Short and long run causality measures: theory and inference*

Jean-Marie Dufour † Abderrahim Taamouti ‡
Université de Montréal Université de Montréal

First version: October 2004
Revised: November 2005, August 2006
This version: December 27, 2008
Compiled: December 27, 2008, 5:08pm

* The authors thank Kilian Lutz, Benoit Perron, Nadège-Désirée Yameogo, the participants of 45th Annual Meeting of the Société canadienne de science économique at Charlevoix (2005), Joint Statistical Meetings of the American Statistical Association at Minneapolis (2005), 22nd CESG Conference at Vancouver (2005), CIREQ Time Series Conference at Montréal (2005), and 40th Annual Meeting of the Canadian Economics Association at Montréal (2006) for their helpful comments. This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal, and Chair in Environmental and Financial Econometric Analysis, Université Laval), the Alexander-von-Humboldt Foundation (Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec).

† William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

‡ Universidad Carlos III de Madrid, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 (514) 343 6111 ext. 5282; FAX: 1 (514) 343 5831; e-mail: abderrahim.taamouti@umontreal.ca .

ABSTRACT

The concept of causality introduced by Wiener (1956) and Granger (1969) is defined in terms of predictability one period ahead. This concept can be generalized by considering causality at a given horizon h, and causality up to any given horizon h [Dufour and Renault (1998)]. This generalization is motivated by the fact that, in the presence of an auxiliary variable vector Z, it is possible that a variable Y does not cause variable X at horizon 1, but causes it at horizon $h > 1$. In this case, there is an indirect causality transmitted by Z. Another related problem consists in measuring the importance of causality between two variables. Existing causality measures have been defined only for the horizon 1 and fail to capture indirect causal effects. This paper proposes a generalization of such measures for any horizon h. We propose nonparametric and parametric measures of unidirectional and instantaneous causality at any horizon h. Parametric measures are defined in the context of autoregressive processes of unknown order and expressed in terms of impulse response coefficients. On noting that causality measures typically involve complex functions of model parameters in VAR and VARMA models, we propose a simple method to evaluate these measures which is based on the simulation of a large sample from the process of interest. We also describe asymptotically valid nonparametric confidence intervals, using a bootstrap technique. Finally, the proposed measures are applied to study causality relations at different horizons between macroeconomic, monetary and financial variables in the U.S. These results show that there is a strong effect of nonborrowed reserves on federal funds rate one month ahead, the effect of real gross domestic product on federal funds rate is economically important for the first three months, the effect of federal funds rate on gross domestic product deflator is economically weak one month ahead, and finally federal funds rate causes the real gross domestic product until 16 months.

Keywords: time series; Granger causality; indirect causality; multiple horizon causality; causality measure; predictability; autoregressive model; vector autoregression; VAR; bootstrap; Monte Carlo; macroeconomics; money; interest rates; output; inflation.

Journal of Economic Literature classification: C1; C12; C15; C32; C51; C53; E3; E4; E52.

Contents

List of Tables

List of definitions and propositions

1. Introduction

The concept of causality introduced by Wiener (1956) and Granger (1969) is now a basic notion for studying dynamic relationships between time series. This concept is defined in terms of predictability at horizon one of a variable X from its own past, the past of another variable $Y,$ and possibly a vector Z of auxiliary variables.

The theory of Wiener-Granger causality has generated a considerable literature; for reviews, see Pierce and Haugh (1977), Newbold (1982), Geweke (1984*a*), Lütkepohl (1991), Boudjellaba, Dufour and Roy (1992, 1994) and Gouriéroux and Monfort (1997, Chapter 10). Most of the work in this field focus on predictability at horizon 1. In Dufour and Renault (1998), the concept of causality in the sense of Granger (1969) is generalized by considering causality at a given (arbitrary) horizon h and causality up to horizon $h,$ where h is a positive integer and can be infinite $(1 \leq h \leq \infty)$; for related work, see also Sims (1980), Hsiao (1982), and Lütkepohl (1993*b*). This generalization is motivated by the fact that, in the presence of auxiliary variables $Z,$ it is possible to have the variable Y not causing variable X at horizon one, but causing it at a longer horizon $h > 1.$ In this case, we have an indirect causality transmitted by the auxiliary variables $Z.$ Necessary and sufficient conditions of noncausality between vectors of variables at any horizon h for stationary and nonstationary processes are also supplied.

The analysis of Wiener-Granger distinguishes among three types of causality: from X to $Y,$ from Y to $X,$ and instantaneous causality. In practice, it is possible that these three types of causality coexist, hence the importance of finding means to measure their degree and determine the most important ones. Unfortunately, existing causality tests fail to accomplish this task, because they only inform us about the presence or the absence of causality. To answer this type of question, Geweke (1982, 1984*b*) has extended the causality concept by defining measures of causality and instantaneous effects, which can be decomposed in time and frequency domains. Gouriéroux, Monfort and Renault (1987) proposed causality measures based on the Kullback information. Polasek (1994) showed how causality measures can be calculated using the Akaike Information Criterion (AIC). Polasek (2002) also introduced new causality measures in the context of univariate and multivariate ARCH models and their extensions based on a Bayesian approach.

Existing causality measures have been established only for the one period horizon and fail to capture indirect causal effects. In this paper, we develop causality measures at different horizons which can detect indirect causality which becomes apparent only after several periods. Specifically, we propose generalizations to any horizon h of the measures proposed by Geweke (1982) for the horizon one. Both nonparametric and parametric measures of unidirectional causality and instantaneous effects at any horizon h are studied. Parametric measures are defined in terms of impulse response coefficients. By analogy with Geweke (1982, 1984*b*), we also define a measure of *dependence* at horizon $h,$ which combines causality measures from X to $Y,$ from Y to $X,$ and an instantaneous effect at horizon $h.$

After noting that analytical formulae for causality measures in VAR and VARMA models typically involve complex functions of model parameters and may be difficult to evaluate, we propose a simple method based on a long simulation of the process of interest and we show that the approach suggested works quite well in practice. For empirical implementation, we propose consistent estimators, derive their asymptotic distribution under standard regularity conditions, and suggest a bootstrap technique to build confidence intervals.

The proposed causality measures can be applied in different contexts and may help to solve some puzzles from the economic and financial literatures. In this paper, we illustrate their use by studying causality relations at different horizons between macroeconomic, monetary and financial variables in the U.S. The data set considered is the one used by Bernanke and Mihov (1998) and Dufour, Pelletier and Renault (2006). This data set consists of monthly observations on nonborrowed reserves, the federal funds rate, the gross domestic product deflator, and real gross domestic product.

The plan of the paper is as follows. Section 2 provides the motivation behind an extension of causality measures at horizon $h > 1$. Section 3 presents the framework allowing the definition of causality at different horizons. In Section 4, we propose nonparametric short-run and long-run causality measures. In Section 5, we give parametric expressions for the proposed causality measures in the context of linear stationary invertible processes, including VARMA processes. In Section 6, we propose consistent estimators of the causality measures. In Section 7, we suggest a simple method to evaluate the measures based on a simulation approach. In Section 8, we establish the asymptotic distribution of measures and the asymptotic validity of their nonparametric bootstrap confidence intervals. Section 9 is devoted to an empirical application and the conclusion relating to the results is given in Section 10. Proofs appear appendix.

2. Motivation

The causality measures proposed in this paper constitute extensions of those developed by Geweke (1982, 1984*b*, 1984*a*) and others. The existing causality measures quantify the effect of a vector of variables on another one at the one period horizon. The significance of such measures is however limited in the presence of auxiliary variables, since it is possible that a vector Y causes another vector X at an horizon h strictly higher than 1 even if there is no causality at horizon 1. In this case, we speak of an indirect effect induced by the auxiliary variables Z. Causality measures defined for the horizon 1 do not capture this indirect effect. This paper proposes causality measures at different horizons to quantify short- and long-run causality between random vectors. Such causality measures detect and quantify the indirect effects due to auxiliary variables. To see the importance of such causality measures, consider the following examples.

Example 2.1 Suppose we have two variables X and Y. $(X, Y)'$ follows a stationary VAR(1) model:

$$\begin{bmatrix} X(t+1) \\ Y(t+1) \end{bmatrix} = \begin{bmatrix} 0.5 & 0.7 \\ 0.4 & 0.35 \end{bmatrix} \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} + \begin{bmatrix} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \end{bmatrix}, \tag{2.1}$$

so that $X(t+1)$ is given by the equation

$$X(t+1) = 0.5\, X(t) + 0.7\, Y(t) + \varepsilon_X(t+1). \tag{2.2}$$

Since the coefficient of $Y(t)$ in (2.2) is equal to 0.7, we can conclude that Y causes X in the sense of Granger. However, this does not give any information on causality at horizons larger than 1 nor on its strength. To study causality at horizon 2, consider the system (2.1) at time $t+2$:

$$\left[\begin{array}{c} X(t+2) \\ Y(t+2) \end{array}\right] = \left[\begin{array}{cc} 0.53 & 0.595 \\ 0.34 & 0.402 \end{array}\right] \left[\begin{array}{c} X(t) \\ Y(t) \end{array}\right] + \left[\begin{array}{cc} 0.5 & 0.7 \\ 0.4 & 0.35 \end{array}\right] \left[\begin{array}{c} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \end{array}\right] + \left[\begin{array}{c} \varepsilon_X(t+2) \\ \varepsilon_Y(t+2) \end{array}\right].$$

In particular, $X(t+2)$ is given by

$$X(t+2) = 0.53\ X(t) + 0.595Y(t) + 0.5\varepsilon_X(t+1) + 0.7\varepsilon_Y(t+1) + \varepsilon_X(t+2)\,. \tag{2.3}$$

The coefficient of $Y(t)$ in equation (2.3) is equal to 0.595, so Y causes X at horizon 2. But, *how can one measure the importance of this "long-run" causality? Existing measures do not answer this question.*

Example 2.2 Suppose now that the information set contains not only the two variables of interest X and Y but also an auxiliary variable Z. Consider a trivariate stationary process $(X,Y,Z)'$ which follows a VAR(1) model:

$$\left[\begin{array}{c} X(t+1) \\ Y(t+1) \\ Z(t+1) \end{array}\right] = \left[\begin{array}{ccc} 0.6 & 0 & 0.8 \\ 0 & 0.4 & 0 \\ 0 & 0.6 & 0.1 \end{array}\right] \left[\begin{array}{c} X(t) \\ Y(t) \\ Z(t) \end{array}\right] + \left[\begin{array}{c} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \\ \varepsilon_Z(t+1) \end{array}\right] \tag{2.4}$$

hence

$$X(t+1) = 0.6\ X(t) + 0.8\ Z(t) + \varepsilon_X(t+1)\,. \tag{2.5}$$

Since the coefficient of $Y(t)$ in equation (2.5) is 0, we can conclude that Y does not cause X at horizon 1. If we consider model (2.4) at time $t+2$, we get:

$$\begin{aligned} \left[\begin{array}{c} X(t+2) \\ Y(t+2) \\ Z(t+2) \end{array}\right] &= \left[\begin{array}{ccc} 0.6 & 0 & 0.8 \\ 0 & 0.4 & 0 \\ 0 & 0.6 & 0.1 \end{array}\right]^2 \left[\begin{array}{c} X(t) \\ Y(t) \\ Z(t) \end{array}\right] \\ &\quad + \left[\begin{array}{ccc} 0.6 & 0 & 0.8 \\ 0 & 0.4 & 0 \\ 0 & 0.6 & 0.1 \end{array}\right] \left[\begin{array}{c} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \\ \varepsilon_Z(t+1) \end{array}\right] + \left[\begin{array}{c} \varepsilon_X(t+2) \\ \varepsilon_Y(t+2) \\ \varepsilon_Z(t+2) \end{array}\right], \end{aligned} \tag{2.6}$$

so that $X(t+2)$ is given by

$$\begin{aligned} X(t+2) &= 0.36\ X(t) + 0.48Y(t) + 0.56\ Z(t) + 0.6\varepsilon_X(t+1) \\ &\quad +0.8\varepsilon_Z(t+1) + \varepsilon_X(t+2). \end{aligned} \tag{2.7}$$

The coefficient of $Y(t)$ in equation (2.7) is equal to 0.48, which implies that Y causes X at horizon 2. This shows that the absence of causality at $h=1$ does not exclude the possibility of a causality at horizon $h > 1$. This indirect effect is transmitted by the variable Z:

$$Y \underbrace{\rightarrow}_{0.6} Z \underbrace{\rightarrow}_{0.8} X$$

where 0.60 and 0.80 are the coefficients of the one period effect of Y on Z and the one period effect of Z on $X,$ respectively. So, *how can one measure the importance of this indirect effect?* Again, *existing measures do not answer this question.*

3. Framework

The notion of noncausality studied here is defined in terms of orthogonality conditions between subspaces of a Hilbert space of random variables with finite second moments. We denote $L^2 \equiv L^2(\Omega, \mathcal{A}, Q)$ a Hilbert space of real random variables with finite second moments, defined on a common probability space $(\Omega, \mathcal{A}, Q),$ with covariance as the inner product. If E and F are two Hilbert subspaces of L^2, we denote $E + F$ the smallest subspace of L^2 which contains both E and F, while $E \backslash F$ represents the smallest Hilbert subspace of L^2 which contains the difference $E - F = E \cap F' = \{x : x \in E, x \notin F\}$.[If $E - F$ is empty, we set $E \backslash F = \{0\}$.]

"Information" is represented here by nondecreasing sequences of Hilbert subspaces of L^2. In particular, we consider a sequence I of "reference information sets" $I(t)$,

$$I = \{I(t) : t \in \mathbb{Z}\,,\, t > \omega\} \text{ with } t < t' \Rightarrow I(t) \subseteq I(t') \text{ for all } t > \omega\,, \tag{3.1}$$

where $I(t)$ is a Hilbert subspace of L^2, $\omega \in \mathbb{Z} \cup \{-\infty\}$ represents a "starting point", and $\mathbb{Z}$ is the set of the integers. The "starting point" ω is typically equal to a finite initial date (such as $\omega = -1$, 0 or 1) or to $-\infty$; in the latter case $I(t)$ is defined for all $t \in \mathbb{Z}$. We also consider three multivariate stochastic processes

$$X = \{X(t) : t \in \mathbb{Z},\, t > \omega\}\ ,\ \ Y = \{Y(t) : t \in \mathbb{Z},\, t > \omega\}\ ,\ Z = \{Z(t) : t \in \mathbb{Z},\, t > \omega\}\ , \tag{3.2}$$

where

$$\begin{aligned} X(t) &= (x_1(t), \ldots, x_{m_1}(t))',\ x_i(t) \in L^2, \quad i = 1, \ldots, m_1\,,\ m_1 \geq 1, \\ Y(t) &= (y_1(t), \ldots, y_{m_2}(t))',\ y_i(t) \in L^2, \quad i = 1, \ldots, m_2\,,\ m_2 \geq 1, \\ Z(t) &= (z_1(t), \ldots, z_{m_3}(t))',\ z_i(t) \in L^2, \quad i = 1, \ldots, m_3\,,\ m_3 \geq 0, \end{aligned}$$

and a (possibly empty) Hilbert subspace H of L^2, whose elements represent information available at any time, such as time independent variables (*e.g.*, the constant in a regression model) and deterministic processes (e.g., deterministic trends). We denote $X(\omega,\, t]$ the Hilbert space spanned by the components $x_i(\tau), i = 1, \ldots, m_1$, of $X(\tau)$, $\omega < \tau \leq t$, and similarly for $Y(\omega,\, t]$ and $Z(\omega,\, t]$: $X(\omega,\, t]$, $Y(\omega,\, t]$ and $Z(\omega,\, t]$ represent the information contained in the history of the variables X, Y and Z respectively up to time t. Finally, the information sets obtained by "adding" $X(\omega,\, t]$ to

$I(t)$ and $Y(\omega, t]$ to $I_X(t)$ are defined as

$$I_X(t) = I(t) + X(\omega, t]\,,\ I_{XY}(t) = I_X(t) + Y(\omega, t]\,, \tag{3.3}$$

and similarly for $I_Z(t)$, $I_Z(t)$, I_{XZ}, etc. In most cases considered below, the information set $I(t)$ contains $Z(\omega, t]$ but may not contain $X(\omega, t]$ or $Y(\omega, t]$.

For any information set B_t [some Hilbert subspace of L^2] and positive integer h, we denote $P[x_i(t+h)\,|\,B_t]$ the best linear forecast of $x_i(t+h)$ based on the information set B_t,

$$u[x_i(t+h)|\,B_t] = x_i(t+h) - P[x_i(t+h)\,|\,B_t]$$

the corresponding prediction error, and $\sigma^2[x_i(t+h)\,|\,B_t] = \mathsf{E}\big\{u[x_i(t+h)|\,B_t]^2\big\}$. Then, the best linear forecast of $X(t+h)$ is

$$P[X(t+h)|\,B_t] = \big(P[x_1(t+h)\,|\,B_t],\ \ldots\ ,\ P[x_{m_1}(t+h)\,|\,B_t]\big)',$$

the corresponding vector of prediction errors is

$$U[X(t+h)\,|\,B_t] = \big(u[x_1(t+h)\,|\,B_t]',\ \ldots\ ,\ u[x_{m_1}(t+h)\,|\,B_t]\big)', \tag{3.4}$$

and the corresponding matrix of second moments is

$$\varSigma[X(t+h)\,|\,B_t] = \mathsf{E}\big\{U[X(t+h)\,|\,B_t]\,U[X(t+h)\,|\,B_t]'\big\}. \tag{3.5}$$

Provided B_t contains a constant, $\varSigma[X(t+h)\,|\,B_t]$ is covariance matrix of $U[X(t+h)\,|\,B_t]$. Each component $P[x_i(t+h)\,|\,B_t]$ of $P[X(t+h)\,|\,B_t]$ is the orthogonal projection of $x_i(t+h)$ on the subspace B_t.

Following Dufour and Renault (1998), noncausality at horizon h is defined as follows, given an information set I.

