Short and long run causality measures: theory and inference*

Jean-Marie Dufour †
McGill University

Abderrahim Taamouti ‡
Universidad Carlos III de Madrid

June 2009

* The authors thank Huberto Ennis, Kilian Lutz, Benoit Perron, Nadège-Désirée Yameogo, two anonymous referees, an Associate Editor, and the Editor Cheng Hsiao for several useful comments. Earlier versions of this paper were presented at the 45th Annual Meeting of the Société canadienne de science économique at Charlevoix (2005), the 2005 Joint Statistical Meetings of the American Statistical Association (Minneapolis), the 22nd CESG Conference at Vancouver (2005), the 2005 CIREQ Time Series Conference in Montréal, and the 40th Annual Meeting of the Canadian Economics Association at Montréal (2006). This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (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 Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec). Financial support from the Spanish Ministry of Education through grants SEJ 2007-63098 is also acknowledged.

† 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

‡ Economics Department, Universidad Carlos III de Madrid. Adresse and e-mail: Departamento de Economía Universidad Carlos III de Madrid Calle Madrid, 126 28903 Getafe (Madrid) España. TEL: +34-91 6249863; FAX: +34-91 6249329; e-mail: ataamout@eco.uc3m.es.

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 any given horizon h as well as tests for the corresponding noncausality [Dufour and Renault (1998), Dufour, Pelletier and Renault (2006)]. Instead of tests for noncausality at a given horizon, we study the problem of measuring causality between two vector processes. Existing causality measures have been defined only for the horizon 1 and fail to capture indirect causality. We propose generalizations to any horizon h of the measures introduced by Geweke (1982). Nonparametric and parametric measures of unidirectional causality and instantaneous effects are considered. On noting that the causality measures typically involve complex functions of model parameters in VAR and VARMA models, we propose a simple simulation-based method to evaluate these measures for any VARMA model. We also describe asymptotically valid nonparametric confidence intervals, based on 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.

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.

1. Introduction

The concept of causality introduced by Wiener (1956) and Granger (1969) constitutes a basic notion for studying dynamic relationships between time series. This concept is defined in terms of predictability at horizon one of a (vector) variable X from its own past, the past of another (vector) variable Y, and possibly a vector Z of auxiliary variables. In particular, Granger (1969) defines causality from Y to X in terms of *predictability one period ahead*: Y causes X if observations on Y up to time $t-1$ can help to predict $X(t)$ given the past of X and Z up to time $t-1$. More formally, Y causes X if the variance of the forecast error of X obtained by using the past of Y is smaller than the variance of the forecast error of X obtained without using the past of Y. In such a setup, the time order is used to distinguish between "input" variables (past values of different variables) and "output" variables (current values). Further, mere "correlations" that could be driven a third set of variables can be eliminated by allowing for the presence of auxiliary variables (Z). 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 focuses 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 extension is motivated by the following observation: in the presence of auxiliary variables (Z), even if Y does not cause X at horizon one, Y may cause X at a longer horizon $h > 1$. In such case, we have *indirect causality* transmitted by Z. Necessary and sufficient conditions of noncausality between vectors of variables at any horizon h for stationary and nonstationary processes have also been derived, but they are notably more complex for horizons longer than one, even in simple VAR models [Dufour and Renault (1998)].

This type of analysis distinguishes between three basic types of causality: from X to Y, from Y to X, and instantaneous causality. In practice, it is possible that all three causality relations coexist, hence the importance of finding means to quantify their degree. Unfortunately, causality tests fail to accomplish this task, because they only provide evidence on the presence or the absence of causality, and statistical significance depends on the available data and test power. A large effect may not be statistically significant (at a given level), and a statistically significant effect may not be "large" from an economic viewpoint (or more generally from the viewpoint of the subject at hand) or relevant for decision making. As emphasized by McCloskey and Ziliak (1996), it is crucial to distinguish between the numerical value of a parameter and its statistical significance. Indeed, the importance of this distinction was well understood by Neyman and Pearson (1933, p. 296): "*Is it more serious to convict an innocent man or to acquit a guilty? That will depend on the consequences of the error; is the punishment death or fine; what is the danger to the community of released criminals; what are the current ethical views on punishment? From the point of view of mathematical theory all that we can do is to show how the risk of errors may be controlled and minimized. The use of these statistical tools in any given case, in determining just how the balance should be struck, must be left to the investigator*".

In studying Wiener-Granger causality, predictability is the central issue. So, beyond accepting or rejecting noncausality hypotheses – which state that certain variables do not help forecasting other variables – we wish to assess the magnitude of the forecast improvement, where the latter is defined in terms of some loss function (such as mean-square error). Even if the hypothesis of no improvement (noncausality) cannot be rejected from looking at the available data (for example, because the sample size or the structure of the process do allow for high test power), sizeable improvements may remain consistent with the same data. Or by contrast, a statistically significant improvement – which may easily be produced by a large data set – may not be relevant from a practical viewpoint. Furthermore, in comparing causality properties at different horizons, we may wish to compare the magnitudes of improvement for different horizons. Reporting the results of causality tests at different horizons simply does not provide this type of information. This suggests that building causality measures at different horizons along with associated confidence intervals would yield a much more informative analysis of Granger-Wiener causality than tests of noncausality.

As pointed out by Geweke (1982), much research has been devoted to building and applying tests of noncausality. However, once it is admitted that a "causal relation" (in the sense of Granger) may be present, it is usually important to assess the strength of the relationship. This topic has attracted much less attention. To answer this type of question, Geweke (1982, 1984*b*) introduced measures of causality, based on mean-square forecast errors. 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 apparent only after several periods. Specifically, we propose generalizations to any horizon h of the measures proposed by Geweke (1982) for the horizon one. Important properties of these measures include: (1) they are nonnegative, and (2) they cancel only when there is no causality at the horizon considered. 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. When both causalities from X to Y and from Y to X do exist, the effect (or predictability) may be stronger in one of these directions. The causality measures studied do allow one to assess this for each horizon considered.

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 well in practice.

For empirical implementation, both parametric and nonparametric empirical estimates of the causality measures are considered. Parametric estimates can be derived - using the long simulation approach – from consistent estimates of the parameters of a VAR or VARMA model of known order and the associated impulse response coefficients. Nonparametric estimates involve approximating the process of interest by a long VAR whose degree increases at an appropriate rate.

Because of its simplicity, we focus on the second approach for empirical application. We show that the proposed nonparametric estimates are consistent, derive their asymptotic distribution under standard regularity conditions, and suggest a bootstrap technique to build confidence intervals based on these estimates.

