. Pelletier and E. Renault, 2003, Short Run and Long Run Causality in Time Series: Inference, Journal of Econometrics (forthcoming).

[18] Florens, J.P. and M Mouchart, 1982, A Note on Noncausality, Econometrica 50, 583-606.

[19] Friedman, B. and K. Kuttner, 1992, Money, Income, Prices, and Interest Rates, American Economics Review 82, 472-492.

[20] Friedman, B. and K. Kuttner, 1993, Another Look at the Evidence on Money-Income Relationship, Journal of Economics 57, 189-203.

[21] Geiger, D. and J. Pearl, 1990, On the Logic of Causal Models, Uncertainty in Artificial Intelligence 4, 3-14.

[22] Granger, C.W.J., 1963, Economics Processes Involving Feedback, Information and Control 6, 28-48.

[23] Granger, C.W.J., 1969, Investigating Causal Relations by econometric Models and Cross-Spectral Methods, Econometrica 37, 424-438.

[24] Granger, C.W.J., 1980, Long-Memory Relationships and the Aggregation of Dynamic Models, Journal of Econometrics 14, 227-238.

[25] Granger, C.W.J., R. Robins and R. Engle, 1986, Wholesale and Retail Prices: Bivariate Time Series Modeling with Forecastable Error Variances. in Model Reliability, D.A. Belsey and E. Kuh (eds.) (MIT Press: Boston).

[26] Granger, C.W.J., 1988, Some Recent Developments in a Concept of Causality, Journal of Econometrics 39, 199-211.
[27] Hafer, R.W. and D.W. Jensen, 1991, The Demand for Money in the United States: Evidence from Cointegration Tests, Journal of Money, Credit and Banking 23, 155-168.
[28] Hill, Jonathan B., 2004, Consistent LM-Tests for Linearity Against Compound Smooth Transition Alternatives, Dept. of Economics, Florida International University, Working Paper #04-06; available at http://econwpa.wustl.edu:80/eps/em/papers/0402/0402004.pdf.
[29] Hiemstra, C. and J. D. Jones, 1994, Testing for Linear and Nonlinear Causality in the Stock Price-Volume Relation, Journal of Finance 49, 1639-1664.
[30] Hong, Yongmiao, 2001, A Test for Volatility Spillover with Application to Exchange Rates, Journal of Econometrics 103, 183-224.
[31] Lütkepohl, H., 1991, Introduction to Multiple Time Series Analysis (Springer-Verlag: Berlin).
[32] Lütkepohl, H., 1993, Testing for Causation Between Two Variables in Higher Dimensional VAR Models, in Studies in Applied Econometrics, ed. H. Schneeweiss and K. Zimmerman (Springer-Verlag: Heidelberg).
[33] Lütkepohl, H. and M. Müller, 1994, Testing for Multi-Step Causality in Time Series, Sonderforschungsbereich, Humboldt Universitaet Berlin.
[34] Lütkepohl, H. and D.S. Poskitt, 1996, Testing for Causality Using Infinite Order Vector Autoregressive Processes, Econometric Theory 12, 61-87.
[35] Renault, E. and A. Szafarz, 1991, True Versus Spurious Instantaneous Causality, Technical Report 9103, Centre d'Economie Mathématique, Université Libre de Bruxelles.
[36] Rothman, P., D. van Dijk and P.H. Franses, 2001, A multivariate STAR analysis of the relationship between money and output, Macroeconomic Dynamics 5, 506-532.
[37] Sims, C., 1972, Money, Income and Causality, American Economics Review 62, 540-552.
[38] Sims, C., 1980, Macroeconomics and Reality, Econometrica 48, 1-48.
[39] Sims, C., Stock, J.H., and M.W. Watson, 1990, Inference in Linear Time Series Models with Some Unit Roots, Econometrica 58, 113-144.
[40] Stock, J.H., and M.W. Watson, 1989, Interpreting the Evidence on Money-Income Causality, Journal of Econometrics 40, 161-181.
[41] Stock, J.H., and M.W. Watson, 1993, A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems, Econometrica 61, 783-820.
[42] Studený, M. and R.R. Bouckaert, 1998, On Chain Graph Models for Description of Conditional Independence Structures, Annals of Statistics 26, 1434-1495.
[43] Swanson, N., 1998, Money and Output Viewed through a Rolling Window, Journal of Monetary Economics 41, 455-473.
[44] Swanson, N. and C.W.J. Granger, 1997, Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization, Journal of the American Statistical Association 92, 357-367.
[45] Swanson, N., A. Ozyildirim and M. Pisu, 1996, A Comparison of Alternative Causality and Predictive Accuracy Tests in the Presence of Integrated and Cointegrated Economics Variables, Working Paper, Pennsylvania State University.
[46] Thoma, M.A., 1994, Subsample Instability and Asymmetries in Money-Income Causality, Journal of Econometrics 64, 279-306.
[47] Toda, H.Y. and P.C.B. Phillips, 1993, Vector Autoregressions and Causality, Econometrica 61, 1367-1393.
[48] Toda, H.Y. and P.C.B. Phillips, 1994, Vector Autoregressions and Causality: A Theoretical Overview and Simulation Study, Econometric Reviews 13, 259-285,
[49] Toda, H.Y., T. Yamamoto, 1995, Statistical Inference in Vector Autoregressions with Possibly Integrated Processes, Journal of Econometrics 66, 225-250.
[50] Wiener, N., 1956, The Theory of Prediction, in The Theory of Prediction, ed. E. F. Beckenback (McGraw-Hill: New York).

Dept. of Economics, Florida International University