Definition 3.1 NON-CAUSALITY AT HORIZON h. *For* $h \geq 1$,

(i) Y *does not cause* X *at horizon* h *given* I [*denoted* $Y \underset{h}{\nrightarrow} X\,|\,I$] *iff*

$$P[X(t+h)|\,I_X(t)] = P[X(t+h)\,|\,I_{XY}(t)],\ \ \forall t > \omega, \tag{3.6}$$

where $I_X(t) = I(t) + X(\omega, t]$ *and* $I_{XY}(t) = I_X(t) + Y(\omega, t]$;

(ii) Y *does not cause* X *up to horizon* h *given* I [*denoted* $Y \underset{(h)}{\nrightarrow} X\,|\,I$] *iff*

$$Y \underset{k}{\nrightarrow} X\,|\,I \text{ for } k = 1,\, 2,\, \ldots\ ,\, h; \tag{3.7}$$

(iii) Y *does not cause* X *at any horizon given* I [*denoted* $Y \underset{(\infty)}{\nrightarrow} X \mid I$] *iff*

$$Y \underset{k}{\nrightarrow} X \mid I \text{ for all } k = 1, 2, \ldots \tag{3.8}$$

This definition corresponds to causality from Y to X. It means that Y causes X at horizon h if the past of Y improves the forecast of $X(t+h)$ based on the information in $I(t)$ and $X(\omega, t]$. It is slightly more general than the one considered in Dufour and Renault (1998, Definition 2.2), because the conformability assumption $X(\omega, t] \subseteq I(t)$ is not imposed. But, clearly if $X(\omega, t] \subseteq I(t)$, then $I_X(t) = I(t)$. So, if the conformability assumption is added, Definition **3.1**(i) is equivalent to the one in Dufour and Renault (1998, Definition 2.2). Below, relaxing the assumption $X(\omega, t] \subseteq I(t)$ will facilitate the definition of causality measures. Given the above definition, the natural specification for $I(t)$ is one where $Z(\omega, t]$ is a subset of $I(t)$, but $X(\omega, t]$ and $Y(\omega, t]$ are not subsets of $I(t)$, *i.e.*

$$X(\omega, t] \nsubseteq I(t)\,,\; Y(\omega, t] \nsubseteq I(t)\,,\; Z(\omega, t] \subseteq I(t)\,. \tag{3.9}$$

An alternative characterization of noncausality can be expressed in terms of the variance-covariance matrix of the forecast errors. The following result is easily deduced from Definition **3.1**.

Proposition 3.2 COVARIANCE CHARACTERIZATION OF NON-CAUSALITY AT HORIZON h. *For* $h \geq 1$,

(i) Y *does not cause* X *at horizon* h *given* I *iff*

$$\det \varSigma[X(t+h) \mid I_X(t)] = \det \varSigma[X(t+h) \mid I_{XY}(t)]\,,\; \forall t > \omega,$$

where $\varSigma[X(t+h) \mid \cdot\,]$ *is defined by* (3.5);

(ii) Y *does not cause* X *up to horizon* h *given* I *iff*

$$\det \varSigma[X(t+k) \mid I_X(t)] = \det \varSigma[X(t+k) \mid I_{XY}(t)]\,,\; \forall t > \omega,\; k = 1,\; 2, \ldots, h\,;$$

(iii) Y *does not cause* X *at any horizon given* I_X, *iff*

$$\det \varSigma[X(t+k) \mid I_X(t)] = \det \varSigma[X(t+k) \mid I_{XY}(t)]\,,\; \forall t > \omega,\; k = 1, 2, \ldots.$$

Below, we also consider unconditional causality properties induced by eliminating the auxiliary variable vector Z from the information set. This suggests considering $Z-$*unconditional non-causality* which is defined as follows.

Definition 3.3 UNCONDITIONAL NON-CAUSALITY AT HORIZON h. *For* $h \geq 1$,

(i) *Y does not cause X at horizon h given I, unconditionally with respect to Z [denoted $Y \underset{h}{\nrightarrow} X \,|\, I_{(Z)}$] iff*

$$P[X(t+h)|\, I_{(Z)X}(t)] = P[X(t+h)|\, I_{(Z)XY}(t)],\ \forall t > \omega,$$

where $I_{(Z)X}(t) = I_{(Z)}(t) + X(\omega, t]$, $I_{(Z)XY}(t) = I_{(Z)X}(t) + Y(\omega, t]$ and $I_{(Z)}(t) = I(t) \backslash Z(\omega, t]$;

(ii) *Y does not cause X up to horizon h given I, unconditionally with respect to Z [denoted $Y \underset{(h)}{\nrightarrow} X \,|\, I_{(Z)}$] iff*

$$Y \underset{k}{\nrightarrow} X \,|\, I_{(Z)} \text{ for } k = 1,\, 2,\, \ldots\, ,\, h;$$

(iii) *Y does not cause X at any horizon given I, unconditionally with respect to Z [denoted $Y \underset{(\infty)}{\nrightarrow} X \,|\, I_{(Z)}$] iff*

$$Y \underset{k}{\nrightarrow} X \,|\, I_{(Z)} \text{ for all } k = 1,\, 2,\, \ldots$$

If Z is empty $(m_3 = 0)$, there is no effective conditioning and we use the conventions $I_{(Z)X}(t) = I_X(t)$ and $I_{(Z)XY}(t) = I_{XY}(t)$. On replacing I by $I_{(Z)}$, it is straightforward to see that Proposition **3.2** also holds for $Z-$unconditional non-causality.

4. Causality measures

We will now develop extensions of the causality measures introduced by Geweke (1982, 1984*b*, 1984*a*) for the horizon 1. Important properties of these measures include: (1) they are nonnegative, and (2) they cancel only when there is no causality at the horizon considered. Specifically, we propose the following causality measures at horizon $h \geq 1$, where by convention $\ln(0/0) = 0$ and $\ln(x/0) = +\infty$ for $x > 0$.

Definition 4.1 MEAN-SQUARE CAUSALITY MEASURE AT HORIZON h RELATIVE TO AN INFORMATION SET. *For $h \geq 1$,*

$$C_L(Y \underset{h}{\rightarrow} X \,|\, I) = \ln \left[\frac{\det \Sigma[X(t+h) \,|\, I_X(t)]}{\det \Sigma[X(t+h) \,|\, I_{XY}(t)]} \right] \tag{4.1}$$

is the mean-square causality measure [alt., *the* intensity *of the causality*] *from Y to X at horizon h, given I.*

Since we consider here only mean-square measures, the term "mean square causality measure" will be abbreviated to "causality measure". Clearly, $C_L(Y \underset{h}{\rightarrow} X \,|\, I) = 0$ if $Y(\omega, t] \subseteq I_X(t)$, so $C_L(Y \underset{h}{\rightarrow} X \,|\, I)$ provides useful information mainly when $Y(\omega, t] \nsubseteq I(t)$. For $m_1 = m_2 = 1$, Definition **4.1** reduces to

$$C_L(Y \underset{h}{\rightarrow} X \,|\, I) = \ln\left[\frac{\sigma^2[X(t+h)\,|I_X(t)]}{\sigma^2[X(t+h)\,|\,I_{XY}(t)]}\right].$$

$C_L(Y \underset{h}{\rightarrow} X \,|\, I)$ measures the causal effect from Y to X at horizon h given I and the past of X. In terms of predictability, this can be viewed as the amount of information brought by the past of Y which can improve the forecast of $X(t+h)$. Following Geweke (1982), this measure can be also interpreted as the proportional reduction in the variance of the forecast error of $X(t+h)$ obtained by taking into account the past of Y. This proportion is equal to:

$$\frac{\sigma^2[X(t+h)\,|\,I_X(t)] - \sigma^2[X(t+h)\,|\,I_{XY}(t)]}{\sigma^2[X(t+h)\,|\,I_X(t)]} = 1 - \exp[-C_L(Y \underset{h}{\rightarrow} X \,|\, I)].$$

It can be useful to consider unconditional causality properties induced by eliminating the auxiliary variable vector Z from the information set. Such unconditional causality measures can be defined as follows.

Definition 4.2 UNCONDITIONAL MEAN-SQUARE CAUSALITY MEASURE AT HORIZON h. *For* $h \geq 1$,

$$C_L(Y \underset{h}{\rightarrow} X \,|\, I_{(Z)}) = \ln\left[\frac{\det \varSigma[X(t+h)\,|\,I_{(Z)X}(t)]}{\det \varSigma[X(t+h)\,|\,I_{(Z)XY}(t)]}\right] \tag{4.2}$$

is the Z-unconditional mean-square causality measure *from* Y *to* X *at horizon* h, *given* I.

When there is no ambiguity concerning the reference information I, we shall also use the more intuitive notation:

$$C(X \underset{h}{\rightarrow} Y \,|\, Z) = C_L(Y \underset{h}{\rightarrow} X \,|\, I_{(Z)}). \tag{4.3}$$

As in Geweke (1984*b*), we can rewrite the (conditional) causality measures given by Definition **4.1** in terms of unconditional causality measures where Z is eliminated form the reference information set:

$$\begin{aligned} C_L(X \underset{h}{\rightarrow} Y \,|\, I) &= C_L((Y,\, Z) \underset{h}{\rightarrow} X \,|\, I_{(Z)X}) - C_L(Z \underset{h}{\rightarrow} X \,|\, I_{(Z)X}) \\ &= C((Y,\, Z) \underset{h}{\rightarrow} X \,|\, Z) - C(Z \underset{h}{\rightarrow} X \,|\, Z)\,, \end{aligned} \tag{4.4}$$

$$\begin{aligned} C_L(X \underset{h}{\rightarrow} Y \,|\, I) &= C_L((X,\, Z) \underset{h}{\rightarrow} Y \,|\, I_{(Z)Y}) - C_L(Z \underset{h}{\rightarrow} Y \,|\, I_{(Z)Y}) \\ &= C((Y,\, Z) \underset{h}{\rightarrow} X \,|\, Z) - C(Z \underset{h}{\rightarrow} X \,|\, Z)\,, \end{aligned} \tag{4.5}$$

where $(Y,\, Z)$ and $(X,\, Z)$ represent the joint process $\{\big(X(t)',\, Z(t)'\big)' : t \in \mathbb{Z},\, t > \omega\}$ and $\{\big(Y(t)',\, Z(t)'\big)' : t \in \mathbb{Z}, t > \omega\}$.

We now define an instantaneous causality measure between X and Y at horizon h as follows.

Definition 4.3 MEASURE OF INSTANTANEOUS CAUSALITY AT HORIZON h. *For* $h \geq 1$,

$$C_L(X \underset{h}{-} Y \,|\, I) = \ln\left[\frac{\det \Sigma[X(t+h) \,|\, I_{XY}(t)] \ \det \Sigma[Y(t+h) \,|\, I_{XY}(t)]}{\det \Sigma[X(t+h),\, Y(t+h) \,|\, I_{XY}(t)]}\right],$$

where $\Sigma[X(t+h),\, Y(t+h) \,|\, I_{XY}(t)] = \mathsf{E}\{U[W(t+h) \,|\, I_{XY}(t)]U[W(t+h) \,|\, I_{XY}(t)]'\}$ *and* $W(t) = \left(X(t)',\, Y(t)'\right)'$, *is the* mean-square instantaneous causality measure [*alt.,the* intensity *of the instantaneous causality*] *between* Y *and* X *at horizon* h.

For $m_1 = m_2 = 1$ and provided $I(t)$ includes a constant variable, we have:

$$\begin{aligned} \det \Sigma[\left(X(t+h),\, Y(t+h)\right) \,|\, I_{XY}(t)] &= \sigma^2[X(t+h) \,|\, I_{XY}(t)]\, \sigma^2[Y(t+h) \,|\, I_{XY}(t)] \\ &\quad - \left(\mathsf{cov}[(X(t+h),\, Y(t+h) \,|\, I_{XY}(t)])\right)^2, \end{aligned} \tag{4.6}$$

so that

$$\begin{aligned} C_L(X \underset{h}{-} Y \,|\, I) &= \ln\left[\frac{1}{1 - \rho[X(t+h),\, Y(t+h) \,|\, I_{XY}(t)]^2}\right] \\ &= \ln\left[\frac{\sigma^2[X(t+h) \,| I_{XY}(t)]}{\sigma^2[X(t+h) \,|\, I_{XY}(t) + I_{Y(t+h)}]}\right] \\ &= \ln\left[\frac{\sigma^2[Y(t+h) \,| I_{XY}(t)]}{\sigma^2[Y(t+h) \,|\, I_{XY}(t) + I_{X(t+h)}]}\right] \end{aligned} \tag{4.7}$$

where

$$\rho[X(t+h),\, Y(t+h) \,|\, I_{XY}(t)] = \frac{\mathsf{cov}[X(t+h),\, Y(t+h) \,|\, I_{XY}(t)]}{\sigma[X(t+h) \,|\, I_{XY}(t)]\sigma[Y(t+h) \,|\, I_{XY}(t)]} \tag{4.8}$$

is the conditional correlation coefficient between $X(t+h)$ and $Y(t+h)$ given the information set $I_{XY}(t)$, $I_{Y(t+h)}$ represents the Hilbert subspace spanned by the components of $Y(t+h)$ and similarly for $I_{X(t+h)}$. Thus, instantaneous causality increases with the absolute value of the conditional correlation coefficient.

We also define a measure of dependence between X and Y at horizon h. This will enable one to check whether, at a given horizon h, the processes X and Y must be considered together or whether they can be treated separately.

Definition 4.4 DEPENDENCE MEASURE AT HORIZON h. *For* $h \geq 1$,

$$C_L^{(h)}(X,\, Y \,|\, I) = C_L(X \underset{h}{\rightarrow} Y \,|\, I) + C_L(Y \underset{h}{\rightarrow} X \,|\, I) + C_L(X \underset{h}{-} Y \,|\, I) \tag{4.9}$$

is the intensity *of the dependence between* X *and* Y *at horizon* h, *given* I.

It is easy to see that the intensity of the dependence between X and Y at horizon h can be written in the alternative form:

$$C_L^{(h)}(X, Y \mid I) = \ln\left[\frac{\det \Sigma[X(t+h) \mid I_X(t)] \ \det \Sigma[Y(t+h) \mid I_Y(t)]}{\det \Sigma[X(t+h),\, Y(t+h) \mid I_{XY}(t)]}\right]. \tag{4.10}$$

When there is no ambiguity on the definition of the reference information set $I(t)$, we shall also use the following notations:

$$C(Y \underset{h}{\rightarrow} X) = C_L(Y \underset{h}{\rightarrow} X \mid I)\,,\ C(Y \underset{h}{\rightarrow} X \mid Z) = C_L(Y \underset{h}{\rightarrow} X \mid I_{(Z)}), \tag{4.11}$$

$$C(X \underset{h}{-} Y) = C_L(X \underset{h}{-} Y \mid I)\,,\ C^{(h)}(X,\, Y) = C_L^{(h)}(X,\, Y \mid I)\,. \tag{4.12}$$

Now, it is possible to build a recursive formulation of causality measures. This one will depend on the predictability measure introduced by Diebold and Kilian (2001). These authors proposed a predictability measure based on the ratio of expected losses of short and long run forecasts:

$$\bar{P}(L,\, \Omega_t,\, j,\, k) = 1 - \frac{\mathsf{E}\big[L(U[X(t+j) \mid \Omega_t])\big]}{\mathsf{E}\big[L(U[X(t+k) \mid \Omega_t])\big]}$$

where Ω_t is the information set at time t, L is a loss function, j and k represent respectively the short and the long-run, $e_{t+s,\,t} = X(t+s) - P[X(t+s) \mid \Omega_t]$, for $s = j, k,$ is the forecast error at horizon $t+s$. This predictability measure can be constructed according to the horizons of interest and it allows for general loss functions as well as univariate or multivariate information sets. In this paper, we focus on the case of a quadratic loss function,

$$L(e_{t+s,\,t}) = U[X(t+s) \mid \Omega_t]^2, \text{ for } s = j,\, k.$$

Then, we have the following relationships.

Proposition 4.5 RELATION OF CAUSALITY MEASURES WITH PREDICTABILITY MEASURES. *Let h_1 and h_2 be two different horizons, $m_1 = m_2 = 1$, and*

$$\bar{P}_X\big(I_X(t),\, h_1,\, h_2\big) = 1 - \frac{\sigma^2\big(X(t+h_1) \mid I_X(t)\big)}{\sigma^2\big(X(t+h_2) \mid I_X(t)\big)},$$

$$\bar{P}_X\big(I_{XY}(t),\, h_1,\, h_2\big) = 1 - \frac{\sigma^2\big(X(t+h_1) \mid I_{XY}(t)\big)}{\sigma^2\big(X(t+h_2) \mid I_{XY}(t)\big)},$$

the predictability measures for X based on the information sets $I_X(t)$ and $I_{XY}(t)$. Then, for $h_2 > h_1 \geq 1$,

$$C_L(Y \underset{h_1}{\rightarrow} X \mid I) - C_L(Y \underset{h_2}{\rightarrow} X \mid I) = \ln\{1 - \bar{P}_X[I_X(t),\, h_1,\, h_2]\} - \ln\{1 - \bar{P}_X[I_{XY}(t),\, h_1,\, h_2]\}\,.$$

The following identity follows immediately from the latter proposition: for $h \geq 2$ and $m_1 = m_2 = 1$,

$$C_L(Y \underset{h}{\rightarrow} X \mid I) = C_L(Y \underset{1}{\rightarrow} X \mid I) + \ln[1 - \bar{P}_X\big(I_X(t) + Y(\omega, t], 1, h\big)] - \ln[1 - \bar{P}_X\big(I_X(t), 1, h\big)] .$$

Predictability measures look at the effect of changing the forecast horizon, for a *given information set*, while causality measures look at the joint effect of changing the information set and the forecast horizon.

5. Causality measures for VARMA models

We now consider a more specific set of linear invertible processes which includes vector autoregressive (VAR), moving average (VMA), and mixed (VARMA) models of finite order as special cases. It is possible to provide parametric expressions for short-run and long-run causality measures in terms of impulse response coefficients.

We consider in turn two distinct cases. First, we calculate parametric measures of short-run and long-run causality in the context of an autoregressive moving average model. We assume that the process $\{W(s) = (X(s)', Y(s)', Z(s)')' : s \leq t\}$ is a VARMA(p, q) model, hereafter the unconstrained model, where p and q can be infinite. The structure of the process $\{W_0(s) = (X(s)^{'}, Z(s)')^{'} : s \leq t\}$, hereafter the *constrained* model, can be deduced from the unconstrained model using Corollary 6.1.1 in Lütkepohl (1993*b*). This model is a VARMA$(\bar{p}, \bar{q})$ with $\bar{p} \leq mp$ and $\bar{q} \leq (m-1)p + q$. Second, we provide a characterization of the parametric causality measures in the context of VMA(q) model, where q is finite.