The proposed causality measures can be applied in different contexts and may help 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 general theoretical framework which underlies 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 the measures and the asymptotic validity of 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 in the 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 causality induced by the auxiliary variables Z. Causality measures defined for the horizon 1 do not capture this indirect causality. This paper proposes causality measures at different horizons to quantify short- and long-run causality between random vectors. Such causality measures can detect and quantify the indirect causalities 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:

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

Since the coefficient of $Y(t)$ in the first equation of (2.1) is 0.7, we can conclude that Y causes X in the sense of Granger. However, this does not provide information on causality at horizons larger than 1. To study causality at horizon 2, consider the system (2.1) at time $t+2$:

$$\begin{bmatrix} X(t+2) \\ Y(t+2) \end{bmatrix} = \begin{bmatrix} 0.53 & 0.595 \\ 0.34 & 0.402 \end{bmatrix} \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} + \begin{bmatrix} 0.5 & 0.7 \\ 0.4 & 0.35 \end{bmatrix} \begin{bmatrix} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \end{bmatrix} + \begin{bmatrix} \varepsilon_X(t+2) \\ \varepsilon_Y(t+2) \end{bmatrix}.$$

The coefficient of $Y(t)$ in the equation for $X(t+2)$ is 0.595, so Y causes X at horizon 2. But, *how can one measure this "longer-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:

$$\begin{bmatrix} X(t+1) \\ Y(t+1) \\ Z(t+1) \end{bmatrix} = \begin{bmatrix} 0.60 & 0.00 & 0.80 \\ 0.00 & 0.40 & 0.00 \\ 0.00 & 0.60 & 0.10 \end{bmatrix} \begin{bmatrix} X(t) \\ Y(t) \\ Z(t) \end{bmatrix} + \begin{bmatrix} \varepsilon_X(t+1) \\ \varepsilon_Y(t+1) \\ \varepsilon_Z(t+1) \end{bmatrix}. \tag{2.2}$$

Since the coefficient of $Y(t)$ in the first equation for $X(t+1)$ is 0, Y does not cause X at horizon 1. At time $t+2$, then $X(t+2)$ is given by

$$X(t+2) = 0.36\, X(t) + 0.48 Y(t) + 0.56\, Z(t) + 0.6\varepsilon_X(t+1) + 0.8\varepsilon_Z(t+1) + \varepsilon_X(t+2). \tag{2.3}$$

The coefficient of $Y(t)$ in the latter equation is 0.48, which implies that Y causes X at horizon 2. This shows that the absence of causality at horizon 1 does not preclude causality at a longer horizon. This indirect causality is transmitted by the variable Z:

$$Y \underbrace{\rightarrow}_{0.60} Z \underbrace{\rightarrow}_{0.80} X$$

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 degree of this indirect causality?* 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 $X(t) = (x_1(t), \ldots, x_{m_1}(t))'$, $Y(t) = (y_1(t), \ldots, y_{m_2}(t))'$, $Z(t) = (z_1(t), \ldots, z_{m_3}(t))'$, $m_1 \geq 1$, $m_2 \geq 1$, $m_3 \geq 0$, and $x_i(t) \in L^2$, $y_i(t) \in L^2$, $z_i(t) \in L^2$, for all i. Further, we let H be a (possibly empty) Hilbert subspace 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 $I_Y(t)$, $I_Z(t)$, 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 $\sigma^2[x_i(t+h)\,|\,B_t] = \mathsf{E}\{u[x_i(t+h)|\,B_t]^2\}$. 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

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

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. *Let $h \geq 1$. (i) Y does not cause X at horizon h given I [denoted $Y \underset{h}{\nrightarrow} X\,|\,I$] iff $P[X(t+h)|\,I_X(t)] = P[X(t+h)\,|\,I_{XY}(t)],\ \forall t > \omega,$ 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 $Y \underset{k}{\nrightarrow} X\,|\,I$ for $k = 1,\, 2,\, \ldots\,,\, h$; (iii) Y does not cause X at any horizon given I [denoted $Y \underset{(\infty)}{\nrightarrow} X\,|\,I$] iff $Y \underset{k}{\nrightarrow} X\,|\,I$ for all $k = 1,\, 2,\, \ldots$.*

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. Clearly if $X(\omega,\, t] \subseteq I(t)$, then $I_X(t) = I(t)$. So, if the conformability assumption is added, Definition 3.1 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.* $X(\omega,\, t] \nsubseteq I(t)\,,\ Y(\omega,\, t] \nsubseteq I(t)\,,\ Z(\omega,\, t] \subseteq I(t)\,.$

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 3.1 COVARIANCE CHARACTERIZATION OF NON-CAUSALITY AT HORIZON h. *Let $h \geq 1$. (i) Y does not cause X at horizon h given I iff*

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

where $\varSigma[X(t+h)\,|\,\cdot\,]$ is defined by (3.5); *(ii) Y does not cause X up to horizon h given I iff*

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

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

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

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 3.2 UNCONDITIONAL NON-CAUSALITY AT HORIZON h. *Let $h \geq 1$. (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*

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

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 $Y \underset{k}{\nrightarrow} X\,|\,I_{(Z)}$ 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 $Y \underset{k}{\nrightarrow} X\,|\,I_{(Z)}$ 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 3.1 also holds for $Z-$unconditional non-causality.

4. Causality measures

We will now develop multi-horizon 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 also be 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* $h \geq 1$,

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

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:

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

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

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

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

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

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* $W(t) = \big(X(t)',\,Y(t)'\big)'$, *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:

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

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

where

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

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 $h \geq 1$,*

$$
C_L^{(h)}(X,\, Y \mid I) = C_L(X \underset{h}{\rightarrow} Y \mid I) + C_L(Y \underset{h}{\rightarrow} X \mid I) + C_L(X \underset{h}{-} Y \mid I) \tag{4.2}
$$

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:

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

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

$$
\begin{aligned}
C(Y \underset{h}{\rightarrow} X) = C_L(Y \underset{h}{\rightarrow} X \mid I)\,,\ C(Y \underset{h}{\rightarrow} X \mid Z) = C_L(Y \underset{h}{\rightarrow} X \mid I_{(Z)}), \\
C(X \underset{h}{-} Y) = C_L(X \underset{h}{-} Y \mid I)\,,\ C^{(h)}(X,\, Y) = C_L^{(h)}(X,\, Y \mid I)\,.
\end{aligned}
$$

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). For $m_1 = m_2 = 1$, 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, $U[X(t+s) \mid \Omega_t] = X(t+s) - P[X(t+s) \mid \Omega_t]$, 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:

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

Then, we have the following relationships.

Proposition 4.1 RELATION OF CAUSALITY MEASURES WITH PREDICTABILITY MEASURES. *Let h_1 and h_2 be two different horizons, $m_1 = m_2 = 1$, and*

$$\bar{P}_X\big(I_X(t),\, h_1,\, h_2\big) = 1 - \frac{\sigma^2\big(X(t+h_1)\,|\, I_X(t)\big)}{\sigma^2\big(X(t+h_2)\,|\, I_X(t)\big)}\,, \tag{4.3}$$

$$\bar{P}_X\big(I_{XY}(t),\, h_1,\, h_2\big) = 1 - \frac{\sigma^2\big(X(t+h_1)\,|\, I_{XY}(t)\big)}{\sigma^2\big(X(t+h_2)\,|\, I_{XY}(t)\big)}\,, \tag{4.4}$$

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

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

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

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

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 reduced-form impulse responses.

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 $W(t) = (X(t)', Y(t)', Z(t)')'$ is a VARMA(p, q) model, hereafter the *unconstrained* model, where p and q can be infinite. The structure of the process $W_0(t) = (X(t)', Z(t)')'$, hereafter the *constrained* model, can be deduced from the unconstrained model using Corollary 6.1.1 in Lütkepohl (1993*a*, pages 308-309). 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.

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

$$\varPhi(L)W(t) = \varTheta(L)u(t) \tag{5.1}$$

where $m = m_1 + m_2 + m_3$,

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

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

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

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

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

$$W(t) = \Psi(L)u(t)\,, \tag{5.2}$$

$$\Psi(L) = \Phi(L)^{-1}\Theta(L) = \sum_{j=0}^{\infty} \Psi_j L^j = \sum_{j=0}^{\infty} \begin{bmatrix} \psi_{XXj} & \psi_{XYj} & \psi_{XZj} \\ \psi_{YXj} & \psi_{YYj} & \psi_{YZj} \\ \psi_{ZXj} & \psi_{ZYj} & \psi_{ZZj} \end{bmatrix} L^j, \; \Psi_0 = I_m. \tag{5.3}$$

From Section 4, 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 noncausality). This one can be deduced from the structure of the unconstrained model (5.1) using the following proposition and corollary [Lütkepohl (1993*a*, pages 231-232)].

Lemma 5.1 LINEAR TRANSFORMATION OF A VMA(q) PROCESS. *Let $u(t)$ be a m-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 m-dimensional invertible VMA(q) process. Furthermore, let F be an $(\bar{m} \times m)$ matrix of rank $\bar{m}$. Then the $\bar{m}$-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 $\bar{m}$-dimensional white noise with nonsingular variance-covariance matrix Σ_ε, the $\bar{\theta}_j$, $j = 1, \ldots, \bar{q}$, are $\bar{m} \times \bar{m}$ coefficient matrices and $\bar{q} \leq q$.

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

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

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 $\bar{p} \leq mp$ and $\bar{q} \leq (m-1)p$. Suppose now we are interested in measuring causality from Y to X at a given horizon h. We need to apply Lemma 5.2 to obtain the structure of process $W_0(t) = (X(t)', Z(t)')'$. If we left-multiply equation

(5.1) by the adjoint matrix of $\Phi(L)$, say $\Phi(L)^*$, we get

$$\Phi(L)^*\Phi(L)W(t) = \Phi(L)^*\Theta(L)u(t) \tag{5.4}$$

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 $\bar{p} \le mp$. 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 (5.4) can be written as follows:

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

This equation is another stationary invertible VARMA representation of process $W(t)$, called the final equation form. The model of the process $W_0(t) = (X(t)',\ Z(t)')'$ can be obtained by choosing F in Lemma 5.2 as

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

On premultiplying (5.5) by F, we get

$$\det\{\Phi(L)\} W_0(t) = F\Phi(L)^*\Theta(L)u(t). \tag{5.6}$$

The right-hand side of (5.6) is a linearly transformed finite-order VMA process which, by Lemma 5.1, has a VMA($\bar{q}$) representation with $\bar{q} \le p(m-1)+q$. Thus, we get the model:

$$\det\{\Phi(L)\} W_0(t) = \bar{\theta}(L)\varepsilon(t) = \begin{bmatrix} \bar{\theta}_{XX}(L) & \bar{\theta}_{XZ}(L) \\ \bar{\theta}_{ZX}(L) & \bar{\theta}_{ZZ}(L) \end{bmatrix} \varepsilon(t), \tag{5.7}$$

$$\bar{\theta}_{ll}(L) = I_{m_l} + \sum_{j=1}^{\bar{q}} \bar{\theta}_{llj} L^j,\ \bar{\theta}_{lk}(L) = \sum_{j=1}^{\bar{q}} \bar{\theta}_{lkj} L^j, \text{ for } l \ne k \text{ and } l,\ k = X,\ Z, \tag{5.8}$$

$$\mathsf{E}[\varepsilon(t)] = 0,\ \mathsf{E}\big[\varepsilon(t)\varepsilon(s)'\big] = \begin{cases} \Sigma_\varepsilon & \text{for } s = t \\ 0 & \text{for } s \ne t \end{cases}. \tag{5.9}$$

The coefficients $\bar{\theta}_{lkj}$ and the elements of Σ_ε are functions of Σ_u, φ_{lki}, Θ_{lkj}, $l, k = X,\ Z,\ Y$, $1 \le i \le p$, $1 \le j \le q$. This is possible by solving the following system:

$$\gamma_\varepsilon(v) = \gamma_u(v),\ v = 0,\ 1,\ 2,\ \ldots, \tag{5.10}$$

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.

Example 5.1 Suppose we have two variables X and Y. If we assume that the joint process $(X,Y)'$ follows a stationary VAR(1) model, then the marginal process X is an ARMA$(2,1)$. The parameters of the autoregressive part of the ARMA$(2,1)$ model are functions of the VAR(1) parameters. However, the moving average coefficient [say $\bar{\theta}$] of the ARMA$(2,1)$ and the variance of new error terms [say $\sigma^2_{\varepsilon_X}$] can be determined by solving the system (5.10) for $v = 0$ and $v = 1$. These parameters must satisfy the constraints $|\,\bar{\theta}\,| < 1$ and $\sigma^2_{\varepsilon_X} > 0$.

The VMA(∞) representation of model (5.7) is given by

$$W_0(t) = \det\{\Phi(L)\}^{-1}\bar{\theta}(L)\varepsilon(t) = \sum_{j=0}^{\infty}\bar{\Psi}_j\varepsilon(t-j) = \sum_{j=0}^{\infty}\begin{bmatrix} \bar{\psi}_{XX_j} & \bar{\psi}_{XZ_j} \\ \bar{\psi}_{ZX_j} & \bar{\psi}_{ZZ_j} \end{bmatrix}\begin{bmatrix} \varepsilon_X(t-j) \\ \varepsilon_Z(t-j) \end{bmatrix} \tag{5.11}$$

where $\bar{\Psi}_0 = I_{m_1+m_2}$. To quantify the degree of causality from Y to X at horizon h, we need to calculate the variance-covariance matrices of the unconstrained and constrained forecast errors of $X(t+h)$. From (5.2), the forecast error of $W(t+h)$ and its covariance matrix are given by

$$U\left[W(t+h)\,|\,I_W(t)\right] = \sum_{i=0}^{h-1}\Psi_i u(t+h-i)\,, \tag{5.12}$$

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

Consequently, the variance-covariance matrix of the unconstrained forecast error of $X(t+h)$ is

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

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

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

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

Then the variance-covariance matrix of the constrained forecast error of $X(t+h)$ is

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

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 [Definition 4.1].

Theorem 5.1 REPRESENTATION OF CAUSALITY MEASURE IN TERMS OF REDUCED-FORM IMPULSE RESPONSES. *Under Assumptions* 5.1 *and* 5.2*,*

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

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. For a bivariate VAR(1) model, it is relatively easy to analytically compute the causality measures at any horizon h using only the unconstrained parameters [see Example 5.1].

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

$$\Sigma\big(X(t+h),\, Y(t+h)\,|\, I_W(t)\big) = \sum_{i=0}^{h-1} G\,\Psi_i \Sigma_u \Psi_i' G'\,,\ G = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & I_{m_2} & 0 \end{bmatrix}.$$

Consequently,

$$\Sigma\big(X(t+h)\,|\ I_W(t)\big) = \sum_{i=0}^{h-1} J_1 \Psi_i \Sigma_u \Psi_i' J_1'\,, \quad \Sigma\big(Y(t+h)\,|\ I_W(t)\big) = \sum_{i=0}^{h-1} J_2 \Psi_i \Sigma_u \Psi_i' J_2',$$

where $J_1 = \begin{bmatrix} I_{m_1} & 0 & 0 \end{bmatrix}$ and $J_2 = \begin{bmatrix} 0 & I_{m_2} & 0 \end{bmatrix}$. We can immediately deduce the following result by using the definition of the instantaneous causality measure [see Definition 4.3].

Theorem 5.2 REPRESENTATION OF THE INSTANTANEOUS CAUSALITY MEASURE IN TERMS OF REDUCED-FORM IMPULSE RESPONSES. *Under Assumptions* 5.1 *and* 5.2*,*

$$C_L(X \underset{h}{-} Y\,|\, I) = \ln\left[\frac{\det\big\{\sum_{i=0}^{h-1}\big(J_1\Psi_i\Sigma_u\Psi_i'J_1'\big)\big\}\ \det\big\{\sum_{i=0}^{h-1}\big(J_2\Psi_i\Sigma_u\Psi_i'J_2'\big)\big\}}{\det\big\{\sum_{i=0}^{h-1}\big(G\,\Psi_i\Sigma_u\Psi_i'G'\big)\big\}}\right]$$

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 (4.2).