5.1. Parametric causality measures in the context of a VARMA(p, q) process

Without loss of generality, let us consider the discrete $m \times 1$ vector process with zero mean $\{W(s) = (X(s)^{'}, Y^{'}(s), Z(s)')^{'} : s \leq t\}$ defined on L^2 and characterized by the following autoregressive moving average representation:

$$W(t) = \sum_{i=1}^{p} \Phi_i W(t-i) + \sum_{j=1}^{q} \Theta_j u(t-j) + u(t) \tag{5.1}$$

where $m = m_1 + m_2 + m_3$ and

$$\Phi_i = \begin{bmatrix} \varphi_{XXi} & \varphi_{XYi} & \varphi_{XZi} \\ \varphi_{YXi} & \varphi_{YYi} & \varphi_{YZi} \\ \varphi_{ZXi} & \varphi_{ZYi} & \varphi_{ZZi} \end{bmatrix}, \quad \Theta_j = \begin{bmatrix} \theta_{XXj} & \theta_{XYj} & \theta_{XZj} \\ \theta_{YXj} & \theta_{YYj} & \theta_{YZj} \\ \theta_{ZXj} & \theta_{ZYj} & \theta_{ZZj} \end{bmatrix}, \tag{5.2}$$

$$\mathsf{E}\left[u(t)\right] = 0, \ \mathsf{E}\big[u(t)u(s)^{'}\big] = \begin{cases} \Sigma_u, & \text{for } s = t \\ 0, & \text{for } s \neq t\,. \end{cases}$$

More compactly,

$$\Phi(L)W(t) = \Theta(L)u(t) \tag{5.3}$$

with

$$\Phi(L)= \begin{bmatrix} \varphi_{XX}(L) & \varphi_{XY}(L) & \varphi_{XZ}(L) \\ \varphi_{YX}(L) & \varphi_{YY}(L) & \varphi_{YZ}(L) \\ \varphi_{ZX}(L) & \varphi_{ZY}(L) & \varphi_{ZZ}(L) \end{bmatrix}, \quad \Theta(L)= \begin{bmatrix} \theta_{XX}(L) & \theta_{XY}(L) & \theta_{XZ}(L) \\ \theta_{YX}(L) & \theta_{YY}(L) & \theta_{YZ}(L) \\ \theta_{ZX}(L) & \theta_{ZY}(L) & \theta_{ZZ}(L) \end{bmatrix},$$

$$\varphi_{ll}(L) = I_{m_l} - \sum_{i=1}^{p} \varphi_{lli} L^i, \ \varphi_{lk}(L) = -\sum_{i=1}^{p} \varphi_{lki} L^i,$$

$$\theta_{ll}(L) = I_{m_l} + \sum_{j=1}^{q} \theta_{llj} L^j, \ \theta_{lk}(L) = \sum_{j=1}^{q} \theta_{lkj} L^j, \ \text{ for } l \neq k, \ l, \, k = X, \ Y, \ Z.$$

We assume that $u(t)$ is orthogonal to the Hilbert subspace spanned $\{W(s) : \, s \leq (t-1)\}$ with Σ_u is symmetric positive definite matrix. Under stationarity, $W(t)$ has a VMA(∞) representation:

$$W(t) = \Psi(L)u(t) \tag{5.4}$$

where

$$\Psi(L) = \Phi(L)^{-1}\Theta(L) = \sum_{j=0}^{\infty} \Psi_j L^j = \sum_{j=0}^{\infty} \begin{bmatrix} \psi_{XXj} & \psi_{XYj} & \psi_{XZj} \\ \psi_{YXj} & \psi_{YYj} & \psi_{YZj} \\ \psi_{ZXj} & \psi_{ZYj} & \psi_{ZZj} \end{bmatrix} L^j, \ \Psi_0 = I_m.$$

From the previous section, measures of dependence and causality are defined in terms of variance-covariance matrices of the constrained and unconstrained forecast errors. Thus, to calculate these measures, we need to know the structure of the constrained model (imposing non-causality). This one can be deduced from the structure of the unconstrained model (5.1) using the following proposition and corollary [Lütkepohl (1993*b*, pages 231-232)].

Lemma 5.1 LINEAR TRANSFORMATION OF A $\text{VMA}(q)$ PROCESS. *Let $u(t)$ be a K-dimensional white noise process with nonsingular variance-covariance matrix Σ_u and let*

$$W(t) = \mu + \sum_{j=1}^{q} \Psi_j u(t-j) + u(t)$$

be a K-dimensional invertible VMA(q) process. Furthermore, let F be an $(M \times K)$ matrix of rank M. Then the M-dimensional process $W_0(t) = FW(t)$ has an invertible VMA($\bar{q}$) representation:

$$W_0(t) = F\mu + \sum_{j=1}^{\bar{q}} \bar{\theta}_j \varepsilon(t-j) + \varepsilon(t)$$

where $\varepsilon(t)$ is M-dimensional white noise with nonsingular variance-covariance matrix Σ_ε, the $\bar{\theta}_j$,

$j = 1, \ldots, \bar{q}$, *are* $M \times M$ *coefficient matrices and* $\bar{q} \leq q$.

Lemma 5.2 LINEAR TRANSFORMATION OF A VARMA(p, q) PROCESS. *Let* $W(t)$ *be a* K*-dimensional, stable, invertible VARMA*(p, q) *process and let* F *be an* $M \times K$ *matrix of rank* M. *Then the process* $W_0(t) = FW(t)$ *has a VARMA*$(\bar{p}, \bar{q})$ *representation with*

$$\bar{p} \leq Kp, \;\; \bar{q} \leq (K-1)p + q.$$

If we assume that $W(t)$ follows a VAR(p) [or VARMA$(p, 0)$] model, then its linear transformation $W_0(t) = FW(t)$ has a VARMA$(\bar{p}, \bar{q})$ representation with $\bar{p} \leq Kp$ and $\bar{q} \leq (K-1)p$. Suppose now that we are interested in measuring the causality from Y to X at a given horizon h. We need to apply Lemma **5.2** to obtain the structure of process $\{W_0(s) = (X(s)', Z(s)')' : s \leq t\}$. If we left-multiply equation (5.3) by the adjoint matrix of $\Phi(L)$, denoted $\Phi(L)^*$, we get

$$\Phi(L)^* \Phi(L) W(t) = \Phi(L)^* \Theta(L) u(t) \tag{5.5}$$

where $\Phi(L)^* \Phi(L) = \det[\Phi(L)]$. Since the determinant of $\Phi(L)$ is a sum of products involving one operator from each row and each column of $\Phi(L)$, the degree of the VAR polynomial, here $\det[\Phi(L)]$, is at most mp. We write:

$$\det[\Phi(L)] = 1 - \alpha_1 L - \cdots - \alpha_{\bar{p}} L^{\bar{p}}$$

where $\bar{p} \leq mp$. It is also easy to check that the degree of the operator $\Phi(L)^* \Theta(L)$ is at most $p(m-1) + q$. Thus, equation (5.5) can be written as follows:

$$\det[\Phi(L)] \, W(t) = \Phi(L)^* \Theta(L) u(t). \tag{5.6}$$

This equation is another stationary invertible VARMA representation of process $W(t)$, called the final equation form. The model of the process $\{W_0(s) = (X(s)', \; Z(s)')' : s \leq t\}$ can be obtained by choosing

$$F = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & 0 & I_{m3} \end{bmatrix}.$$

On premultiplying (5.6) by F, we get

$$\det[\Phi(L)] \, W_0(t) = F\Phi(L)^* \Theta(L) u(t). \tag{5.7}$$

The right-hand side of (5.7) is a linearly transformed finite-order VMA process which, by Lemma **5.1**, has a VMA$(\bar{q})$ representation with $\bar{q} \leq p(m-1) + q$. Thus, we get the model:

$$\det[\Phi(L)] \, W_0(t) = \bar{\theta}(L)\varepsilon(t) = \begin{bmatrix} \bar{\theta}_{XX}(L) & \bar{\theta}_{XZ}(L) \\ \bar{\theta}_{ZX}(L) & \bar{\theta}_{ZZ}(L) \end{bmatrix} \varepsilon(t) \tag{5.8}$$

where

$$\mathsf{E}[\varepsilon(t)] = 0, \;\; \mathsf{E}\big[\varepsilon(t)\varepsilon(s)'\big] = \begin{cases} \Sigma_\varepsilon & \text{for } s = t \\ 0 & \text{for } s \neq t \end{cases},$$

$$\bar{\theta}_{ll}(L) = I_{m_l} + \sum_{j=1}^{\bar{q}} \bar{\theta}_{llj} L^j,\ \bar{\theta}_{lk}(L) = \sum_{j=1}^{\bar{q}} \bar{\theta}_{lkj} L^j,\ \text{ for } l \neq k,\ \ l,\, k = X,\ Z.$$

Note that, in theory, the coefficients $\bar{\theta}_{lkj}$ and elements of the variance-covariance matrix Σ_ε, can be computed from coefficients φ_{lki}, Θ_{lkj}, $l, k = X,\ Z,\ Y$, $i = 1, \ldots, p$, $j = 1, \ldots, q$, and elements of the variance-covariance matrix Σ_u. This is possible by solving the following system:

$$\gamma_\varepsilon(v) = \gamma_u(v),\ \ v = 0,\ 1,\ 2,\ \ldots \tag{5.9}$$

where $\gamma_\varepsilon(v)$ and $\gamma_u(v)$ are the autocovariance functions of the processes $\bar{\theta}(L)\varepsilon(t)$ and $F\Phi(L)^*\Theta(L)u(t)$, respectively. The following example shows how one can calculate the theoretical parameters of the constrained model in terms of those of the unconstrained model in the context of a bivariate VAR(1) model.

Example 5.3 Consider the following bivariate VAR(1) model:

$$\begin{aligned}\begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} &= \begin{bmatrix} \varphi_{XX} & \varphi_{XY} \\ \varphi_{YX} & \varphi_{YY} \end{bmatrix} \begin{bmatrix} X(t-1) \\ Y(t-1) \end{bmatrix} + \begin{bmatrix} u_X(t) \\ u_Y(t) \end{bmatrix} \\ &= \Phi \begin{bmatrix} X(t-1) \\ Y(t-1) \end{bmatrix} + u(t).\end{aligned} \tag{5.10}$$

We assume that all the roots of $\det[\Phi(z)] = \det[I_2 - \Phi z]$ are outside of the unit circle. Under this assumption, model (5.10) has the following VMA(∞) representation:

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \sum_{j=0}^{\infty} \Psi_j \begin{pmatrix} u_X(t-j) \\ u_Y(t-j) \end{pmatrix} = \sum_{j=0}^{\infty} \begin{bmatrix} \psi_{XXj} & \psi_{XYj} \\ \psi_{YXj} & \psi_{YYj} \end{bmatrix} \begin{pmatrix} u_X(t-j) \\ u_Y(t-j) \end{pmatrix}$$

where$\Psi_j = \Phi^j$. If we are interested in determining the model for the marginal process $X(t)$, then by Lemma **5.2** and for $F = [1,\ 0]$, we have

$$\det[\Phi(L)] X(t) = [1,\ 0]\, \Phi(L)^* u(t)$$

where

$$\Phi(L)^* = \begin{bmatrix} 1 - \varphi_{YY} L & \varphi_{XY} L \\ \varphi_{YX} L & 1 - \varphi_{XX} L \end{bmatrix}$$

and

$$\det[\Phi(L)] = 1 - (\varphi_{YY} + \varphi_{XX})L - (\varphi_{YX}\varphi_{XY} - \varphi_{XX}\varphi_{YY})L^2.$$

Thus,

$$X(t) - \varphi_1 X(t-1) - \varphi_2 X(t-2) = \varphi_{XY} u_Y(t-1) - \varphi_{YY} u_X(t-1) + u_X(t). \tag{5.11}$$

where $\varphi_1 = \varphi_{YY} + \varphi_{XX}$ and $\varphi_2 = \varphi_{YX}\varphi_{XY} - \varphi_{XX}\varphi_{YY}$. The right-hand side of equation (5.11), denoted $\varpi(t)$, is the sum of an MA(1) process and a white noise process. By Lemma **5.1**,

$\varpi(t)$ has an MA(1) representation, $\varpi(t) = \varepsilon_X(t) + \bar{\theta}\varepsilon_X(t-1)$. To determine parameters $\bar{\theta}$ and $\mathsf{V}(\varepsilon_X(t)) = \sigma^2_{\varepsilon_X}$ in terms of the parameters of the unconstrained model, we can solve system (5.9) for $v = 0$ and $v = 1$,

$$\begin{aligned} \mathsf{V}\left[\varpi(t)\right] &= \mathsf{V}\left[u_X(t) - \varphi_{YY}u_X(t-1) + \varphi_{XY}u_Y(t-1)\right], \\ \mathsf{E}\left[\varpi(t)\varpi(t-1)\right] &= \mathsf{E}[(u_X(t) - \varphi_{YY}u_X(t-1) + \varphi_{XY}u_Y(t-1)) \\ &\quad \times (u_X(t-1) - \varphi_{YY}u_X(t-2) + \varphi_{XY}u_Y(t-2))], \end{aligned}$$

which is equivalent to solve the following system:

$$(1+\bar{\theta}^2)\sigma^2_{\varepsilon_X} = (1+\varphi^2_{YY})\sigma^2_{u_X} + \varphi^2_{XY}\sigma^2_{u_Y} - 2\varphi_{YY}\varphi_{XY}\sigma_{u_Y u_X} \quad \text{and} \quad \bar{\theta}\sigma^2_{\varepsilon_X} = -\varphi_{YY}\sigma^2_{u_X}.$$

Here we have two equations and two unknown parameters $\bar{\theta}$ and $\sigma^2_{\varepsilon_X}$. These parameters must satisfy the constraints $|\,\bar{\theta}\,| < 1$ and $\sigma^2_{\varepsilon_X} > 0$.

The VMA(∞) representation of model (5.8) is given by

$$\begin{aligned} W_0(t) &= \det\left[\Phi(L)\right]^{-1}\bar{\theta}(L)\varepsilon(t) = \sum_{j=0}^{\infty}\bar{\Psi}_j\varepsilon(t-j) \\ &= \sum_{j=0}^{\infty}\begin{bmatrix} \bar{\psi}_{XX_j} & \bar{\psi}_{XZ_j} \\ \bar{\psi}_{ZX_j} & \bar{\psi}_{ZZ_j} \end{bmatrix}\begin{bmatrix} \varepsilon_X(t-j) \\ \varepsilon_Z(t-j) \end{bmatrix} \end{aligned} \tag{5.12}$$

where $\bar{\Psi}_0 = I_{m_1+m_2}$. To quantify the degree of causality from Y to X at horizon h, we first consider the unconstrained and constrained models of process X. The unconstrained model is

$$X(t) = \sum_{j=1}^{\infty}\psi_{XXj}u_X(t-j) + \sum_{j=1}^{\infty}\psi_{XYj}u_Y(t-j) + \sum_{j=1}^{\infty}\psi_{XZj}u_Z(t-j) + u_X(t),$$

whereas the constrained model is

$$X(t) = \sum_{j=1}^{\infty}\bar{\psi}_{XXj}\varepsilon_X(t-j) + \sum_{j=1}^{\infty}\bar{\psi}_{XZj}\varepsilon_Z(t-j) + \varepsilon_X(t).$$

Second, we need to calculate the variance-covariance matrices of the unconstrained and constrained forecast errors of $X(t+h)$. From (5.4), the forecast error of $W(t+h)$ is given by

$$U\left[W(t+h)\,|\,I_W(t)\right] = \sum_{i=0}^{h-1}\Psi_i u(t+h-i)$$

so that

$$\Sigma[W(t+h)\,|\,I_W(t)] = \sum_{i=0}^{h-1}\Psi_i\,\mathsf{V}\left[u(t)\right]\,\Psi_i' = \sum_{i=0}^{h-1}\Psi_i\,\Sigma_u\,\Psi_i'. \tag{5.13}$$

The unconstrained forecast error of $X(t+h)$ is given by

$$\begin{aligned} U[X(t+h) \,|\, I_W(t)] &= \sum_{j=1}^{h-1} \psi_{XXj} u_X(t+h-j) + \sum_{j=1}^{h-1} \psi_{XYj} u_Y(t+h-j) \\ &\quad + \sum_{j=1}^{h-1} \psi_{XZj} u_Z(t+h-j) + u_X(t+h), \end{aligned}$$

which is associated with the unconstrained variance-covariance matrix

$$\Sigma[X(t+h) \,|\, I_W(t)] = \sum_{i=0}^{h-1} J_1 \Psi_i \Sigma_u \Psi_i' J_1'$$

where $J_1 = \begin{bmatrix} I_{m_1} & 0 & 0 \end{bmatrix}$. Similarly, the forecast error of $W_0(t+h)$ is given by

$$U_0[W_0(t+h) \,|\, I_{W_0}(t)] = \sum_{i=0}^{h-1} \bar{\Psi}_i \varepsilon(t+h-i)$$

associated with the variance-covariance matrix

$$\Sigma[W_0(t+h) \,|\, I_{W_0}(t)] = \sum_{i=0}^{h-1} \bar{\Psi}_i \, \Sigma_\varepsilon \bar{\Psi}_i' \, .$$

Consequently, the constrained forecast error of $X(t+h)$ is given by

$$U_0\left[X(t+h) \,|\, I_{W_0}(t)\right] = \sum_{j=1}^{h-1} \bar{\psi}_{XXj} \, \varepsilon_X(t+h-j) + \sum_{j=1}^{h-1} \bar{\psi}_{XZj} \, \varepsilon_Z(t+h-j) + \varepsilon_X(t+h)$$

associated with the constrained variance-covariance matrix

$$\Sigma[X(t+h) \,|\, I_{W_0}(t)] = \sum_{i=0}^{h-1} J_0 \bar{\Psi}_i \, \Sigma_\varepsilon \bar{\Psi}_i' \, J_0{}'$$

where $J_0 = \begin{bmatrix} I_{m_1} & 0 \end{bmatrix}$. We can immediately deduce the following result by using the definition of a causality measure from Y to X [see Definition **4.1**].

Theorem 5.4 REPRESENTATION OF CAUSALITY MEASURE IN TERMS OF IMPULSE RESPONSES. *Under assumptions (*5.1*) and (*5.4*),*

$$C(Y \underset{h}{\longrightarrow} X \,|\, Z) = \ln \left[\frac{\det(\sum_{i=0}^{h-1} J_0 \bar{\Psi}_i \, \Sigma_\varepsilon \bar{\Psi}_i' \, J_0')}{\det(\sum_{i=0}^{h-1} J_1 \Psi_i \Sigma_u \Psi_i' J_1')} \right]$$

for $h \geq 1$, *where* $J_1 = \begin{bmatrix} I_{m_1} & 0 & 0 \end{bmatrix}$, *and* $J_0 = \begin{bmatrix} I_{m_1} & 0 \end{bmatrix}$.

We can, of course, repeat the same argument switching the role of the variables X and Y.