Let us finally consider the special case where the process $W(t) = (X(t)', Y(t)', Z(t)')'$ follows an invertible VMA(q) model:

$$W(t) = \Theta(L)u(t) \tag{5.14}$$

with

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

$\theta_{ll}(L) = I_l + \sum_{j=1}^{q}\theta_{llj}L^j,\ \theta_{lk}(L) = \sum_{j=1}^{q}\theta_{lkj}L^j$, for $l \neq k,\ l,\ k = X, Z, Y$, and I_l is an identity matrix. The following result then follows from Proposition 4.1.

Theorem 5.3 CHARACTERIZATION OF CAUSALITY MEASURES FOR VMA(q). *Let* h_1 *and* h_2 *be two different horizons. Under Assumption* 5.14 *and for* $h_2 \geq h_1 \geq q$,

$$C_L(Y \underset{h_1}{\longrightarrow} X\,|\, I) = C_L(Y \underset{h_2}{\longrightarrow} X\,|\, I).$$

6. Estimation

We know that short-run and long-run causality measures depend on the parameters of the model describing the process of interest [see Section 5]. 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 in practice especially when the number of parameters is large, as it is typically the case in VARMA models. The last approach assumes that the process is autoregressive with potentially infinite order, but 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) and 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 $W(t) = (X(t)', Y(t)', Z(t)')'$. 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 $\sum_{j=0}^{\infty} \| \Psi_j \| < \infty$ and $\det\{\Psi(z)\} \neq 0$ for $z \in \mathbb{C}$ and $|z| \leq 1$, where $\| \Psi_j \| = tr(\Psi_j \Psi_j)$ and $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$, with $\Psi_0 = I_m$ an $m \times m$ identity matrix. Under the latter assumptions, $W(t)$ is invertible and can be written as an infinite autoregressive process:

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

where $\sum_{j=1}^{\infty} \| \Phi_j \| < \infty$ and $\Phi(z) = I_m - \sum_{j=1}^{\infty} \Phi_j z^j = \Psi(z)^{-1}$ satisfies $\det\{\Phi(z)\} \neq 0$ 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:

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

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

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

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

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

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 $W_0(t) = (X(t)', Z(t)')'$. 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:

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

Model (6.2) can also be approximated by a a finite-order VAR(k) model, where k depends on the sample size T :

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

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.

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 $\varepsilon_k(t)$ are given by the following equation:

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

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, using the corresponding fitted VAR(k) models. The variance-covariance matrix of the unconstrained forecast error of $W(t+h)$ based on the $VAR(\infty)$ model is given by

$$\Sigma(h) = \sum_{j=0}^{h-1} \Psi_j \Sigma_u \Psi_j' \tag{6.3}$$

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

$$\Phi_1^{(j+1)} = \Phi_2^{(j)} + \Phi_1^{(j)}\, \Phi_1,\ \Phi_1^{(1)} = \Phi_1,\ \Phi_1^{(0)} = I_m,\ \text{ for } j \geq 1; \tag{6.4}$$

see Dufour and Renault (1998). An estimator of the variance-covariance matrix (6.3) based on the fitted $VAR(k)$ model is given by

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

where $\hat{\Psi}_{jk} = \hat{\Phi}_{1k}^{(j)}$ and $\hat{\Phi}_{1k}^{(j)}$ are calculated using (6.4) [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

$$\varSigma_0(h) = \sum_{j=0}^{h-1} \bar{\varPsi}_j \varSigma_\varepsilon \bar{\varPsi}'_j \tag{6.6}$$

where $\bar{\varPsi}_j = \bar{\varPhi}_1^{(j)}$ and $\bar{\varPhi}_1^{(j)}$ are defined in similar way as in (6.4). Furthermore, an estimator of the variance-covariance matrix (6.6) based on the fitted $VAR(k)$ model is given by

$$\tilde{\varSigma}_{0|k}(h) = \sum_{j=0}^{h-1} \tilde{\varPsi}_{jk} \tilde{\varSigma}_{\varepsilon|k} \tilde{\varPsi}'_{jk} \tag{6.7}$$

where $\tilde{\varPsi}_{jk}$ is an estimate of the corresponding population parameter $\bar{\varPsi}_j$. Consequently, from Theorem **5.1** an estimator of the causality measure from Y to X at horizon h is given by

$$\hat{C}_L(Y \underset{h}{\longrightarrow} X \mid I) = \ln \left[\frac{\det\left\{ \left[J_0 \tilde{\varSigma}_{0|k}(h) J'_0 \right] \right\}}{\det\left\{ \left[J_1 \hat{\varSigma}_k(h) J'_1 \right] \right\}} \right]. \tag{6.8}$$

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 considered.

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* $k^{1/2} \sum_{j=k+1}^{\infty} \parallel \varPhi_j \parallel \to 0$ *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 $\hat{\varPhi}(k)$ 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 the consistency of $\hat{\varPhi}(k)$. Furthermore, their Theorem 4 derives the asymptotic distribution for $\hat{\varPhi}(k)$ under Assumption 6.1 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 $\tilde{\bar{\varPhi}}(k)$ converges to $\bar{\varPhi}(k)$ and asymptotically follows a normal distribution . Finally, we note that $\hat{\varSigma}_{u|k}$ $(\tilde{\varSigma}_{\varepsilon|k})$ converges to $\varSigma_{u|k}$ $(\varSigma_{\varepsilon|k})$, as k and $T \to \infty$ [Lütkepohl (1993*a*, pages 308-309)].

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

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 for very simple specifications, it is quite difficult to derive 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 unconstrained and constrained forecast errors at horizon h [see Section 6].

4. Calculate the causality measure at horizon h using for example (6.8).

To see better how this works, consider again Example 2.1 of Section 2. Our illustration involves two steps. First, we calculate the theoretical values of the causality measures at horizons 1 and 2. We know from Example 5.1 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:

$$\sigma^2[X(t+1) \mid I_{XY}(t)] = 1, \; \sigma^2[X(t+2) \mid I_{XY}(t)] = 1.74.$$

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 of X. This one is given by the following equation:

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

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

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

Table 1. Evaluation by simulation of causality measures at horizons 1 and 2 for Model (2.1)

p	$C_L(Y \underset{1}{\longrightarrow} X)$	$C_L(Y \underset{2}{\longrightarrow} X)$
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

Table 2. Evaluation by simulation of causality measures at horizons 1 and 2 for Model (2.2)

p	$C_L(Y \underset{1}{\longrightarrow} X \mid I)$	$C_L(Y \underset{2}{\longrightarrow} X \mid I)$
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

The set of possible solutions is $\big\{(\bar{\theta},\, \sigma^2_{\varepsilon_X}) = (-4.378,\, 0.08),\ (-0.2285,\, 1.53)\big\}$. 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 $\sigma^2[X(t+1) \mid I_X(t)] = 1.53,$ and the variance of the forecast error of X at horizon 2 is $\sigma^2[X(t+2) \mid I_X(t)] = 2.12$. Consequently,

$$C_L(Y \underset{1}{\longrightarrow} X) = 0.425,\ \ C_L(Y \underset{2}{\longrightarrow} X) = 0.197. \tag{7.1}$$

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 correspond to the theoretical values of the causality measures calculated in (7.1) and confirm the convergence ensured by the law of large numbers.

Now consider Example 2.2 of Section 2. In this example, analytical calculation of causality measures is not easy. In model (2.2) 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 $W(t) = (X(t)', Y(t)', Z(t)')'$ follows a VAR(p) model:

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

where $\{u(t)\}_{t=0}^{\infty}$ is a sequence of *i.i.d.* random variables with zero mean and a positive definite variance-covariance matrix Σ_u, 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 I_m is an $m \times m$ identity matrix.[1]

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

$$\hat{\Phi} = \hat{\Gamma}_1' \hat{\Gamma}^{-1}, \;\; \hat{\Sigma}_u = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}(t)\hat{u}(t)', \tag{8.1}$$

$$\hat{\Gamma} = \frac{1}{T-p} \sum_{t=p+1}^{T} w(t)w(t)', \;\; \hat{\Gamma}_1 = \frac{1}{T-p} \sum_{t=p+1}^{T} w(t)W(t+1)', \tag{8.2}$$

where $w(t) = (W(t)', \ldots, W(t-p+1)')'$, 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 $W_0(t) = (X(t)', Z(t)')'$. This one has a VARMA($\bar{p}$, $\bar{q}$) representation with $\bar{p} \leq mp$ and $\bar{q} \leq (m-1)p$,

$$\Phi^c(L) W_0(t) = \theta^c(L)\varepsilon(t) \tag{8.3}$$

where $\Phi^c(L) = I_{\bar{m}} - \Phi_1^c L - \cdots - \Phi_{\bar{p}}^c L^{\bar{p}}$, $\theta^c(L) = I_{\bar{m}} + \theta_1^c L + \cdots + \theta_{\bar{q}}^c L^{\bar{q}}$, for $\bar{m} = m_3 + 1$, $I_{\bar{m}}$ an $\bar{m} \times \bar{m}$ identity matrix, and $\{\varepsilon(t)\}_{t=0}^{\infty}$ is a sequence of uncorrelated random variables with zero mean and a positive definite variance-covariance matrix Σ_ε. 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:

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

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

$$W_0(t) = \sum_{j=1}^{k} \bar{\Phi}_{jk} W_0'(t-j) + \varepsilon_k(t).$$

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 $\varepsilon_k(t)$ are given by the following equa-

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

tion:

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

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

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

$$C_L(Y \underset{h}{\longrightarrow} X \mid I) = \ln\left[\frac{\det\left\{J_0\Sigma_0'(h)J_0'\right\}}{\det\left\{J_1\Sigma(h)J_1'\right\}}\right], \ \Sigma_0(h) = \sum_{j=0}^{h-1}\bar{\Psi}_j\Sigma_\varepsilon\bar{\Psi}_j', \ \Sigma(h) = \sum_{j=0}^{h-1}\Psi_j\Sigma_u\Psi_j',$$

where $\bar{\Psi}_j = \bar{\Phi}_1^{(j)}$, $\Psi_j = \Phi_1^{(j)}$ and $\Phi_1^{(j)}$ and $\bar{\Phi}_1^{(j)}$ are defined in similar way as in (6.4). Using Lemma 5.2 $C_L(Y \underset{h}{\longrightarrow} X \mid I)$ may be written as follows:

$$C_L(Y \underset{h}{\longrightarrow} X \mid I) = \ln\left[\frac{\det\left\{G\big(\Phi, \Sigma_u\big)\right\}}{\det\left\{H\big(\Phi, \Sigma_u\big)\right\}}\right],$$

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

where $G(\cdot)$ and $H(\cdot)$ are continuous and differentiable *functions of* $\big(\Phi, \Sigma_u\big)$. A consistent estimator of $C_L(Y \underset{h}{\longrightarrow} X \mid I)$ is given by

$$\hat{C}_L(Y \underset{h}{\longrightarrow} X \mid I) = \ln\left[\frac{\det\left\{J_0\tilde{\Sigma}_{0|k}(h)J_0'\right\}}{\det\left\{J_1\hat{\Sigma}(h)J_1'\right\}}\right], \ \tilde{\Sigma}_{0|k}(h) = \sum_{j=0}^{h-1}\tilde{\Psi}_{jk}\tilde{\Sigma}_{\varepsilon|k}\tilde{\Psi}_{jk}', \ \hat{\Sigma}(h) = \sum_{j=0}^{h-1}\hat{\Psi}_j\hat{\Sigma}_u\hat{\Psi}_j',$$

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

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

$$T^{1/2}\begin{pmatrix} \mathrm{vec}(\hat{\Phi}) - \mathrm{vec}(\Phi) \\ \mathrm{vech}(\hat{\Sigma}_u) - \mathrm{vech}(\Sigma_u) \end{pmatrix} \xrightarrow{d} N[0, \ \Omega] \tag{8.5}$$

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

$$\Omega = \begin{bmatrix} \Gamma^{-1}\otimes\Sigma_u & 0 \\ 0 & 2(D_m'D_m)^{-1}D_m'(\Sigma_u\otimes\Sigma_u)D_m(D_m'D_m)^{-1} \end{bmatrix}, \tag{8.6}$$

and 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. Thereafter, we will consider the following assumptions.

Assumption 8.1 *The following conditions are satisfied:*

(1) $E\,|\,\varepsilon_h(t)\varepsilon_i(t)\varepsilon_j(t)\varepsilon_l(t)\,| \le \gamma_4 < \infty$, *for* $1 \le h, i, j, l \le \bar{m}$; *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 $k^{1/2}\sum_{j=k+1}^{\infty}\parallel \bar{\Phi}_j \parallel \rightarrow 0$ as $k, T \rightarrow \infty$;*

(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 8.1 ASYMPTOTIC DISTRIBUTION OF CAUSALITY MEASURES. *Under Assumptions 8.1, we have:*

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

where $\sigma_c(h)^2 = D_C \Omega D_C^{'}$, $D_C = \partial C_L(Y\underset{h}{\longrightarrow}X \mid I)\,/\,\partial\theta'$, $\theta = \big(\mathrm{vec}(\Phi)',\, \mathrm{vech}(\Sigma_u)'\big)'$ *and* Ω *is given by* (8.6).

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 of 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

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

$\hat{\Phi} = [\hat{\Phi}_1, \ldots, \hat{\Phi}_p]$ is given by (8.1) and the OLS estimate of Σ_u is given by $\tilde{\Sigma}_u = \sum_{t=p+1}^{T}\tilde{u}(t)\tilde{u}(t)'/(T-p)$, where $\tilde{u}(t) = \hat{u}(t) - \sum_{t=p+1}^{T}\hat{u}(t)/(T-p)$ and $\hat{u}(t) = W(t) - \sum_{i=1}^{p}\hat{\Phi}_i W(t-i)$.

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 $w^*(0) = (W(1)', \ldots, W(p)')'$. [2]

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:

$$W^*(t) = \sum_{i=1}^{p}\hat{\Phi}_i W^*(t-i) + u^*(t), \text{ for } t = p+1, \ldots, T. \quad (8.7)$$

[2]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)].

5. Calculate the bootstrap OLS regression estimates

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

where

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

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

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

7. Calculate the causality measure at horizon h, say $\hat{C}_L^{(j)*}(Y \underset{h}{\longrightarrow} X \mid I)$, using equation (6.8) and the bootstrap sample.