Example 5.5 For a bivariate VAR(1) model [see Example **5.3**], we can analytically compute the causality measures at any horizon h using only the unconstrained parameters. For example, the measures of causality from Y to X at horizons 1 and 2 are given by[1]

$$C(Y \underset{1}{\longrightarrow} X) = \ln\left[\frac{(1+\varphi_{YY}^2)\sigma_{u_X}^2 + \varphi_{XY}^2\sigma_{u_Y}^2 + \sqrt{((1+\varphi_{YY}^2)\sigma_{u_X}^2 + \varphi_{XY}^2\sigma_{u_Y}^2)^2 - 4\varphi_{YY}^2\sigma_{u_X}^4}}{2\sigma_{u_X}^2}\right], \tag{5.14}$$

$$C(Y \underset{2}{\longrightarrow} X) = \ln\left[\frac{4\varphi_{YY}^2\sigma_{u_X}^4 + [(1+\varphi_{YY}^2)\sigma_{u_X}^2 + \varphi_{XY}^2\sigma_{u_Y}^2 - \Delta - 2\varphi_{YY}\sigma_{u_X}^2]^2}{2[(1+\varphi_{XX}^2)\sigma_{u_X}^2 + \varphi_{XY}^2\sigma_{u_Y}^2][(1+\varphi_{YY}^2)\sigma_{u_X}^2 + \varphi_{XY}^2\sigma_{u_Y}^2 - \Delta]}\right] \tag{5.15}$$

where $\Delta = \sqrt{((1+\varphi_{YY})^2\sigma_{u_X}^2 + \varphi_{XY}^2\sigma_{u_Y}^2)^2 - 4\varphi_{YY}^2\sigma_{u_X}^4}$.

Now, we will determine the parametric measure of instantaneous causality between X and Y at given horizon h. We know from Section 4 that a measure of instantaneous causality is defined only in terms of the variance-covariance matrices of unconstrained forecast errors [see Definition **4.3**]. The variance-covariance matrix of the unconstrained forecast error of joint process $\big(X(t+h)', Y'(t+h)'\big)'$ is given by

$$\Sigma\big(X(t+h),\, Y(t+h) \mid I_W(t)\big) = \sum_{i=0}^{h-1} G\,\Psi_i \Sigma_u \Psi_i' G'$$

where $G = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & I_{m_2} & 0 \end{bmatrix}$. Consequently,

$$\begin{aligned} \Sigma\big(X(t+h) \mid I_W(t)\big) &= \sum_{i=0}^{h-1} [J_1 \Psi_i \Sigma_u \Psi_i' J_1'] \\ \Sigma\big(Y(t+h) \mid I_W(t)\big) &= \sum_{i=0}^{h-1} [J_2 \Psi_i \Sigma_u \Psi_i' J_2'], \end{aligned}$$

where $J_2 = \begin{bmatrix} 0 & I_{m_2} & 0 \end{bmatrix}$. We can immediately deduce the following result by using the definition of the instantaneous causality measure [see Definition **4.3**].

[1]Equations (5.14)-(5.15) are obtained under assumptions $cov(u_X(t), u_Y(t)) = 0$ and

$$\left[(1+\pi_{YY}^2)\sigma_{u_X}^2 + \pi_{XY}^2\sigma_{u_Y}^2\right]^2 - 4\pi_{YY}^2\sigma_{u_X}^4 \geq 0.$$

Theorem 5.6 REPRESENTATION OF THE INSTANTANEOUS CAUSALITY MEASURE IN TERMS OF IMPULSE RESPONSES. *Under assumptions (*5.1*) and (*5.4*),*

$$C(X\underset{h}{\longleftrightarrow}Y \mid Z) = \ln\left[\frac{\det(\sum_{i=0}^{h-1}[J_1\Psi_i\Sigma_u\Psi_i'J_1'])\ \det(\sum_{i=0}^{h-1}[J_2\Psi_i\Sigma_u\Psi_i'J_2'])}{\det(\sum_{i=0}^{h-1}[G\,\Psi_i\Sigma_u\Psi_i'G'])}\right]$$

for $h \geq 1$, *where* $G = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & I_{m_2} & 0 \end{bmatrix}$, $J_1 = \begin{bmatrix} I_{m_1} & 0 & 0 \end{bmatrix}$, *and* $J_2 = \begin{bmatrix} 0 & I_{m_2} & 0 \end{bmatrix}$.

The parametric measure of dependence between X and Y at horizon h can be deduced from its decomposition given by equation (4.9).

5.2. Characterization of causality measures for VMA(q) processes

Now, assume that the process $\{W(s) = (X(s)', Y(s)', Z(s)')' : s \leq t\}$ follows an invertible VMA(q) model:

$$W(t) = \sum_{j=1}^{q} \Theta_j u(t-j) + u(t) \tag{5.16}$$

where

$$\Theta_j = \begin{bmatrix} \theta_{XXj} & \theta_{XYj} & \theta_{XZj} \\ \theta_{YXj} & \theta_{YYj} & \theta_{YZj} \\ \theta_{ZXj} & \theta_{ZYj} & \theta_{ZZj} \end{bmatrix}$$

or, more compactly,

$$W(t) = \Theta(L)u(t)$$

where

$$\Theta(L) = \begin{bmatrix} \theta_{XX}(L) & \theta_{XY}(L) & \theta_{XZ}(L) \\ \theta_{YX}(L) & \theta_{YY}(L) & \theta_{YZ}(L) \\ \theta_{ZX}(L) & \theta_{ZY}(L) & \theta_{ZZ}(L) \end{bmatrix},$$

$$\theta_{ll}(L) = I_{m_l} + \sum_{j=1}^{q} \theta_{llj}L^j, \ \ \theta_{lk}(L) = \sum_{j=1}^{q} \theta_{lkj}L^j \text{ , for } l \neq k,\ \ l,\ k = X,\ \ Z,\ \ Y.$$

From Lemma **5.1** and letting $F = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix}$, the model of the constrained process $W_0(t) = FW(t)$ is an MA($\bar{q}$) with $\bar{q} \leq q$. We write,

$$W_0(t) = \bar{\theta}(L)\varepsilon(t) = \sum_{j=0}^{\bar{q}} \bar{\theta}_j \varepsilon(t-j) = \sum_{j=0}^{\bar{q}} \begin{bmatrix} \bar{\theta}_{XX,j} & \bar{\theta}_{XZ\,j} \\ \bar{\theta}_{ZX,j} & \bar{\theta}_{ZZ,j} \end{bmatrix} \begin{pmatrix} \varepsilon_X(t-j) \\ \varepsilon_Z(t-j) \end{pmatrix}$$

We have the following result.

Theorem 5.7 CHARACTERIZATION OF CAUSALITY MEASURES FOR $\mathrm{VMA}(q)$. *Let h_1 and h_2 be two different horizons. Under assumption (*5.16*) we have,*

$$C(Y \underset{h_1}{\longrightarrow} X \mid Z) = C(Y \underset{h_2}{\longrightarrow} X \mid Z), \ \forall \, h_2 \geq h_1 \geq q.$$

This result follows immediately from Proposition **4.5**.

6. Estimation

From Section 5, we know that short-run and long-run causality measures depend on the parameters of the model describing the process of interest. Consequently, these measures can be estimated by replacing the unknown parameters by their estimates from a finite sample.

There are at least three different approaches to the estimation of causality measures. The first and simplest approach assumes that the process of interest follows a finite-order VAR(p) model which can be estimated by OLS. The second approach assumes that the process follows a finite-order VARMA model. But standard methods for the estimation of VARMA models, such as maximum likelihood and nonlinear least squares, require nonlinear optimization. This is difficult to implement because the number of parameters can increase quickly. To circumvent this problem, several authors have developed a relatively simple approach based only on linear regression [see Hannan and Rissanen (1982), Hannan and Kavalieris (1984*a*, 1984*b*), Koreisha and Pukkila (1989), Dufour and Pelletier (2005), and Dufour and Jouini (2004)]. This approach enables estimation of VARMA models using a long VAR whose order depends on the sample size. The last approach assumes that the process is autoregressive with potentially infinite order, but can can be approximated by a VAR(k) model, where $k = k(T)$ depends on the sample size. It is the focus of this section.

The precise form of the parametric model appropriate for a process is typically unknown. For this reason, several authors have considered a nonparametric approach to predicting future values using an autoregressive model fitted to a series of T observations; see, for example, Parzen (1974), Bhansali (1978), Lewis and Reinsel (1985). This approach is based on assuming the process considered has an infinite-order autoregressive model, which can be approximated in finite samples by a finite-order autoregressive model. In particular, stationary invertible VARMA processes belongs to this class. We will now describe how this approach can be applied to estimate causality measures at different horizons. We first discuss the estimation of the fitted autoregressive constrained and unconstrained models. Then we construct a consistent estimator of the short-run and long-run causality measures.

Consider a stationary vector process $\{W(s) = (X(s)', Y(s)', Z(s)')' : s \leq t)\}$. By Wold's theorem, this process can be written in the form of a VMA(∞) model:

$$W(t) = u(t) + \sum_{j=1}^{\infty} \Psi_j u(t-j).$$

We assume that $\sum_{j=0}^{\infty} \parallel \Psi_j \parallel < \infty$ and $\det\{\Psi(z)\} \neq 0$ for $z \in \mathbb{C}$ and $\mid z \mid \ \leq 1$, where $\parallel \Psi_j \parallel =$

$tr(\varPsi_j \varPsi_j)$ and $\varPsi(z) = \sum_{j=0}^{\infty} \varPsi_j z^j$, with $\varPsi_0 = I_m$ an $m \times m$ identity matrix. Under the latter assumptions, $W(t)$ is invertible and can be written as an infinite autoregressive process:

$$W(t) = \sum_{j=1}^{\infty} \varPhi_j W(t-j) + u(t) \tag{6.1}$$

where $\sum_{j=1}^{\infty} \parallel \varPhi_j \parallel < \infty$ and $\varPhi(z) = I_m - \sum_{j=1}^{\infty} \varPhi_j z^j = \varPsi(z)^{-1}$ satisfies $\det\{\varPhi(z)\} \neq 0$ for $z \in \mathbb{C}$ and $|\, z \,| \leq 1$.

Given a realization $\{W(1), \ldots, W(T)\}$, we can approximate (6.1) by a finite-order VAR(k) model, where k depends on the sample size T:

$$W(t) = \sum_{j=1}^{k} \varPhi_{jk} W(t-j) + u_k(t).$$

The least squares estimators of the coefficients $\varPhi(k) = [\varPhi_{1k}, \varPhi_{2k}, \ldots, \varPhi_{kk}]$ of the fitted VAR(k) model and variance-covariance matrix $\varSigma_{u|k}$ of the error term are given by

$$\hat{\varPhi}(k) = [\hat{\varPhi}_{1k}, \hat{\varPhi}_{2k}, \ldots, \hat{\varPhi}_{kk}] = \hat{\varGamma}_{k1}' \hat{\varGamma}_k^{-1}, \; \hat{\varSigma}_{u|k} = \sum_{t=k+1}^{T} \hat{u}_k(t) \hat{u}_k(t)' / (T-k)$$

where

$$\hat{\varGamma}_k = (T-k)^{-1} \sum_{t=k+1}^{T} w(t) w(t)', \; \hat{\varGamma}_{k1} = (T-k)^{-1} \sum_{t=k+1}^{T} w(t) W(t+1)',$$

$w(t) = (W(t)', \ldots, W(t-k+1)')'$ and $\hat{u}_k(t) = W(t) - \sum_{j=1}^{k} \hat{\varPhi}_{jk} W(t-j)$.

Suppose now we are interested in measuring causality from Y to X at a given horizon h. For that, we need to define the structure of the marginal process $\{W_0(s) = (X(s)', Z(s)')' : s \leq t\}$. Under general condition [and as there is $W(t)$ follows a VARMA(p, q) model as in Lemma **5.2**], $W_0(t)$ has a VAR(∞) representation:

$$W_0(t) = \sum_{j=1}^{\infty} \bar{\varPhi}_j W_0(t-j) + \varepsilon(t). \tag{6.2}$$

(6.2) can be approximated by VAR(k) model, where k depends on the sample size T :

$$W_0(t) = \sum_{j=1}^{k} \bar{\varPhi}_{jk} W(t-j) + \varepsilon_k(t).$$

It is more convenient to calculate the causality measure by considering the same order k for the constrained and unconstrained models. This is to ensure a relevant comparison of the determinants of the variance-covariance matrices of the constrained and unconstrained forecast errors at horizon

h .

The estimators of the autoregressive coefficients $\bar{\Phi}(k) = [\bar{\Phi}_{1k}, \bar{\Phi}_{2k}, \ldots, \bar{\Phi}_{kk}]$ of the fitted constrained VAR(k) model and variance-covariance matrix $\Sigma_{\varepsilon|k}$ of the error term are given by the following equation:

$$\tilde{\Phi}(k) = [\tilde{\Phi}_{1k}, \tilde{\Phi}_{2k}, \ldots, \tilde{\Phi}_{kk}] = \tilde{\Gamma}'_{k1}\tilde{\Gamma}_k^{-1}, \ \tilde{\Sigma}_{\varepsilon|k} = \sum_{t=k+1}^{T} \tilde{\varepsilon}_k(t)\tilde{\varepsilon}_k(t)'/(T-k)$$

where $\tilde{\Gamma}_{k1}$, $\tilde{\Gamma}_k$, and $\tilde{\varepsilon}_k(t)$ are defined as for unconstrained model.

Now to estimate the degree of causality from Y to X at horizon $h,$ we need to estimate the variance-covariance matrices of the unconstrained and constrained forecast errors. The forecast error of the unconstrained process $W(t+h)$ based on the $VAR(\infty)$ model is given by

$$U(h) = \sum_{j=0}^{h-1} \Psi_j u(t+h-j)$$

with the variance-covariance matrix

$$\Sigma(h) = \sum_{j=0}^{h-1} \Psi_j \Sigma_u \Psi_j'$$

where $\Psi_j = \Phi_1^{(j)}$ and

$$\Phi_1^{(j+1)} = \Phi_2^{(j)} + \Phi_1^{(j)}\,\Phi_1, \ \ \Phi_1^{(1)} = \Phi_1, \ \ \Phi_1^{(0)} = I_m, \ \ \text{for } j \geq 1; \tag{6.3}$$

see Dufour and Renault (1998). An estimator of the variance-covariance matrix of the forecast error of $W(t+h)$ based on the $VAR(k)$ model is given by

$$\hat{\Sigma}_k(h) = \sum_{j=0}^{h-1} \hat{\Psi}_{jk} \hat{\Sigma}_{u|k} \hat{\Psi}'_{jk} \tag{6.4}$$

where $\hat{\Psi}_{jk} = \hat{\Phi}_{1k}^{(j)}$ and $\hat{\Phi}_{1k}^{(j)}$ are calculated using (6.3) (with $\Phi_1^{(j)}$ replaced by $\hat{\Phi}_{1k}^{(j)}$). Similarly, the variance-covariance matrix of the forecast error of $W_0(t+h)$ is given by

$$\Sigma_0(h) = \sum_{j=0}^{h-1} \bar{\Psi}_j \Sigma_\varepsilon \bar{\Psi}_j{}'$$

where $\bar{\Psi}_j = \bar{\Phi}_1^{(j)}$ and $\bar{\Phi}_1^{(j)}$ are defined in similar way as in (6.3). Furthermore, an estimator of the variance-covariance matrix of the forecast error of $W_0(t+h)$ based on the $VAR(k)$ model is given by

$$\tilde{\Sigma}_{0|k}(h) = \sum_{j=0}^{h-1} \tilde{\Psi}_{jk} \tilde{\Sigma}_{\varepsilon|k} \tilde{\Psi}_{jk}{}' \tag{6.5}$$

where $\tilde{\Psi}_{jk}$ is an estimate of the corresponding population parameter $\bar{\Psi}_j$. Consequently, from Theorem **5.4** an estimator of the causality measure from Y to X at horizon h is given by

$$\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z) = \ln \left[\frac{\det \left[J_0 \tilde{\Sigma}_{0|k}(h) J_0' \right]}{\det \left[J_1 \hat{\Sigma}_k(h) J_1' \right]} \right]. \tag{6.6}$$

The most basic property that the above estimator should have is *consistency*. To prove consistency, additional regularity assumptions are needed. We consider here the set of standard assumptions originally considered by Lewis and Reinsel (1985) to derive consistency of parameter estimates for a VAR (∞) model. Of course, alternative – eventually weaker – assumptions could also be studied.

Assumption 6.1 *The following conditions are satisfied:*

(1) $E \mid u_h(t) u_i(t) u_j(t) u_l(t) \mid \ \leq \gamma_4 < \infty,$ *for* $1 \leq h, i, j, l \leq m$; *where* $u_h(t)$, $u_i(t)$, $u_j(t)$, *and* $u_l(t)$ *are elements of the vector of the error term* $u(t)$*;*

(2) k *is chosen as a function of* T *such that* $k^3/T \to 0$ *as* $k, T \to \infty$;

(3) k *is chosen as a function of* T *such that* $k^{1/2} \sum_{j=k+1}^{\infty} \parallel \Phi_j \parallel \to 0$ *as* $k, T \to \infty$;

(4) *the series used to estimate parameters of* $VAR(k)$ *and the series used for prediction are generated from two independent processes having the same stochastic structure.*

Theorem 1 in Lewis and Reinsel (1985) ensures convergence of $\hat{\Phi}(k)$ under conditions 1 and 3 of Assumption **6.1** and by choosing k as a function of T such that $k^2/T \to 0$ as $k, T \to \infty$. The latter is an implication of condition 2 of Assumption **6.1**. Consequently, Assumption **6.1** is sufficient for convergency of $\hat{\Phi}(k)$. Furthermore, their Theorem 4 derives the asymptotic distribution for $\hat{\Phi}(k)$ under Assumption **6.1**1 and by assuming that there exists $\{l(k)\}$ a sequence of $km^2 \times 1$ vectors such that $0 < M_1 \leq \parallel l(k) \parallel^2 = l(k)' l(k) \leq M_2 < \infty$, for $k = 1, 2, \ldots$ Under similar conditions the estimator $\tilde{\Phi}(k)$ converges to $\bar{\Phi}(k)$ and asymptotically follows a normal distribution . Finally, we note that $\hat{\Sigma}_{u|k}$ converges to $\Sigma_{u|k}$, as k and $T \to \infty$ [Lütkepohl (1993*a*, pages 308-309)].

Proposition 6.2 CONSISTENCY OF CAUSALITY MEASURES. *Under Assumption* **6.1**, $\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)$ *is a weakly consistent estimator of* $C(Y \underset{h}{\longrightarrow} X \mid Z)$.

Finally, we note that in practice one must choose the value of k to use for any given series T. Lewis and Reinsel (1985, pages 408-409) suggest to use Akaike's information criterion, which was originally proposed to select the order of a finite autoregressive process by choosing the value of k which minimizes the determinant of the estimated one-step ahead mean square prediction error matrix, to determine a finite-order approximation to a true infinite order autoregressive process [see also Bhansali (1978) and Parzen (1974)].

7. Evaluation by simulation of causality measures

Except in very simple specifications, it is quite difficult analytical expressions for causality measures. To bypass this difficulty, we propose here a simple simulation-based technique to calculate causality measures at any horizon h, for $h \geq 1$. To illustrate the proposed technique we consider the examples of Section 2 and limit ourselves to horizons 1 and 2. Since one source of bias in autoregressive coefficients is sample size, the proposed technique consists of simulating a large sample from the unconstrained model whose parameters are assumed to be either known or estimated from a real data set. Once the large sample, hereafter large simulation, is simulated, we use it to estimate the parameters of the constrained model (imposing noncausality). In what follows, we describe an algorithm to calculate the causality measure at given horizon h using a simulation technique.

1. Given the parameters of the unconstrained model and its initial values, simulate a large sample of T observations under the assumption that the probability distribution of the error term $u(t)$ is completely specified [in our work, we have used values of T as high as 1000000]. Note that the form of the probability distribution of $u(t)$ does not affect the value of causality measures.
2. Estimate the constrained model using a large simulation.
3. Calculate the variance-covariance matrices of the constrained and unconstrained forecast errors at horizon h [see Section 6].
4. Calculate the causality measure at horizon h using (6.6).

To see better how this works, consider again Example **2.1**:

$$\begin{bmatrix} X(t+1) \\ Y(t+1) \end{bmatrix} = \Phi \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} + u(t) \tag{7.1}$$

where

$$\Phi = \begin{bmatrix} 0.5 & 0.7 \\ 0.4 & 0.35 \end{bmatrix}, \ \mathsf{E}[u(t)] = 0, \ \mathsf{E}[u(t)u(s)'] = \begin{cases} I_2 \text{ if } s = t \\ 0 \text{ if } s \neq t\,. \end{cases}$$

Our illustration involves two steps. First, we calculate the theoretical values of the causality measures at horizons 1 and 2. We know from Example **5.5** that for a bivariate VAR(1) model it is relatively easy to compute the causality measure at any horizon h using only the unconstrained parameters. Second, we evaluate the causality measures using a large simulation technique and we compare them with theoretical values from step 1. The latter are recovered as follows.

1. We compute the variances of the forecast errors of X at horizons 1 and 2 using its own past and the past of Y. We have:

$$\varSigma(h) = \sum_{i=0}^{h-1} (\varPhi^{i})(\varPhi^{i})'. \tag{7.2}$$

Table 1. Evaluation by simulation of $C(Y \underset{1}{\longrightarrow} X)$ and $C(Y \underset{2}{\longrightarrow} X)$ for Model 7.1

p	$C(Y \underset{1}{\longrightarrow} X)$	$C(Y \underset{2}{\longrightarrow} X)$
1	0.519	0.567
2	0.430	0.220
3	0.427	0.200
4	0.425	0.199
5	0.426	0.198
10	0.425	0.197
15	0.426	0.199
20	0.425	0.197
25	0.425	0.199
30	0.426	0.198
35	0.425	0.198

From (7.2), we get

$$\mathsf{V}[X(t+1) \,|\, I_X(t),\, I_Y(t)] = 1,\; \mathsf{V}[X(t+2) \,|\, I_X(t),\, I_Y(t)] = 1.74.$$

2. We compute the variances of the forecast errors of X at horizons 1 and 2 using only its own past. To do that we need to determine the structure of the constrained model. This one is given by the following equation [see Example **5.3**]:

$$X(t+1) = 0.85X(t) + 0.105X(t-1) + \varepsilon_X(t+1) + \bar{\theta}\varepsilon_X(t).$$

The parameters $\bar{\theta}$ and $\mathsf{V}(\varepsilon_X(t)) = \sigma^2_{\varepsilon_X}$ are the solutions to the following system:

$$(1+\bar{\theta}^2)\sigma^2_{\varepsilon_X} = 1.6125\,, \quad \bar{\theta}\sigma^2_{\varepsilon_X} = -0.35.$$

The set of possible solutions is $\{(\bar{\theta},\, \sigma^2_{\varepsilon_X}) = (-4.378,\, 0.08),\; (-0.2285,\, 1.53)\}$. To get an invertible solution we must choose the combination which satisfies the condition $|\,\bar{\theta}\,| < 1$, *i.e.* the combination $(-0.2285, 1.53)$. Thus, the variance of the forecast error of X at horizon 1 using only its own past is $\varSigma[X(t+1) \,|\, I_X(t)] = 1.53,$ and the variance of the forecast error of X at horizon 2 is $\varSigma[X(t+2) \,|\, I_X(t)] = 2.12$. Consequently,

$$C(Y \underset{1}{\longrightarrow} X) = 0.425,\;\; C(Y \underset{2}{\longrightarrow} X) = 0.197.$$

In a second step we use the algorithm described at the beginning of this section to evaluate the causality measures using a large simulation technique. Table 1 shows results that we get for different lag orders p in the constrained model (using $T = 600000$). These results confirm the convergence ensured by the law of large numbers and that we have proved is Section 5

Table 2. Evaluation by simulation of $C(Y \underset{1}{\longrightarrow} X \mid Z)$ and $C(Y \underset{2}{\longrightarrow} X \mid Z)$ for Model 7.3

p	$C(Y \underset{1}{\longrightarrow} X \mid Z)$	$C(Y \underset{2}{\longrightarrow} X \mid Z)$
1	0.000	0.121
2	0.000	0.123
3	0.000	0.122
4	0.000	0.123
5	0.000	0.124
10	0.000	0.122
15	0.000	0.122
20	0.000	0.122
25	0.000	0.124
30	0.000	0.122
35	0.000	0.122

Now consider Example **2.2**:

$$\begin{bmatrix} X(t+1) \\ Y(t+1) \\ Z(t+1) \end{bmatrix} = \begin{bmatrix} 0.60 & 0.00 & 0.80 \\ 0.00 & 0.40 & 0.00 \\ 0.00 & 0.60 & 0.10 \end{bmatrix} \begin{bmatrix} X(t) \\ Y(t) \\ Z(t) \end{bmatrix} + \begin{bmatrix} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \\ \varepsilon_Z(t+1) \end{bmatrix}. \tag{7.3}$$

In this example, analytical calculation of the causality measures is not easy. In model (7.3) Y does not cause X at horizon one, but causes it at horizon 2 (indirect causality). Consequently, we expect that causality measure from Y to X will be equal to zero at horizon 1 and different from zero at horizon 2. Using a large simulation technique and by considering different lag orders p in the constrained model, we get the results in Table 2. These results confirm our expectation and show clearly the presence of an indirect causality from Y to X.

8. Confidence intervals

In this section, we assume that X and Y are univariate processes ($m_1 = m_2 = 1$) while Z can be multivariate ($m_3 \geq 0$). This corresponds to the case of most practical interest. Furthermore and for simplicity of exposition, we assume that the process $W \equiv \{W(s) = (X(s)', Y(s)', Z(s)')' : s \leq t)\}$ follows a VAR(p) model:

$$W(t) = \sum_{i=1}^{p} \Phi_i W(t-i) + u(t) \tag{8.1}$$

or equivalently,

$$\big(I_m - \sum_{i=1}^{p} \Phi_i L^i\big) W(t) = u(t)$$

where I_m is an $m \times m$ identity matrix, the polynomial $\Phi(z) = I_m - \sum_{i=1}^{p} \Phi_i z^i$ satisfies $\det[\Phi(z)] \neq 0$, for $z \in \mathbb{C}$ with $|\,z\,| \leq 1$, and $\{u(t)\}_{t=0}^{\infty}$ is a sequence of *i.i.d.* random variables.[2]

For a realization $\{W(1), \ldots , W(T)\}$ of process W, estimates of $\Phi = [\Phi_1, \ldots , \Phi_p]$ and the variance-covariance matrix Σ_u of $u(t)$ are given by the following equations:

$$\hat{\Phi} = \hat{\Gamma}_1' \hat{\Gamma}^{-1}, \;\; \hat{\Sigma}_u = \sum_{t=p+1}^{T} \hat{u}(t)\hat{u}(t)'/(T-p), \tag{8.2}$$

where

$$\hat{\Gamma} = (T-p)^{-1} \sum_{t=p+1}^{T} w(t)w(t)', \;\; \hat{\Gamma}_1 = (T-p)^{-1} \sum_{t=p+1}^{T} w(t)W(t+1)',$$

$w(t) = (W(t)', \ldots , W(t-p+1)')'$, and $\hat{u}(t) = W(t) - \sum_{i=1}^{p} \hat{\Phi}_i W(t-i)$.

Suppose that we are interested in measuring causality from Y to X at given horizon h. To do that we need to know the structure of the marginal process $\{W_0(s) = (X(s), Z(s)')' : s \leq t)\}$. This one has a VARMA($\bar{p}$, $\bar{q}$) representation with $\bar{p} \leq mp$ and $\bar{q} \leq (m-1)p$,

$$W_0(t) = \sum_{i=1}^{\bar{p}} \Phi_i^c W_0(t-i) + \sum_{i=1}^{\bar{q}} \theta_i^c \varepsilon(t-i) + \varepsilon(t) \tag{8.3}$$

where $\{\varepsilon(t)\}_{t=0}^{\infty}$ is a sequence of uncorrelated random variables that satisfies

$$\mathsf{E}\left[\varepsilon(t)\right] = 0, \;\; \mathsf{E}\left[\varepsilon(t)\varepsilon'(s)\right] = \begin{cases} \Sigma_\varepsilon & \text{if } s = t \\ 0 & \text{if } s \neq t \end{cases},$$

and Σ_ε is a positive definite matrix. Equation (8.3) can be rewritten in the following reduced form:

$$\Phi^c(L) W_0(t) = \theta^c(L) \varepsilon(t)$$

where $\Phi^c(L) = I_{\bar{m}} - \Phi_1^c L - \cdots - \Phi_{\bar{p}}^c L^{\bar{p}}$ and $\theta^c(L) = I_{\bar{m}} + \theta_1^c L + \cdots + \theta_{\bar{q}}^c L^{\bar{q}}$, for $\bar{m} = m_3 + 1$ and $I_{\bar{m}}$ an $\bar{m} \times \bar{m}$ identity matrix. We assume that $\theta^c(z) = I_{\bar{m}} + \sum_{j=1}^{\bar{q}} \theta_j^c z^j$ satisfies $\det[\theta^c(z)] \neq 0$ for $z \in \mathbb{C}$ and $|\,z\,| \leq 1$. Under the latter assumption, the VARMA($\bar{p}$, $\bar{q}$) process is invertible and has a VAR(∞) representation:

$$W_0(t) - \sum_{j=1}^{\infty} \bar{\Phi}_j W_0(t-j) = \theta^c(L)^{-1} \Phi^c(L) W_0(t) = \varepsilon(t). \tag{8.4}$$

[2] If W follows a VAR(∞) model, then one can use Inoue and Kilian's (2002) approach to get results that are similar to those developed in this section.

We approximate (8.4) by a finite-order VAR(k) model, where k depends on sample size T :

$$W_0(t) = \sum_{j=1}^{k} \bar{\Phi}_{jk} W_0(t-j) + \varepsilon_k(t). \tag{8.5}$$

The estimators of the coefficients $\bar{\Phi}(k) = [\bar{\Phi}_{1k}, \bar{\Phi}_{2k}, \ldots, \bar{\Phi}_{kk}]$ of the fitted constrained VAR(k) model and variance-covariance matrix $\Sigma_{\varepsilon|k}$ of the error term are given by the following equation:

$$\tilde{\Phi}(k) = [\tilde{\Phi}_{1k}, \tilde{\Phi}_{2k}, \ldots, \tilde{\Phi}_{kk}] = \tilde{\Gamma}_{k1}' \tilde{\Gamma}_k^{-1}, \;\; \tilde{\Sigma}_{\varepsilon|k} = \sum_{t=k+1}^{T} \tilde{\varepsilon}_k(t) \tilde{\varepsilon}_k(t)' / (T-k),$$

where $\tilde{\Gamma}_k$, $\tilde{\Gamma}_{k1}$, and $\tilde{\varepsilon}_k(t)$ are calculated as for the unconstrained model.

The theoretical value of the causality measure from Y to X at horizon h is given by

$$C(Y \underset{h}{\longrightarrow} X \mid Z) = \ln \left(\frac{\det\left(J_0 \Sigma_0(h) J_0'\right)}{\det\left(J_1 \Sigma(h) J_1'\right)} \right)$$

where

$$\Sigma_0(h) = \sum_{j=0}^{h-1} \bar{\Psi}_j \Sigma_\varepsilon \bar{\Psi}_j', \quad \Sigma(h) = \sum_{j=0}^{h-1} \Psi_j \Sigma_u \Psi_j{}',$$

$\bar{\Psi}_j = \bar{\Phi}_1^{(j)}$, $\Psi_j = \Phi_1^{(j)}$ and $\Phi_1^{(j)}$ is defined in similar way as in (6.3). Using Lemma **5.2** $C(Y \underset{h}{\longrightarrow} X \mid Z)$ may be written as follows:

$$C(Y \underset{h}{\longrightarrow} X \mid Z) = \ln \left(\frac{\det\left(G(\Phi, \Sigma_u)\right)}{\det\left(H(\Phi, \Sigma_u)\right)} \right),$$

$$G(\Phi, \Sigma_u) = \sum_{j=0}^{h-1} J_0 \bar{\Psi}_j \Sigma_\varepsilon \bar{\Psi}_j' J_0', \;\; H(\Phi, \Sigma_u) = \sum_{j=0}^{h-1} J_1 \Psi_j \Sigma_u \Psi_j{}' J_1',$$

$G(\cdot)$ and $H(\cdot)$ are continuous and differentiable *functions of* (Φ, Σ_u). A consistent estimator of $C(Y \underset{h}{\longrightarrow} X \mid Z)$ is given by

$$\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z) = \ln \left(\frac{\det\left(J_0 \tilde{\Sigma}_{0|k}(h) J_0'\right)}{\det\left(J_1 \hat{\Sigma}(h) J_1'\right)} \right) \tag{8.6}$$

where

$$\tilde{\Sigma}_{0|k}(h) = \sum_{j=0}^{h-1} \tilde{\Psi}_{jk} \, \tilde{\Sigma}_{\varepsilon|k} \tilde{\Psi}_{jk}{}', \;\; \hat{\Sigma}(h) = \sum_{j=0}^{h-1} \hat{\Psi}_j \hat{\Sigma}_u \hat{\Psi}_j{}',$$

$\hat{\Psi}_j$, $\hat{\Sigma}_u, \tilde{\Psi}_{jk}$, and $\hat{\Sigma}_{\varepsilon|k}$ are estimates of the corresponding population quantities Ψ_j, $\Sigma_u, \bar{\Psi}_{jk}$, and $\Sigma_{\varepsilon|k}$.

To establish the asymptotic distribution of $\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)$, we recall the following result [see Lütkepohl (1993*a*, Chapter 3) and Kilian (1998, page 221)]:

$$T^{1/2} \begin{pmatrix} \text{vec}(\hat{\Phi}) - \text{vec}(\Phi) \\ \text{vech}(\hat{\Sigma}_u) - \text{vech}(\Sigma_u) \end{pmatrix} \xrightarrow{d} N[0,\ \Omega] \tag{8.7}$$

where vec denotes the column stacking operator, vech is the column stacking operator that stacks the elements on and below the diagonal only,

$$\Omega = \begin{bmatrix} \Gamma^{-1} \otimes \Sigma_u & 0 \\ 0 & 2(D_m^{'} D_m)^{-1} D_m^{'} (\Sigma_u \otimes \Sigma_u) D_m (D_m^{'} D_m)^{-1} \end{bmatrix}, \tag{8.8}$$

and D_m is the duplication matrix, defined such that $\text{vech}(F) = D_m \text{vech}(F)$ for any symmetric $m \times m$ matrix F. Thereafter, we will consider the following assumptions.

Assumption 8.1 *The following conditions are satisfied:*

(1) $E \mid \varepsilon_h(t)\varepsilon_i(t)\varepsilon_j(t)\varepsilon_l(t) \mid \ \leq \gamma_4 < \infty$, *for* $1 \leq h, i, j, l \leq \bar{m}$; *where* $\varepsilon_h(t)$, $\varepsilon_i(t)$, $\varepsilon_j(t)$, *and* $\varepsilon_l(t)$ *are elements of the vector of the error term* $\varepsilon(t)$;

(2) k *is chosen as a function of* T *such that* $k^3/T \to 0$ *as* $k, T \to \infty$;

(3) k *is chosen as a function of* T *such that* $k^{1/2} \sum_{j=k+1}^{\infty} \parallel \bar{\Phi}_j \parallel \to 0$ *as* $k, T \to \infty$;

(4) *the series used to estimate parameters of* $VAR(k)$ *and the series used for prediction are generated from two independent processes having the same stochastic structure.*

Proposition 8.2 ASYMPTOTIC DISTRIBUTION OF CAUSALITY MEASURES. *Under Assumptions* **6.1**, *we have:*

$$T^{1/2}[\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z) - C(Y \underset{h}{\longrightarrow} X \mid Z)] \xrightarrow{d} N\big[0,\ \sigma_c(h)^2\big]$$

where $\sigma_c(h)^2 = D_C \Omega D_C^{'}$, $D_C = \partial C(Y \underset{h}{\longrightarrow} X \mid Z) \,/\, \partial\theta'$, $\theta = \big(\text{vec}(\Phi)', \text{vech}(\Sigma_u)'\big)'$ *and* Ω *is given by* (8.8).

Differentiating analytically the causality measures with respect to θ is typically difficult. One way to build confidence intervals for causality measures is to use a large simulation technique [see Section 7] to calculate the derivative numerically. Another way consists in building bootstrap confidence intervals. As mentioned by Inoue and Kilian (2002), for bounded measures, as in our case, the bootstrap approach is more reliable than the delta-method. One reason is because the delta-method interval is not range respecting and may produce confidence intervals that are logically invalid. In contrast, the bootstrap percentile interval preserves by construction these constraints [see Inoue and Kilian (2002, pages 315-318) and Efron and Tibshirani (1993)].

Let us consider the following bootstrap approximation to the distribution of the causality measure at given horizon h.

1. Estimate a VAR(p) process and save the residuals

$$\tilde{u}(t) = W(t) - \sum_{i=1}^{p} \hat{\Phi}_i W(t-i), \text{ for } t = p+1, \ldots, T,$$

$\hat{\Phi}_i$, for $i = 1, \ldots, p$, are given by (8.2) and the OLS estimate of Σ_u is given by $\hat{\Sigma}_u = \sum_{t=p+1}^{T} \hat{u}(t)\hat{u}(t)^{'}/(T-p)$, where $\hat{u}(t) = \tilde{u}(t) - \sum_{t=p+1}^{T} \tilde{u}(t)/(T-p)$ and $\tilde{u}(t) = W(t) - \sum_{i=1}^{p} \hat{\Phi}_i W(t-i)$.