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

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

Proposition 8.2 ASYMPTOTIC VALIDITY OF THE RESIDUAL-BASED BOOTSTRAP. *Under Assumptions 8.1, we have*

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

where $\sigma_c(h)^2$ and Ω are defined in Proposition 8.1.

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 can consider a similar algorithm to the one 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 $Bias^* = \mathsf{E}^*[\hat{\Phi}^* - \hat{\Phi}]$, 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 (8.7) 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, one can consider another bias correction applied directly on the measure itself,

$$\tilde{C_L}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid I) = \hat{C}_L^{(j)*}(Y \underset{h}{\longrightarrow} X \mid I) - \left[\bar{C}_L^{*}(Y \underset{h}{\longrightarrow} X \mid I) - \hat{C}_L(Y \underset{h}{\longrightarrow} X \mid I)\right],$$

$$\bar{C}_L^{*}(Y \underset{h}{\longrightarrow} X \mid I) = \frac{1}{B}\sum_{j=1}^{B}\tilde{C_L}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid I).$$

In practice, specially when the true value of causality measure is close to zero, it is possible that for some bootstrap samples the quantity $\tilde{C_L}^{(j)*}(Y \underset{h}{\longrightarrow} X \mid I)$ becomes negative. In this case we impose the following non-negativity truncation:

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

9. Empirical application

We apply our causality measures to examine whether or not monetary policy can cause at different horizons the real economy and vice versa. Further, we compare the causal effects that different measures of monetary policy can have on the real economy, where the latter is represented by gross domestic product. We follow Bernanke and Blinder (1992) and Bernanke and Mihov (1998) to consider as measures of monetary policy nonborrowed reserves and federal funds rate.

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), federal funds rate (R), gross domestic product deflator (P), and real gross domestic product (GDP). The monthly data on GDP and 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. These variables were also transformed by taking first differences, consequently the causality relations have to be interpreted in terms of growth rates.

We performed augmented Dickey-Fuller tests (hereafter ADF-tests) for nonstationarity of the four variables of interest and their first differences. The results show that all variables in logarithmic form are nonstationary. However, their first differences are stationary except for the GDP deflator. We performed a nonstationarity test for the second differences of the GDP deflator. 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. The critical values in both cases are equal to -2.8695 and -3.4235. Thus, the second differences of the GDP deflator are stationary.

Once the data is made stationary, we use the nonparametric approach described in Section 6 for the estimation and Akaike's information criterion to specify the orders of the corresponding fitted VAR(k) models. Applying Akaike's criterion for the unconstrained VAR model, which corresponds to four variables, we find that it is minimized at order $k = 16$. The orders of the constrained VAR models, which correspond to different combinations of three variables, are all less than or equal to

Figures 1 -8. Causality profiles

These figures present measures of causality between nonborrowed reserves (NBR), federal funds rate (R), the gross domestic product deflator (P), and real gross domestic product (GDP), until 40 months. The sample runs from January 1965 to December 1996 for a total of 384 observations.

Figure 1: Causality measures from Nonborrowed reserves to Federal funds rate

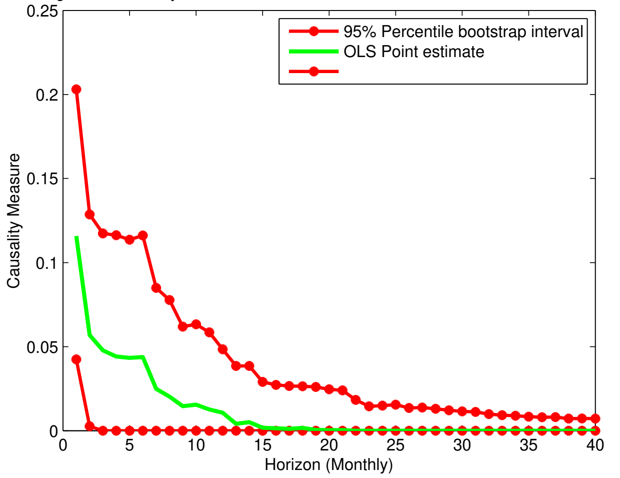


Figure 2: Causality measures from Nonborrowed reserves to GDP Deflator

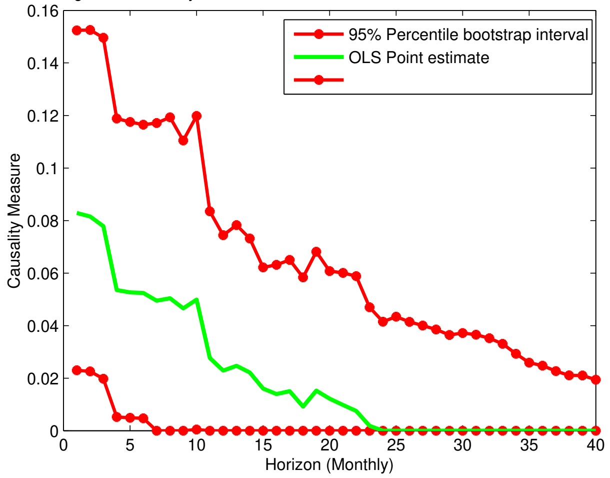


Figure 3: Causality measures from Nonborrowed reserves to Real GDP

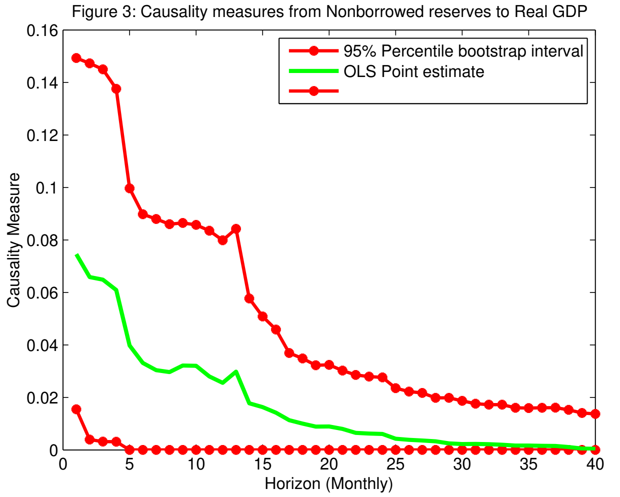


Figure 4: Causality measures from Federal funds rate to GDP Deflator

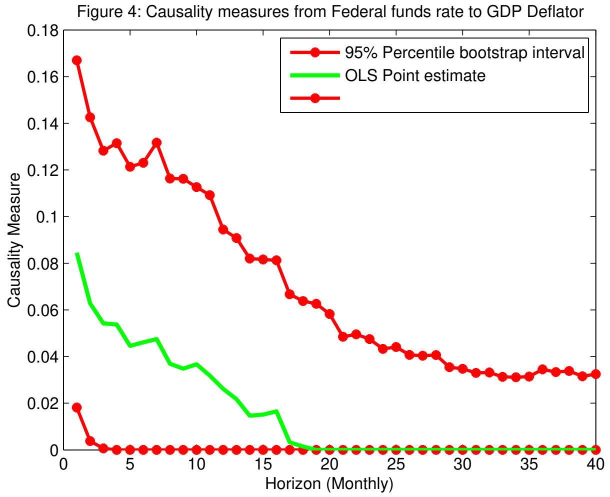


Figure 5: Causality measures from Federal funds rate to Real GDP

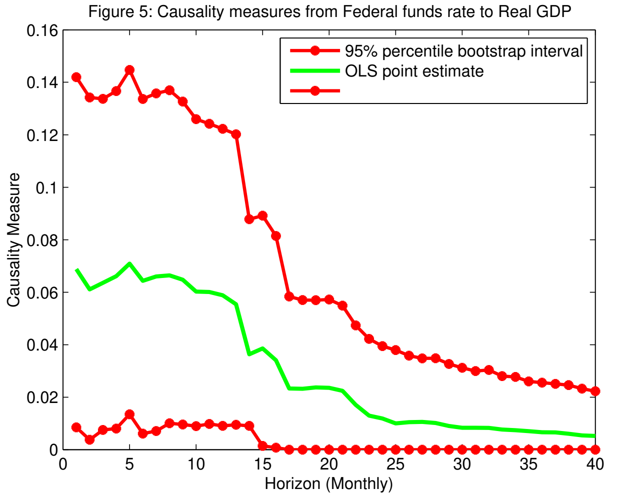


Figure 6: Causality measures from Real GDP to Federal funds rate

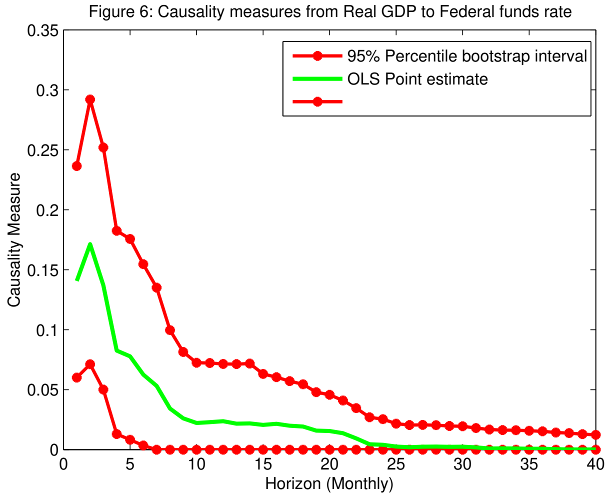


Figure 7: Comparison the impacts of NBR and R on Real GDP

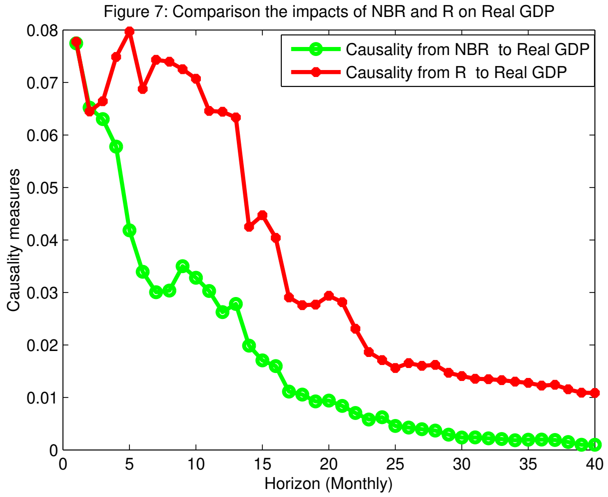
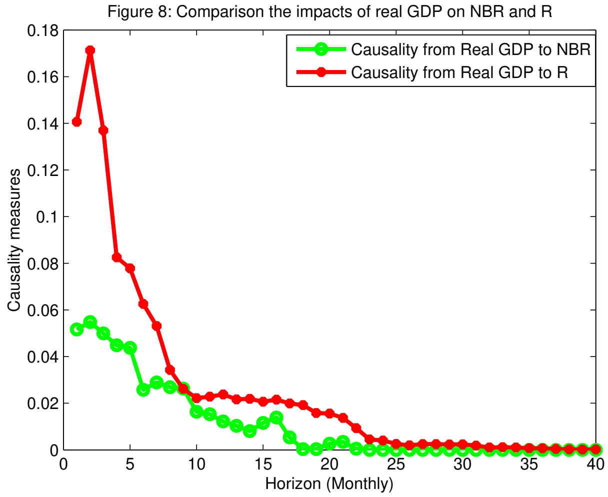


Figure 8: Comparison the impacts of real GDP on NBR and R



16. To calculate the causality measures we consider the same order $k = 16$ for the constrained and unconstrained models. This ensures a relevant comparison of the determinants of the variance-covariance matrices of the constrained and unconstrained forecast errors. We compute the causality measures for horizons $h = 1, \ldots, 40$ [see Figure 1- 8]. Higher values of the measures indicate larger causality. We also calculate the corresponding nominal 95% bootstrap confidence intervals as described in the previous section.

Figure 1 shows that nonborrowed reserves strongly Granger cause the federal funds rate (one period ahead). This causality is well known in the literature and can be explained by the theory of supply and demand for money. The nonborrowed reserves also cause the real GDP and GDP deflator: these causality measures are not significantly different from zero after 4 months for real GDP and after 6 months for the GDP deflator, although the confidence intervals still cover sizeable values. The causality measures from the federal funds rate to the GDP deflator decline gradually, although they are significantly different from zero only for the two-months horizon [see Figure 4]. We also observe that real GDP strongly causes the federal funds rate during the first three months. This causality is statistically significant for the first 4 or 5 months. Another interesting result is the one which corresponds to the causality from the federal funds rate to real GDP. This causality is sizeable (in the range of 0.05) and statistically significant during the first 15 months, after which it becomes statistically non significant although possibly sizeable from an economic viewpoint [see Figure 5]. These findings are consistent with conclusions obtained by Dufour et al. (2006). Statistically significant causality measures – in the sense that the 95% confidence intervals do not cover zero – are summarized in Table 3. The above results do not change substantially when we consider second – rather than first – differences of the GDP deflator.

From the above results, we draw the following conclusions. First, monetary policy – as measured by nonborrowed reserves and federal funds – affects the real economy. Second, nonborrowed reserves cause federal funds rate in the short-term, which then causes real GDP over several months (see figures 1 and 5). These findings are consistent with the literature; see for example Bernanke and Blinder (1992), Bernanke and Mihov (1998), and Ireland (2005). Third, except for the first two months, where the both measures of monetary policy have the same degree of causal effect on real GDP, the federal funds rate causes the real economy more strongly than nonborrowed reserves at longer horizons (see Figure 7). The latter result is also supported by the literature: "*The results are striking: the Federal funds rate is markedly superior to the both monetary aggregates* [M1 and M2 money supplies] *and to most other interest rates as a forecaster of the economy"* [Bernanke and Blinder (1992, page 903)]. Finally, Figure 8 shows that the real economy (real GDP) affects the federal funds rate more than nonborrowed reserves.

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.

A related problem arises when measuring the degree of causality between two variables. Existing causality measures have been established only for horizon 1 and fail to capture indirect causal effects. In this paper, we have proposed a generalization of such measures for any horizon h, both parametric and nonparametric. Parametric measures are defined in terms of impulse response coefficients in the VMA representation. On observing that the relevant parametric expressions can be quite complex, an original approach based on simulating a large sample from the process of interest was suggested. Bootstrap nonparametric confidence intervals were also derived.

The causality measures are applied to examine whether monetary policy causes the real economy at different horizons. The results show that nonborrowed reserves cause the federal funds rate in the short-term, while the federal funds rate causes the real economy (real gross domestic product) for several months ahead. Further, the federal funds rate has more impact on the real economy than nonborrowed reserves. The federal funds rate is a better forecaster of the real economy than nonborrowed reserves. This also suggests that nonborrowed reserves affect the real economy mainly through their effect on interest rates.