2. Generate $(T-p)$ bootstrap residuals $u^*(t)$ by random sampling with replacement from the residuals $\hat{u}(t),\ t = p+1, \ldots, T$.

3. Choose the vector of p initial observations $w(0) = (W(1)^{'}, \ldots, W(p)^{'})^{'}$. [3]

4. Given $\hat{\Phi} = [\hat{\Phi}_1, \ldots, \hat{\Phi}_p]$, $\{u^*(t)\}_{t=p+1}^{T}$, and $w(0)$, generate bootstrap data for the dependent variable $W^*(t)$ from equation:

$$W^*(t) = \sum_{i=1}^{p} \hat{\Phi}_i W^*(t-i) + u^*(t), \text{ for } t = p+1, \ldots, T\,. \tag{8.9}$$

5. Calculate the bootstrap OLS regression estimates

$$\hat{\Phi}^* = [\hat{\Phi}_1^*, \hat{\Phi}_2^*, \ldots, \hat{\Phi}_p^*] = \hat{\Gamma}_1^{*'}\hat{\Gamma}^{*-1}, \ \hat{\Sigma}_u^* = \sum_{t=p+1}^{T} \hat{u}^*(t)\hat{u}^*(t)^{'}/(T-p),$$

where

$$\hat{\Gamma}^* = (T-p)^{-1} \sum_{t=p+1}^{T} w^*(t)w^*(t)^{'}, \ \hat{\Gamma}_1^* = (T-p)^{-1} \sum_{t=p+1}^{T} w^*(t)W^*(t+1)^{'},$$

$w^*(t) = (W^*(t)^{'}, \ldots, W^*(t-p+1)^{'})^{'}$, , $\hat{u}^*(t) = \tilde{u}^*(t) - \sum_{t=p+1}^{T} \tilde{u}^*(t)/(T-p)$, and $\tilde{u}^*(t) = W^*(t) - \sum_{i=1}^{p} \hat{\Phi}_i W^*(t-i)$.

6. Estimate the constrained model of the marginal process (X, Z) using the bootstrap sample $\{W^*(t)\}_{t=1}^{T}$.

[3]The choice of using the initial vectors $(W(1)^{'}, \ldots, W(p)^{'})^{'}$ seems natural, but any block of p vectors from $W \equiv \{W(1), \ldots, W(T)\}$ would be appropriate. Berkowitz and Kilian (2000) note that conditioning each bootstrap replicate on the same initial value will understate the uncertainty associated with the bootstrap estimates, and this choice is randomised in the simulations by choosing the starting value from $W \equiv \{W(1), \ldots, W(T)\}$ [see Patterson (2007)].

7. Calculate the causality measure at horizon $h,$ denoted $\hat{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z)$, using equation (6.6) and the bootstrap sample.

8. Choose B such $\frac{1}{2}\alpha(B+1)$ is an integer and repeat steps (2)-(7) B times.

We have the following result which establish the validity of the percentile bootstrap technique.

Proposition 8.3 ASYMPTOTIC VALIDITY OF THE RESIDUAL-BASED BOOTSTRAP. *Under assumptions* **6.1***, we have*

$$T^{1/2}[\hat{C}^{*}(Y \underset{h}{\longrightarrow} X \mid Z) - \hat{C}(\, Y \underset{h}{\longrightarrow} X \mid Z)] \xrightarrow{d} N\big[0,\, \sigma_c(h)^2\big]$$

where $\sigma_c(h)^2$ and Ω are defined in Proposition **8.2***.*

Kilian (1998) proposes an algorithm to remove the bias in impulse response functions prior to bootstrapping the estimate. As he mentioned, the small sample bias in an impulse response function may arise from bias in slope coefficient estimates or from the nonlinearity of this function, and this can translate into changes in interval width and location. If the ordinary least-squares small-sample bias can be responsible for bias in the estimated impulse response function, then replacing the biased slope coefficient estimates by bias-corrected slope coefficient estimates may help to reduce the bias in the impulse response function. Kilian (1998) shows that the additional modifications proposed in the bias-corrected bootstrap confidence intervals method do not alter its asymptotic validity. The reason is that the effect of bias corrections is negligible asymptotically.

To improve the performance of the percentile bootstrap intervals described above, we almost consider the same algorithm as in Kilian (1998). Before bootstrapping the causality measures, we correct the bias in the VAR coefficients. We approximate the bias term $Bias = E[\hat{\Phi} - \Phi]$ of the VAR coefficients by the corresponding bootstrap bias $Bias^{*} = E^{*}[\hat{\Phi}^{*} - \hat{\Phi}]$, where E^{*} is the expectation based on the bootstrap distribution of $\hat{\Phi}^{*}$. This suggests the bias estimate

$$\widehat{Bias}^{*} = \frac{1}{B}\sum_{j=1}^{B} \hat{\Phi}^{*(j)} - \hat{\Phi}.$$

We substitute $\hat{\Phi} - \widehat{Bias}^{*}$ in equation (8.9) and generate B new bootstrap replications $\hat{\Phi}^{*}$. We use the same bias estimate, $\widehat{Bias}^{*}$, to estimate the mean bias of new $\hat{\Phi}^{*}$[see Kilian (1998)]. Then we calculate the bias-corrected bootstrap estimator $\tilde{\Phi}^{*} = \hat{\Phi}^{*} - \widehat{Bias}^{*}$ that we use to estimate the bias-corrected bootstrap causality measure estimate. Based on the discussion by Kilian (1998, page 219), given the nonlinearity of the causality measure, this procedure will not in general produce unbiased estimates, but as long as the resulting bootstrap estimator is approximately unbiased, the implied percentile intervals are likely to be good approximations. Further, to reduce the bias in the causality measure estimate, in the empirical application we consider another bias correction applied directly on the measure itself, this one is given by

$$\tilde{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z) = \hat{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z) - [\overline{C}^{*}(Y \underset{h}{\longrightarrow} X \mid Z) - \hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)]$$

Table 3. Augmented Dickey-Fuller tests for the variables in level

	With Intercept		With Intercept and Trend	
	ADF test statistic	5% Critical Value	ADF test statistic	5% Critical Value
NBR	-0.510587	-2.8694	-1.916428	-3.4234
R	-2.386082	-2.8694	-2.393276	-3.4234
P	-1.829982	-2.8694	-0.071649	-3.4234
GDP	-1.142940	-2.8694	-3.409215	-3.4234

where

$$\overline{C}^{*}(Y \underset{h}{\longrightarrow} X \mid Z) = \frac{1}{B}\sum_{j=1}^{B}\tilde{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z).$$

In practice, specially when the true value of causality measure is close to zero, it is possible that for some bootstrap samples

$$\hat{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z) \leq [\overline{C}^{*}(Y \underset{h}{\longrightarrow} X \mid Z) - \hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)].$$

In this case we impose the following non-negativity truncation:

$$\tilde{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z) = \max\left\{\tilde{C}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid Z), 0\right\}.$$

9. Empirical illustration

We apply our causality measures to measure the strength of relationships between macroeconomic and financial variables. The data set considered is the one used by Bernanke and Mihov (1998) and Dufour et al. (2006). This data set consists of monthly observations on nonborrowed reserves (NBR), the federal funds rate (r), the gross domestic product deflator (P), and real gross domestic product (GDP). The monthly data on GDP and the GDP deflator were constructed using state space methods from quarterly observations [for more details, see Bernanke and Mihov (1998)]. The sample runs from January 1965 to December 1996 for a total of 384 observations.

All variables are in logarithmic form [see Figures 1-4]. These variables were also transformed by taking first differences [see Figures 5-8], consequently the causality relations have to be interpreted in terms of the growth rate of variables.

We performed an Augmented Dickey-Fuller test (hereafter ADF-test) for nonstationarity of the four variables of interest and their first differences. The values of the test statistics, as well as the critical values corresponding to a 5% significance level, are given in tables 3 and 4. Table 5, below, summarizes the results of the stationarity tests for all variables.

As we can read from Table 5, all variables in logarithmic form are nonstationary. However, their first differences are stationary except for the GDP deflator, P. We performed a nonstationarity test for the second difference of variable P. The test statistic values are equal to -11.04826 and -11.07160 for the ADF-test with only an intercept and with both intercept and trend, respectively.

Figure 1: NBR in logarithmic form

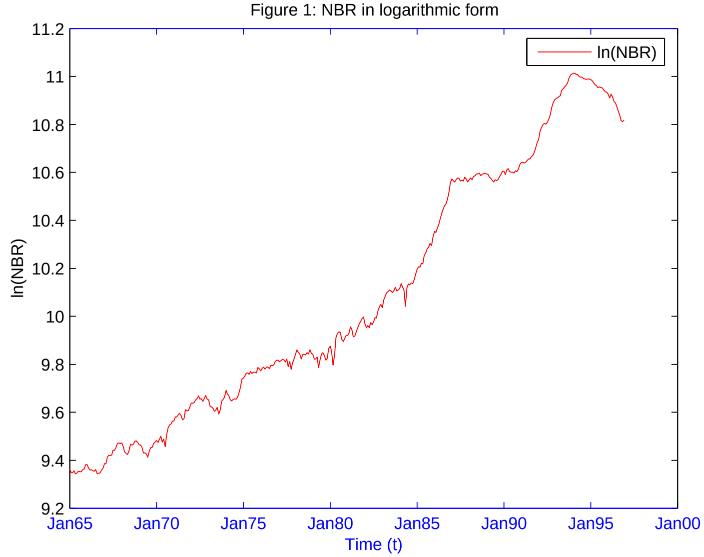


Figure 2: r in logarithmic form

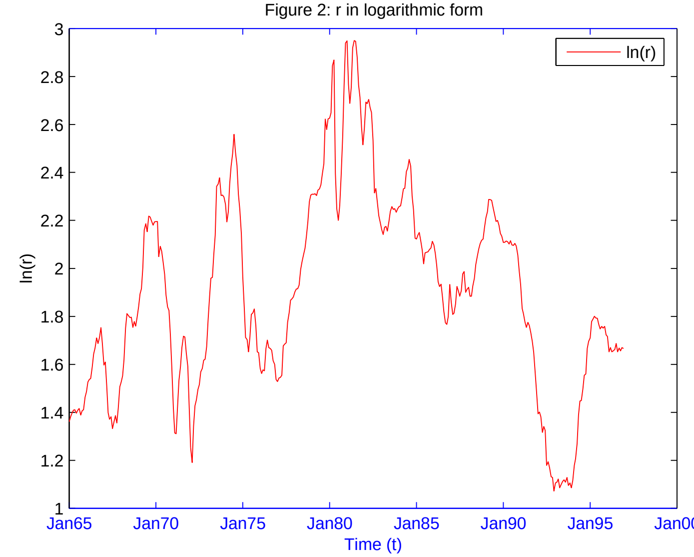


Figure 3: P in logarithmic form

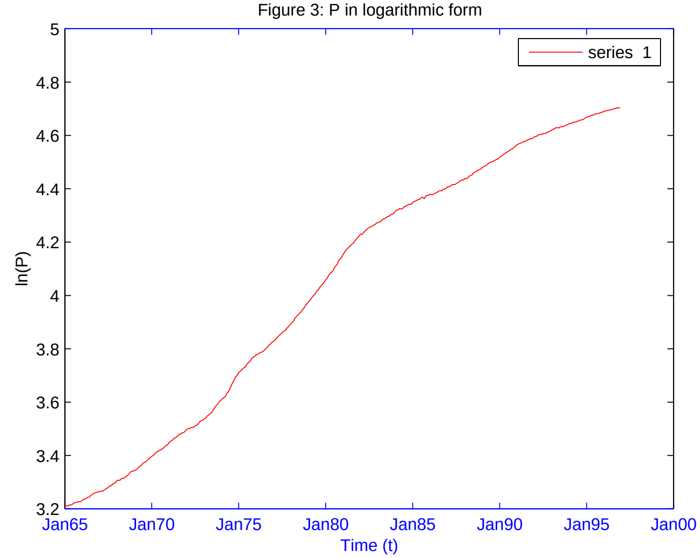


Figure 4: GDP in logarithmic form

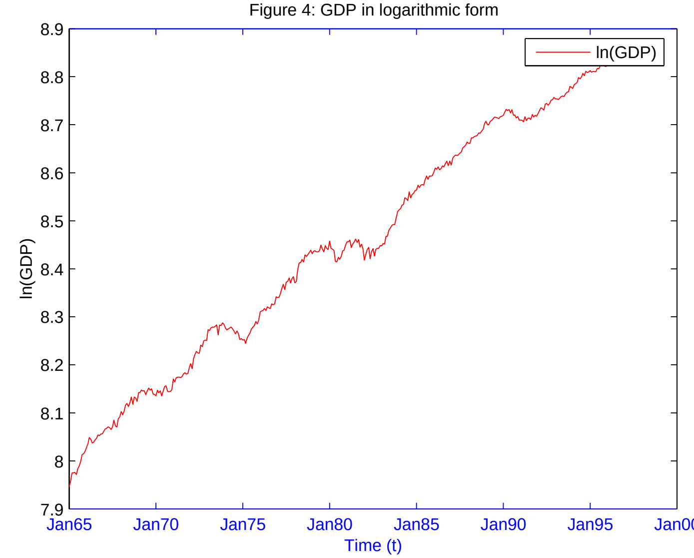


Figure 5: The first dfferentiation of ln(NBR)

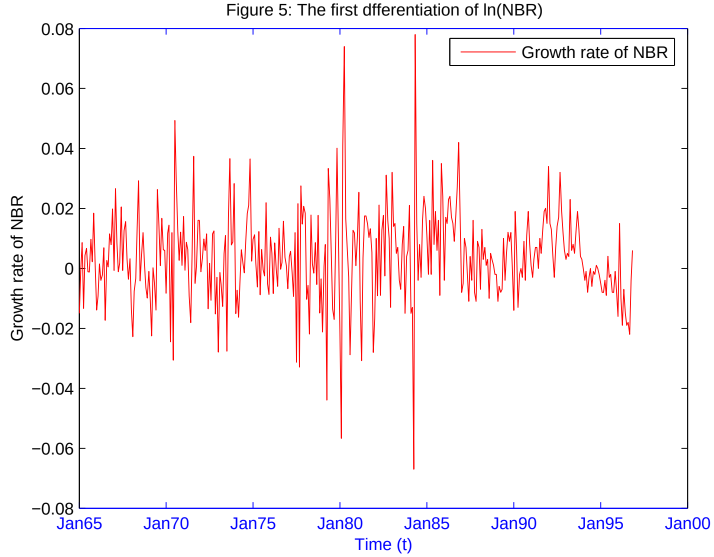


Figure 6: The first dfferentiation of ln(r)

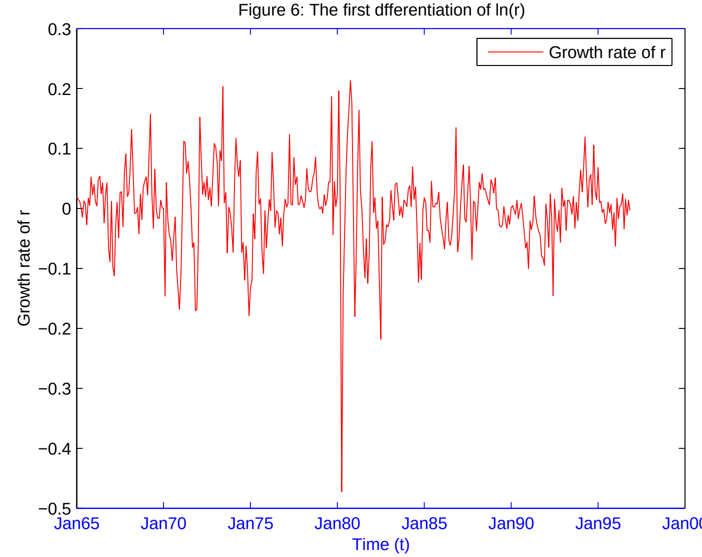


Figure 7: The first dfferentiation of ln(P)

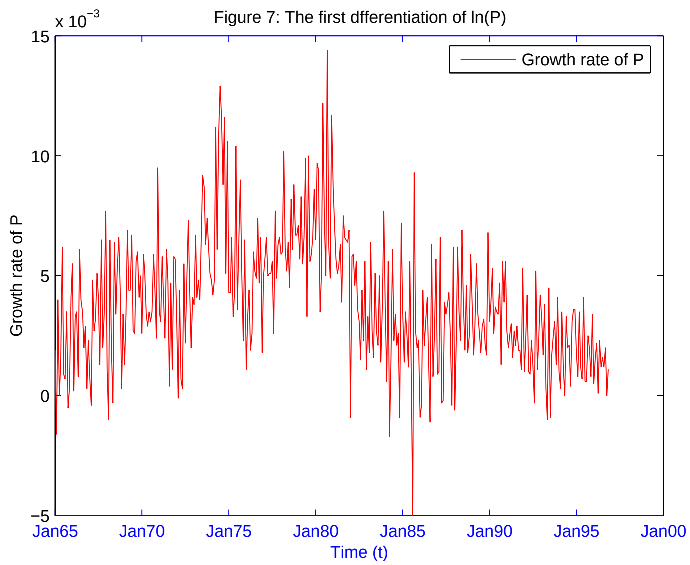


Figure 8: The first dfferentiation of ln(GDP)

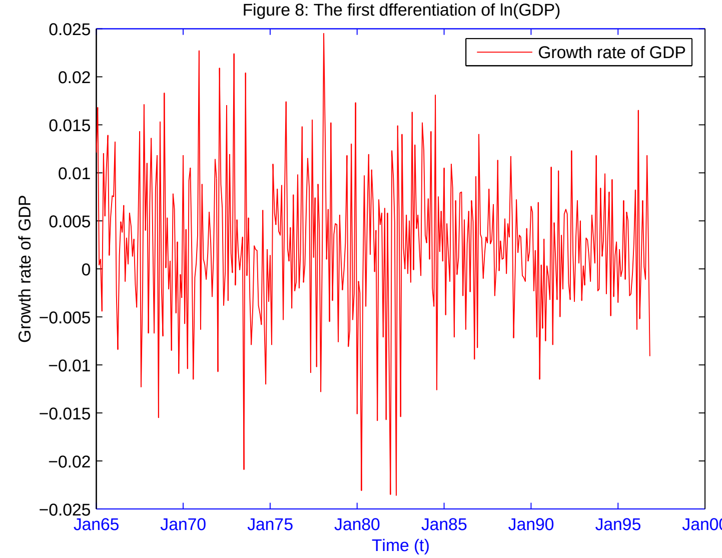


Table 4. Augmented Dickey-Fuller tests for the variables in first difference

	With Intercept		With Intercept and Trend	
	ADF test statistic	5% Critical Value	ADF test statistic	5% Critical Value
NBR	-5.956394	-2.8694	-5.937564	-3.9864
r	-7.782581	-2.8694	-7.817214	-3.9864
P	-2.690660	-2.8694	-3.217793	-3.9864
GDP	-5.922453	-2.8694	-5.966043	-3.9864

Table 5. Unit root test results

	Variables in logarithmic form	First difference
NBR	No	Yes
r	No	Yes
P	No	No
GDP	No	Yes

The critical values in both cases are equal to -2.8695 and -3.4235. Thus, the second difference of variable P is stationary. Once the data is made stationary, we use a nonparametric approach for the estimation and Akaike's information criterion to specify the orders of the long VAR(k) models. Using Akaike's criterion for the unconstrained VAR model, which corresponds to four variables, we observe that it is minimized at $k = 16$. We use same criterion to specify the orders of the constrained VAR models, which correspond to different combinations of three variables, and we find that the orders are all less than or equal to 16. To compare the determinants of the variance-covariance matrices of the constrained and unconstrained forecast errors at horizon h, we take the same order $k = 16$ for the constrained and unconstrained models. We compute different causality measures for horizons $h = 1, \ldots, 40$ [see Figures 9-14]. Higher values of measures indicate greater causality. We also calculate the corresponding nominal 95% bootstrap confidence intervals as described in the previous section.

From Figure 9 we see that nonborrowed reserves have a strong effect on the federal funds rate one month ahead comparatively with other variables [see Figures 10 and 11]. This effect is well known in the literature and can be explained by the theory of supply and demand for money. We also note that nonborrowed reserves have a short-term effect on GDP and can cause the GDP deflator until 5 months. Figure 14 shows the effect of GDP on the federal funds rate is economically important and statistically significant for the first three months. The effect of the federal funds rate on the GDP deflator is economically weak one month ahead [see Figure 12]. Other significant results concern the causality from r to GDP. Figure 13 shows that federal funds rate causes the GDP until 16 months. These results are consistent with conclusions obtained by Dufour et al. (2006).

Table 6 represents results of other causality directions until 20 months. As we can read from

Figure 9: Causality measures from Nonborrowed reserves to Federal funds rate

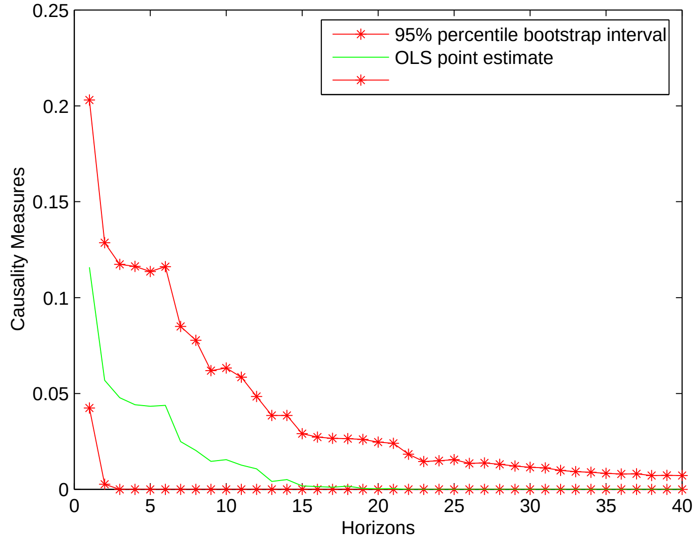


Figure10: Causality measures from Nonborrowed reserves to GDP Deflator

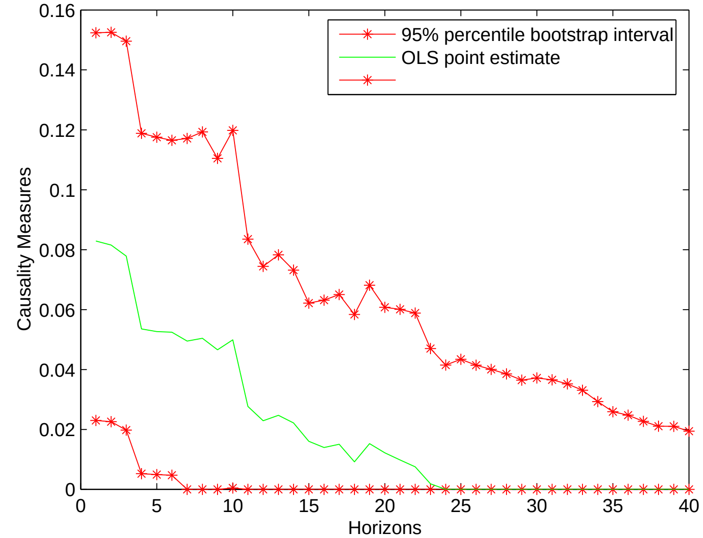


Figure 11: Causality measures from Nonborrowed reserves to Real GDP

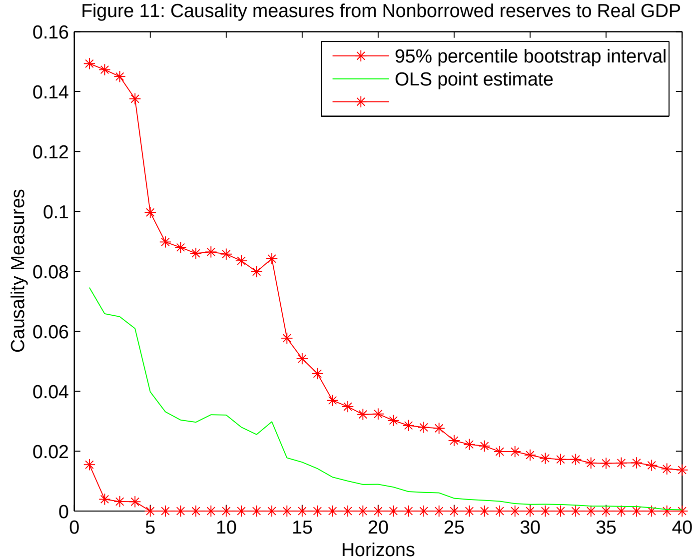


Figure 12: Causality measures from Federal funds rate to GDP Deflator

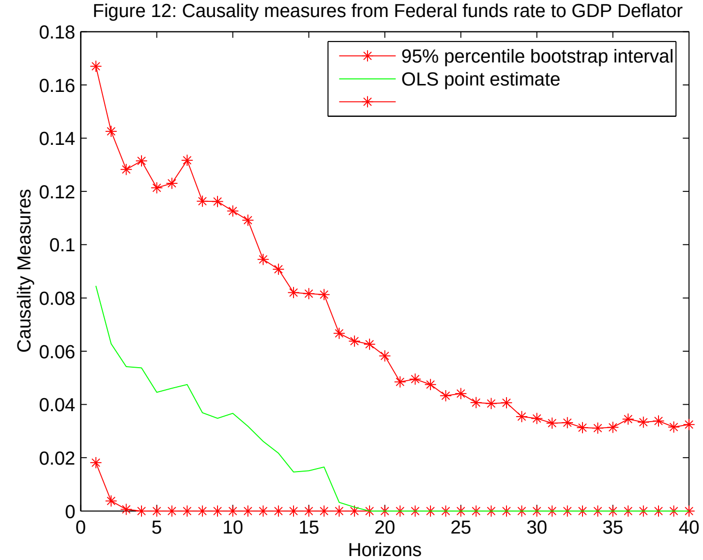


Figure 13: Causality measures from Federal funds rate to Real GDP

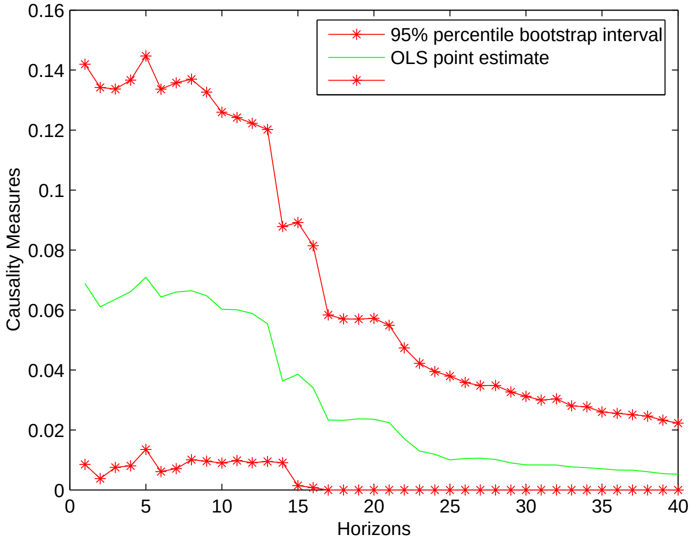
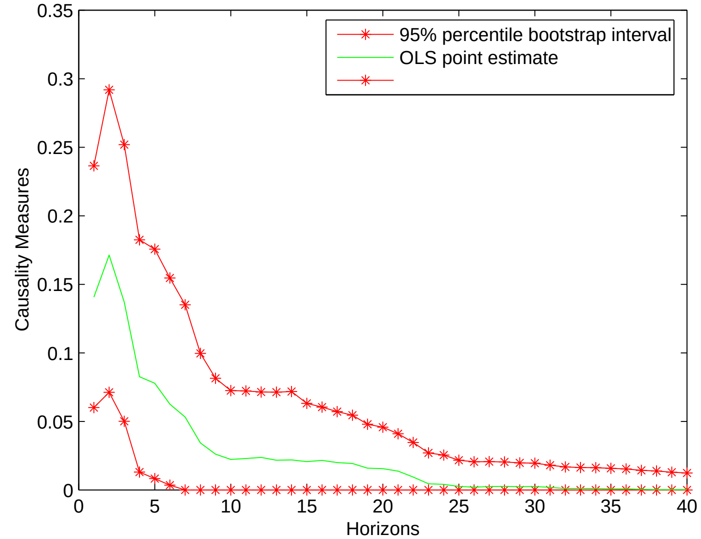


Figure 14: Causality measures from Real GDP to Federal funds rate



this table, there is no causality in these other directions. Finally, note that the above results do not change when we consider the second, rather than first, difference of variable P.

10. Conclusion

New concepts of causality were introduced in Dufour and Renault (1998): causality at a given (arbitrary) horizon h, and causality up to any given horizon h, where h is a positive integer and can be infinite $(1 \leq h \leq \infty)$. These concepts are motivated by the fact that, in the presence of an auxiliary variable Z, it is possible to have a situation in which the variable Y does not cause variable X at horizon 1, but causes it at a longer horizon $h > 1$. In this case, this is an indirect causality transmitted by the auxiliary variable Z.

Another related problem arises when measuring the importance of the causality between two variables. Existing causality measures have been established only for horizon 1 and fail to capture indirect causal effects. This paper proposes a generalization of such measures for any horizon h. We propose parametric and nonparametric measures of causality at any horizon h. Parametric measures are defined in terms of impulse response coefficients in the VMA representation. By analogy with Geweke (1982), we show that it is possible to define a measure of dependence at horizon h which can be decomposed into a sum of causality measures from X to $Y,$ from Y to $X,$ and an instantaneous effect at horizon h. We also show how these causality measures can be related to the predictability measures developed in Diebold and Kilian (2001).

We propose a new approach to estimating these measures based on simulating a large sample from the process of interest. We also propose a valid nonparametric confidence interval, using the bootstrap technique.

From an empirical application we found that there is a strong effect of nonborrowed reserves on federal funds rate one month ahead, the effect of real gross domestic product on federal funds rate is economically important for the first three months, the effect of federal funds rate on gross domestic product deflator is economically weak one month ahead, and finally federal funds rate causes the real gross domestic product until 16 months

Table 6. Summary of causality relations at various horizons for series in first difference

h	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$NBR \to R$	yes																			
$NBR \to P$	yes	yes	yes	yes																
$NBR \to GDP$	yes																			
$R \to NBR$																				
$R \to P$	yes																			
$R \to GDP$	yes		yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes				
$P \to NBR$																				
$P \to R$																				
$P \to GDP$																				
$GDP \to NBR$																				
$GDP \to R$	yes	yes	yes	yes	yes															
$GDP \to P$																				

A. Appendix: Proofs

PROOF OF PROPOSITION **4.5**

$$C(Y \underset{h_2}{\longrightarrow} X \mid Z) = C(Y \underset{h_1}{\longrightarrow} X \mid Z) + \ln\left[\frac{\sigma^2(X(t+h_1) \mid I(t))}{\sigma^2(X(t+h_2) \mid I(t))}\right] - \ln\left[\frac{\sigma^2\big(X(t+h_1) \mid I_X(t)\big)}{\sigma^2\big(X(t+h_2) \mid I_X(t)\big)}\right]$$

According to Diebold and Kilian (2001), the predictability measure of vector X under the information sets $I_X(t)$ and $I_{XY}(t)$ are, respectively, defined as

$$\begin{aligned}\bar{P}_X\big(I_X(t),\, h_1,\, h_2\big) &= 1 - \frac{\sigma^2\big(X(t+h_1) \mid I_X(t)\big)}{\sigma^2\big(X(t+h_2) \mid I_X(t)\big)}\,,\\ \bar{P}_X\big(I_{XY}(t),\, h_1,\, h_2\big) &= 1 - \frac{\sigma^2\big(X(t+h_1) \mid I_{XY}(t)\big)}{\sigma^2\big(X(t+h_2) \mid I_{XY}(t)\big)}\,.\end{aligned}$$

By Definition **4.1**, we then see that

$$\begin{aligned}C_L(Y \underset{h_1}{\rightarrow} X \mid I) - C_L(Y \underset{h_2}{\rightarrow} X \mid I) &= \ln\left[\frac{\sigma^2[X(t+h_1)\,|I_X(t)]}{\sigma^2[X(t+h_1) \mid I_{XY}(t)]}\right] - \ln\left[\frac{\sigma^2[X(t+h_2)\,|I_X(t)]}{\sigma^2[X(t+h_2) \mid I_{XY}(t)]}\right]\\ &= \ln\left[\frac{\sigma^2[X(t+h_1)\,|I_X(t)]}{\sigma^2[X(t+h_2)\,|I_X(t)]}\right] - \ln\left[\frac{\sigma^2[X(t+h_1) \mid I_{XY}(t)]}{\sigma^2[X(t+h_2) \mid I_{XY}(t)]}\right]\\ &= \ln\big[1 - \bar{P}_X\big(I_X(t),\, h_1,\, h_2\big)\big] - \ln\big[1 - \bar{P}_X\big(I_{XY}(t),\, h_1,\, h_2\big)\big].\end{aligned}$$

□

PROOF OF PROPOSITION **6.2** Under Assumption **6.1** and using Theorem 1 in Lewis and Reinsel (1985), we have

$$\hat{\Phi}(k) = \Phi(k) + o_p(1).$$

Using (4.1) of Lewis and Reinsel (1985) and Assumption **6.1**, we have:

$$\hat{\Sigma}_k(h) = (1+\frac{mk}{T})\Sigma(h) + o_p(1) = \Sigma(h) + \Sigma(h)o_p(T^{-\delta}) + o_p(1),\ \ for\ \delta \le \frac{2}{3},$$

hence

$$\hat{\Sigma}_k(h) \underset{T\to\infty}{\overset{p}{\to}} \Sigma(h). \tag{A.1}$$

Similarly, we can show that

$$\tilde{\Sigma}_{0|k}(h) \underset{T\to\infty}{\overset{p}{\to}} \Sigma_0(h). \tag{A.2}$$

Consequently,

$$\ln\left[\frac{\det\big[J_0\tilde{\Sigma}_{0|k}(h)J_0'\big]}{\det\big[J_1\hat{\Sigma}_k(h)J_1'\big]}\right] \underset{T\to\infty}{\overset{p}{\to}} \ln\left[\frac{\det\big[J_0\Sigma_0(h)(h)J_0'\big]}{\det\big[J_1\Sigma(h)J_1'\big]}\right],$$

and

$$\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z) \underset{T\to\infty}{\overset{p}{\to}} C(Y \underset{h}{\longrightarrow} X \mid Z).$$

□

PROOF OF PROPOSITION **8.2** We know that, for $\delta \leq \frac{2}{3}$,

$$G\big(\tilde{\Phi}(k), \tilde{\Sigma}_{\varepsilon|k}\big) = (1 + o_p(T^{-\delta}))G\big(\Phi, \Sigma_u\big) + o_p(1), \tag{A.3}$$

or

$$\ln\big(G\big(\tilde{\Phi}(k), \tilde{\Sigma}_{\varepsilon|k}\big)\big) = \ln\big(G\big(\Phi, \Sigma_u\big)\big) + o_p(T^{-\delta}) + o_p(1). \tag{A.4}$$

By the differentiability of $G(\cdot)$,

$$\ln\big(G\big(\hat{\Phi},\, \hat{\Sigma}_u\big)\big) = \ln\big(G\big(\Phi,\, \Sigma_u\big)\big) + o_p(1). \tag{A.5}$$

From (A.3) and (A.5), we get

$$\ln\big(G\big(\tilde{\Phi}(k), \tilde{\Sigma}_{\varepsilon|k}\big)\big) = \ln\big(G\big(\hat{\Phi}, \hat{\Sigma}_u\big)\big) + o_p(T^{-\delta}) + o_p(1).$$

Consequently,

$$\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z) = \tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) + o_p(T^{-\delta}) + o_p(1)$$

where

$$\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) = \ln\left(\frac{\det\big(G(\hat{\Phi},\, \hat{\Sigma}_u)\big)}{\det\big(H(\hat{\Phi},\, \hat{\Sigma}_u)\big)}\right).$$

Since $\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) = O_p(1)$, the asymptotic distribution of $\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)$ will be the same as that of $\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z)$. Using a first-order Taylor expansion of $\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z)$, we get

$$\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) = C(Y \underset{h}{\longrightarrow} X \mid Z) + D_C \begin{pmatrix} \mathrm{vec}(\hat{\Phi}) - \mathrm{vec}(\Phi) \\ \mathrm{vech}(\hat{\Sigma}_u) - \mathrm{vech}(\Sigma_u) \end{pmatrix} + o_p(T^{-\frac{1}{2}}),$$

where

$$D_C = \frac{\partial C(Y \underset{h}{\longrightarrow} X \mid Z)}{\partial(\mathrm{vec}(\Phi)',\, \mathrm{vech}(\Sigma_u)')} = \frac{\partial C(Y \underset{h}{\longrightarrow} X \mid Z)}{\partial \theta'}$$

hence

$$T^{1/2}[\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) - C(Y \underset{h}{\longrightarrow} X \mid Z)] \simeq D_C \begin{pmatrix} T^{1/2}\, \mathrm{vec}(\hat{\Phi}) - \mathrm{vec}(\Phi) \\ T^{1/2} \mathrm{vech}(\hat{\Sigma}_u) - \mathrm{vech}(\Sigma_u) \end{pmatrix}.$$

Using (8.7),

$$T^{1/2}[\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) - C(Y \underset{h}{\longrightarrow} X \mid Z)] \xrightarrow{d} N(0,\ \sigma_c(h)^2).$$

Consequently,

$$T^{1/2}[\hat{C}\,(Y \ \underset{h}{\longrightarrow}\ X \mid Z) - C(Y \underset{h}{\longrightarrow} X \mid Z)] \xrightarrow{d} N(0,\ \sigma_c(h)^2)$$

where

$$\sigma_c(h)^2 = D_C \varOmega D_C^{'}$$

$$\varOmega = \begin{bmatrix} \varGamma^{-1} \otimes \varSigma_u & 0 \\ 0 & 2(D_m^{'} D_m)^{-1} D_m^{'}(\varSigma_u \otimes \varSigma_u) D_m (D_m^{'} D_m)^{-1} \end{bmatrix}.$$

D_m is the duplication matrix, defined such that $\mathrm{vech}(F) = D_m \mathrm{vech}(F)$ for any symmetric $m \times m$ matrix F. □

PROOF OF PROPOSITION **8.3** We start by showing that conditional on the sample

$$\mathrm{vec}(\hat{\varPhi}^*) \underset{T\to\infty}{\overset{p}{\to}} \mathrm{vec}(\hat{\varPhi})\,, \quad \mathrm{vech}(\hat{\varSigma}_u^*) \underset{T\to\infty}{\overset{p}{\to}} \mathrm{vech}(\hat{\varSigma}_u),$$
$$\mathrm{vec}(\tilde{\Phi}^*(k)) \underset{T\to\infty}{\overset{p}{\to}} \mathrm{vec}(\tilde{\Phi}(k))\,, \quad \mathrm{vech}(\tilde{\varSigma}_{\varepsilon|\,k}^*) \underset{T\to\infty}{\overset{p}{\to}} \mathrm{vech}(\tilde{\varSigma}_{\varepsilon|\,k}).$$

We first note that

$$\begin{aligned} \mathrm{vec}(\hat{\varPhi}^*) &= \mathrm{vec}(\hat{\varGamma}_1^{*'} \hat{\varGamma}^{*-1}) = \mathrm{vec}((T-p)^{-1} \sum_{t=p+1}^{T} W(t{+}1)^* w^*(t)^{'} \hat{\varGamma}^{*-1}) \\ &= \mathrm{vec}((T-p)^{-1} \sum_{t=p+1}^{T} [\hat{\varPhi} w^*(t) + u^*\,(t+1)] w^*(t)^{'} \hat{\varGamma}^{*-1}) \\ &= \mathrm{vec}(\hat{\varPhi}((T-p)^{-1} \sum_{t=p+1}^{T} w^*(t) w^*(t)^{'}) \hat{\varGamma}^{*-1}) \\ &+ \mathrm{vec}((T-p)^{-1} \sum_{t=p+1}^{T} u^*(t+1) w^*(t)^{'} \hat{\varGamma}^{*-1}) \\ &= \mathrm{vec}(\hat{\varPhi}\, \hat{\varGamma}^* \hat{\varGamma}^{*-1}) + \mathrm{vec}((T-p)^{-1} \sum_{t=p+1}^{T} u^*(t+1) w^*(t)^{'} \hat{\varGamma}^{*-1}). \end{aligned}$$

Let $\Im_t^* = \sigma(u^*(1), \ldots, u^*(t))$ denote the σ-algebra generated by $u^*(1), \ldots, u^*(t)$. Then,

$$\begin{aligned} \mathsf{E}^*[u^*(t+1) w^*(t)^{'} \hat{\varGamma}^{*-1}] &= \mathsf{E}^*[\mathsf{E}^*[u^*(t+1) w^*(t)^{'} \hat{\varGamma}^{*-1} \mid \Im_t^*]] \\ &= \mathsf{E}^*[\mathsf{E}^*[u^*(t+1) \mid \Im_t^*] w^*(t)^{'} \hat{\varGamma}^{*-1}] = 0. \end{aligned}$$

By the law of large numbers,

$$(T-p)^{-1}\sum_{t=p+1}^{T} u^*(t+1)w^*(t)^{'}\hat{\Gamma}^{*-1}{=}E^*[u^*(t+1)w^*(t)^{'}\hat{\Gamma}^{*-1}]{+}o_p(1)\,,$$

and

$$\mathrm{vec}(\hat{\Phi}^*)-\mathrm{vec}(\hat{\Phi})\underset{T\to\infty}{\overset{p}{\to}}0\,.$$

Now, to prove that $\mathrm{vech}(\hat{\Sigma}_u^*)\underset{T\to\infty}{\overset{p}{\to}}\mathrm{vech}(\hat{\Sigma}_u),$ we observe that

$$\begin{aligned}\mathrm{vech}(\hat{\Sigma}_u^*{-}\hat{\Sigma}_u) &= (T-p)^{-1}\,\mathrm{vech}\big[\sum_{t=p+1}^{T}u^*(t)\hat{u}^*(t)^{'}{-}\sum_{t=p+1}^{T}\hat{u}(t)\hat{u}(t)^{'}\big]\\ &= (T-p)^{-1}\mathrm{vech}\big[\sum_{t=p+1}^{T}(u^*(t)u^*(t)^{'}{-}\sum_{t=p+1}^{T}\hat{u}(t)\hat{u}(t)^{'})\big]\,.\end{aligned}$$

Conditional on the sample and by the law of iterated expectations, we have:

$$\begin{aligned}\mathsf{E}^*[u^*(t)u^*(t)^{'}{-}(T-p)^{-1}\sum_{t=p+1}^{T}\hat{u}(t)\hat{u}(t)^{'}] &= \mathsf{E}^*[\mathsf{E}^*[u^*(t)u^*(t)^{'}{-}(T-p)^{-1}\sum_{t=p+1}^{T}\hat{u}(t)\hat{u}(t)^{'}\mid\Im_t^*]]\\ &= \mathsf{E}^*[\mathsf{E}^*[u^*(t)u^*(t)^{'}\mid\Im_t^*]-(T-p)^{-1}\sum_{t=p+1}^{T}\hat{u}(t)\hat{u}(t)^{'}].\end{aligned}$$

Because

$$E^*[\mathsf{E}^*[u^*(t)u^*(t)^{'}\mid\Im_{t-1}^*]{=}(T-p)^{-1}\sum_{t=p+1}^{T}\mathsf{E}^*[u^*(t)u^*(t)^{'}],$$

then

$$E^*[u^*(t)u^*(t)^{'}{-}(T-p)^{-1}\sum_{t=p+1}^{T}\hat{u}(t)\hat{u}(t)^{'}]=0.$$

Since

$$\begin{aligned}&(T-p)^{-1}\big[\sum_{t=p+1}^{T}(u^*(t)u^*(t)^{'}{-}(T-p)^{-1}\sum_{t=p+1}^{T}\hat{u}(t)\hat{u}(t)^{'})\big]\\ &=E^*[u^*(t)u^*(t)^{'}{-}(T-p)^{-1}\sum_{t=p+1}^{T}\hat{u}(t)\hat{u}(t)^{'}]+o_p(1),\end{aligned}$$

we get

$$\mathrm{vec}(\hat{\Sigma}_u^*){-}\,\mathrm{vec}(\hat{\Sigma}_u)\underset{T\to\infty}{\overset{p}{\to}}0.$$

Similarly, we can show that

$$\mathrm{vec}(\tilde{\Phi}^*(k)) \underset{T\to\infty}{\overset{p}{\longrightarrow}} \mathrm{vec}(\tilde{\Phi}(k)) \;\text{ and }\; \mathrm{vech}(\tilde{\Sigma}^*_{\varepsilon|\,k}) \underset{T\to\infty}{\overset{p}{\longrightarrow}} \mathrm{vech}(\tilde{\Sigma}_{\varepsilon|\,k}).$$

Since $G(.)$ and $H(.)$ and differentiable functions, we have:

$$\begin{aligned} \ln\big(H\big(\hat{\Phi}^*,\hat{\Sigma}^*_u\big)\big) &= \ln\big(H\big(\hat{\Phi},\hat{\Sigma}_u\big)\big) + o_p(1), \\ \ln\big(G\big(\tilde{\Phi}^*(k),\tilde{\Sigma}^*_{\varepsilon|\,k}\big)\big) &= \ln\big(G\big(\tilde{\Phi}(k),\tilde{\Sigma}_{\varepsilon|\,k}\big)\big) + o_p(1). \end{aligned}$$

By Theorems 2.5–3.4 in Paparoditis (1996) and Theorem 6 in Lewis and Reinsel (1985), we have, for $\delta \le \frac{2}{3}$,

$$\ln\big(G\big(\tilde{\Phi}^*(k), \tilde{\Sigma}^*_{\varepsilon|\,k}\big)\big) = \ln\big(G\big(\bar{\Phi}, \Sigma_\varepsilon\big)\big) + o_p(T^{-\delta}) + o_p(1)$$

Consequently,

$$\hat{C}^*(Y \underset{h}{\longrightarrow} X \mid Z) = \tilde{C}^*(Y \underset{h}{\longrightarrow} X \mid Z) + o_p(T^{-\delta}) + o_p(1)$$

where

$$\tilde{C}^*(Y \underset{h}{\longrightarrow} X \mid Z) = \ln\left(\frac{\det G\big(\hat{\Phi}^*,\, \hat{\Sigma}^*_u\big)}{\det H\big(\hat{\Phi}^*,\, \hat{\Sigma}^*_u\big)}\right).$$

We have shown that for $\delta \le \frac{2}{3}$ [see the proof of Proposition **8.2**],

$$\hat{C}(Y \underset{h}{\longrightarrow} X \mid Z) = \ln\left(\frac{\det\big(G\big(\hat{\Phi},\, \hat{\Sigma}_u\big)\big)}{\det\big(H\big(\hat{\Phi},\, \hat{\Sigma}_u\big)\big)}\right) + o_p(T^{-\delta}) + o_p(1).$$

Consequently

$$\hat{C}^*(Y \to X \mid Z) = \ln\left(\frac{\det\big(G\big(\hat{\Phi},\, \hat{\Sigma}_u\big)\big)}{\det\big(H\big(\hat{\Phi},\, \hat{\Sigma}_u\big)\big)}\right) + o_p(T^{-\delta}) + o_p(1).$$

Conditional on the sample, the first order Taylor expansion of $\tilde{C}^*(Y \to X \mid Z)$ around $\tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z)$ is given by

$$\hat{C}^*(Y \underset{h}{\longrightarrow} X \mid Z) = \tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z) + D_C \begin{pmatrix} \mathrm{vec}(\hat{\Phi}^*) - \mathrm{vec}(\hat{\Phi}) \\ \mathrm{vech}(\hat{\Sigma}^*_u) - \mathrm{vech}(\hat{\Sigma}_u) \end{pmatrix} + o_p(T^{\frac{1}{2}}),$$

hence

$$T^{1/2}[\hat{C}^*(Y \underset{h}{\longrightarrow} X \mid Z) - \tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z)] \simeq D_C \begin{pmatrix} T^{1/2}(\mathrm{vec}(\hat{\Phi}^*) - \mathrm{vec}(\hat{\Phi})) \\ T^{1/2}(\mathrm{vech}(\hat{\Sigma}^*_u) - \mathrm{vech}(\hat{\Sigma}_u)) \end{pmatrix}.$$

Conditional on the sample, we have [see Inoue and Kilian (2002)],

$$T^{1/2}\begin{pmatrix} \mathrm{vec}(\hat{\Phi}^{*}) - \mathrm{vec}(\hat{\Phi}) \\ \mathrm{vech}(\hat{\Sigma}_u^{*}) - \mathrm{vech}(\hat{\Sigma}_u) \end{pmatrix} \xrightarrow{d} N(0,\ \Omega), \tag{A.6}$$

where

$$\Omega = \begin{bmatrix} \Gamma^{-1}\otimes\Sigma_u & 0 \\ 0 & 2(D_m' D_m)^{-1}D_m'(\Sigma_u\otimes\Sigma_u)D_m(D_m' D_m)^{-1} \end{bmatrix},$$

D_m is the duplication matrix defined such that $\mathrm{vech}(F) = D_m\mathrm{vech}(F)$ for any symmetric $m\times m$ matrix F. Thus,

$$T^{1/2}[\hat{C}^{*}(Y \underset{h}{\longrightarrow} X \mid Z) - \tilde{C}(Y \underset{h}{\longrightarrow} X \mid Z)] \xrightarrow{d} N(0, \sigma_c(h)^2),$$

and

$$T^{1/2}[\hat{C}^{*}(Y \underset{h}{\longrightarrow} X \mid Z) - \hat{C}(Y \underset{h}{\longrightarrow} X \mid Z)] \xrightarrow{d} N(0, \sigma_c(h)^2)$$

where

$$\sigma_c(h)^2 = D_C\Omega D_C',\ D_C = \frac{\partial C(Y \underset{h}{\longrightarrow} X \mid Z)}{\partial(\mathrm{vec}(\Phi)',\ \mathrm{vech}(\Sigma_u)')}.$$

□

References

Berkowitz, J. and Kilian, L. (2000), 'Recent developments in bootstrapping time series', *Econometric Reviews* **19**, 1–48. With Comment by Russell Davidson, 49-54.

Bernanke, B. S. and Mihov, I. (1998), 'Measuring monetary policy', *The Quarterly Journal of Economics* **113**(3), 869–902.

Bhansali, R. J. (1978), 'Linear prediction by autoregressive model fitting in the time domain', *The Annals of Statistics* **6**, 224–231.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1992), 'Testing causality between two vectors in multivariate ARMA models', *Journal of the American Statistical Association* **87**(420), 1082–1090.

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

Diebold, F. X. and Kilian, L. (2001), 'Measuring predictability: Theory and macroeconomic applications', *Journal of Applied Econometrics* **16**, 657–669.

Dufour, J.-M. and Jouini, T. (2004), Asymptotic distribution of a simple linear estimator for VARMA models in echelon form, *in* P. Duchesne and B. Rémillard, eds, 'Statistical Modeling and Analysis for Complex Data Problems', Kluwer/Springer-Verlag, Canada, chapter 11, pp. 209–240.

Dufour, J.-M. and Pelletier, D. (2005), Practical methods for modelling weak VARMA processes: Identification, estimation and specification with a macroeconomic application, Technical report, Département de sciences économiques and CIREQ, Université de Montréal, Montréal, Canada.

Dufour, J.-M., Pelletier, D. and Renault, É. (2006), 'Short run and long run causality in time series: Inference', *Journal of Econometrics* **132**(2), 337–362.

Dufour, J.-M. and Renault, E. (1998), 'Short-run and long-run causality in time series: Theory', *Econometrica* **66**, 1099–1125.

Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Vol. 57 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, New York.

Geweke, J. (1982), 'Measurement of linear dependence and feedback between multiple time series', *Journal of the American Statistical Association* **77**, 304–313.

Geweke, J. (1984*a*), Inference and causality in economic time series, *in* Z. Griliches and M. D. Intriligator, eds, 'Handbook of Econometrics, Volume 2', North-Holland, Amsterdam, pp. 1102–1144.

Geweke, J. (1984*b*), 'Measures of conditional linear dependence and feedback between time series', *Journal of the American Statistical Association* **79**(388), 907–915.

Gouriéroux, C. and Monfort, A. (1997), *Time Series and Dynamic Models*, Cambridge University Press, Cambridge, U.K.

Gouriéroux, C., Monfort, A. and Renault, É. (1987), 'Kullback causality measures', *Annales d'Économie et de Statistique* **6/7**, 369–410.

Granger, C. W. J. (1969), 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica* **37**, 424–459.

Hannan, E. J. and Kavalieris, L. (1984*a*), 'A method for autoregressive-moving average estimation', *Biometrika* **71**(2), 273–280.

Hannan, E. J. and Kavalieris, L. (1984*b*), 'Multivariate linear time series models', *Advances in Applied Probability* **16**, 492–561.

Hannan, E. J. and Rissanen, J. (1982), 'Recursive estimation of mixed autoregressive-moving-average order', *Biometrika* **69**(1), 81–94. Errata 70 (1983), 303.

Hsiao, C. (1982), 'Autoregressive modeling and causal ordering of economic variables', *Journal of Economic Dynamics and Control* **4**, 243–259.

Inoue, A. and Kilian, L. (2002), 'Bootstrapping smooth functions of slope parameters and innovation variances in $VAR(\infty)$ models', *International Economic Review* **43**(2), 309–332.

Kilian, L. (1998), 'Small-sample confidence intervals for impulse response functions', *Review of Economics and Statistics* **80**(2), 218–230.

Koreisha, S. G. and Pukkila, T. M. (1989), 'Fast linear estimation methods for vector autoregressive moving-average models', *Journal of Time Series Analysis* **10**(4), 325–339.

Lewis, R. and Reinsel, G. C. (1985), 'Prediction of multivariate time series by autoregressive model fitting', *Journal of Multivariate Analysis* **16**, 393–411.

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

Lütkepohl, H. (1993*a*), *Introduction to Multiple Time Series Analysis*, second edn, Springer-Verlag, Berlin.

Lütkepohl, H. (1993*b*), Testing for causation between two variables in higher dimensional VAR models, *in* H. Schneeweiss and K. Zimmermann, eds, 'Studies in Applied Econometrics', Springer-Verlag, Heidelberg.

Newbold, P. (1982), Causality testing in economics, *in* O. D. Anderson, ed., 'Time Series Analysis: Theory and Practice 1', North-Holland, Amsterdam.

Paparoditis, E. (1996), 'Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes', *Journal of Multivariate Analysis* **57**, 277–296.

Parzen, E. (1974), ‘Some recent advances in time series modelling’, *IEEE Trans. Automat. Control* **AC-19**, 723–730.

Patterson, K. (2007), ‘Bias reduction through first-order mean correction, bootstrapping and recursive mean adjustment’, *Journal of Applied Statistics* **34**, 23–45.

Pierce, D. A. and Haugh, L. D. (1977), ‘Causality in temporal systems: Characterizations and survey’, *Journal of Econometrics* **5**, 265–293.

Polasek, W. (1994), Temporal causality measures based on AIC, *in* H. Bozdogan, ed., ‘Proceedings of the Frontier of Statistical Modeling: An Informal Approach’, Kluwer, Dordrecht, The Netherlands, pp. 159–168.

Polasek, W. (2002), Bayesian causality meaures for multiple ARCH models using marginal likelihoods, Technical report, Institute of Statistics and Econometrics, Univeristy of Basel., Switzerland.

Sims, C. (1980), ‘Macroeconomics and reality’, *Econometrica* **48**, 1–48.

Wiener, N. (1956), The theory of prediction, *in* E. F. Beckenback, ed., ‘The Theory of Prediction’, McGraw-Hill, New York, chapter 8.