The long VAR approach used in this paper to estimate the causality measures can have some drawbacks. Given the dimension of VAR and the large number of lags that we put in the model to capture the structure of the infinite autoregressive model, the estimated slope coefficients and the corresponding reduced-form impulse responses can be imprecise, especially in small sample. Consequently, a small sample bias in the causality measures may arises from bias in the reduced-form impulse coefficients estimates or from the nonlinearity of these measures. To remove this bias, one may replace the biased reduced-form impulse responses in the causality measures by bias-corrected reduced-form impulse responses estimates using bootstrap-after-bootstrap method discussed in Kilian (1998).

The choice of the appropriate order of VAR models can also be an important issue for causality measures. Recently, Kapetanios, Pagan and Scott (2007) argue that the order of VAR needed to reproduce the reduced-form impulse responses for actual economies is likely to be far higher than those suggested in practice and quite infeasible given the sample sizes in macroeconomics. The investigation of the impact of those issue on the small sample estimates of causality measures is the topic of on-going research.

Table 3. 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
$NBR \rightarrow R$	yes																			
$NBR \rightarrow P$	yes	yes	yes	yes	yes	yes														
$NBR \rightarrow GDP$	yes	yes	yes	yes																
$R \rightarrow NBR$																				
$R \rightarrow P$	yes	yes																		
$R \rightarrow GDP$	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes					
$P \rightarrow NBR$																				
$P \rightarrow R$																				
$P \rightarrow GDP$																				
$GDP \rightarrow NBR$																				
$GDP \rightarrow R$	yes	yes	yes	yes	yes															
$GDP \rightarrow P$																				

Note: This table summarizes the results of different causality directions between nonborrowed reserves (NBR), the federal funds rate (R), the gross domestic product deflator (P), and real gross domestic product (GDP), up to 20 months. "yes" means that the corresponding causality measure is significant at level 0.05. (*i.e.*, the 95% bootstrap confidence interval does not cover zero). The sample runs from January 1965 to December 1996 (384 observations).

A. Appendix: Proofs

PROOF OF PROPOSITION 4.1 Let h_1 and h_2 be two different horizons and $m_1 = m_2 = 1$. According to Diebold and Kilian (2001), the predictability measures for X based on the information sets $I_X(t)$ and $I_{XY}(t)$ are, respectively, defined as

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

By Definition 4.1, we then see that

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

□

PROOF OF PROPOSITION 6.1 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:

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

hence

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

Similarly, we have:

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

From (A.1) and (A.2), we get

$$\ln\left[\frac{\det\big\{J_0\tilde{\Sigma}_{0|k}(h)J_0'\big\}}{\det\big\{J_1\hat{\Sigma}_k(h)J_1'\big\}}\right] \underset{T\to\infty}{\overset{p}{\rightarrow}} \ln\left[\frac{\det\big\{J_0\Sigma_0(h)J_0'\big\}}{\det\big\{J_1\Sigma(h)J_1'\big\}}\right], \tag{A.3}$$

$$\hat{C}_L\big(Y \underset{h}{\longrightarrow} X\mid I\big) \underset{T\to\infty}{\overset{p}{\rightarrow}} C_L\big(Y \underset{h}{\longrightarrow} X\mid I\big)\,. \tag{A.4}$$

□

PROOF OF PROPOSITION 8.1 We know that, for $\delta \leq \frac{2}{3}$,

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

$$\ln\left(G\big(\tilde{\Phi}(k), \tilde{\Sigma}_{\varepsilon|\,k}\big)\right) = \ln\left(G\big(\Phi, \Sigma_u\big)\right) + o_p(T^{-\delta}) + o_p(1). \tag{A.5}$$

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

$$\ln\big(G\big(\hat{\Phi},\, \hat{\Sigma}_u\big)\big) = \ln\big(G\big(\Phi,\, \Sigma_u\big)\big) + o_p(1). \tag{A.6}$$

From (A.5) and (A.6), we get

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

Consequently,

$$\hat{C}_L(Y \underset{h}{\longrightarrow} X \,|\, I) = \tilde{C}_L(Y \underset{h}{\longrightarrow} X \,|\, I) + o_p(T^{-\delta}) + o_p(1)$$

where

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

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

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

where

$$D_C = \frac{\partial C_L(Y \underset{h}{\longrightarrow} X \mid I)}{\partial\,(\mathrm{vec}(\Phi)',\, \mathrm{vech}(\Sigma_u)')} = \frac{\partial C_L(Y \underset{h}{\longrightarrow} X \mid I)}{\partial \theta'} \tag{A.7}$$

for $\theta = \big(\mathrm{vec}(\Phi)',\, \mathrm{vech}(\Sigma_u)'\big)'$, hence

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

Using (8.5),

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

Consequently,

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

where $\sigma_c(h)^2 = D_C \Omega D_C^{'}$,

$$\Omega = \begin{bmatrix} \Gamma^{-1} \otimes \Sigma_u & 0 \\ 0 & 2(D_m^{'} D_m)^{-1} D_m^{'}(\Sigma_u \otimes \Sigma_u) D_m (D_m^{'} D_m)^{-1} \end{bmatrix}, \tag{A.8}$$

and 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 8.2 We start by showing that conditional on the sample

$$\mathrm{vec}\big(\hat{\Phi}^*\big) \underset{T\to\infty}{\overset{p}{\to}} \mathrm{vec}\big(\hat{\Phi}\big), \quad \mathrm{vech}\big(\hat{\Sigma}_u^*\big) \underset{T\to\infty}{\overset{p}{\to}} \mathrm{vech}\big(\hat{\Sigma}_u\big),$$

$$\operatorname{vec}\big(\tilde{\Phi}^*(k)\big) \underset{T\to\infty}{\overset{p}{\to}} \operatorname{vec}\big(\tilde{\Phi}(k)\big)\,, \quad \operatorname{vech}\big(\tilde{\Sigma}^*_{\varepsilon|\,k}\big) \underset{T\to\infty}{\overset{p}{\to}} \operatorname{vech}\big(\tilde{\Sigma}_{\varepsilon|\,k}\big).$$

We first note that

$$\begin{aligned}
\operatorname{vec}\big(\hat{\Phi}^*\big) &= \operatorname{vec}\big(\hat{\Gamma}_1^{*\prime}\hat{\Gamma}^{*-1}\big) = \operatorname{vec}\Big(\frac{1}{T-p}\sum_{t=p+1}^{T} W(t+1)^* w^*(t)'\hat{\Gamma}^{*-1}\Big) \\
&= \operatorname{vec}\Big(\frac{1}{T-p}\sum_{t=p+1}^{T}\big[\hat{\Phi}w^*(t) + u^*\,(t+1)\big]w^*(t)'\hat{\Gamma}^{*-1}\Big) \\
&= \operatorname{vec}\Big(\hat{\Phi}\big(\frac{1}{T-p}\sum_{t=p+1}^{T} w^*(t)w^*(t)'\big)\hat{\Gamma}^{*-1}\Big) + \operatorname{vec}\Big(\frac{1}{T-p}\sum_{t=p+1}^{T} u^*(t+1)w^*(t)'\hat{\Gamma}^{*-1}\Big) \\
&= \operatorname{vec}\big(\hat{\Phi}\big) + \operatorname{vec}\Big(\frac{1}{T-p}\sum_{t=p+1}^{T} u^*(t+1)w^*(t)'\hat{\Gamma}^{*-1}\Big)\,.
\end{aligned}$$

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

$$\mathsf{E}^*\big[u^*(t+1)w^*(t)'\hat{\Gamma}^{*-1}\big] = \mathsf{E}^*\big[\mathsf{E}^*\left[u^*(t+1)\,|\,\Im^*_t\right]w^*(t)'\hat{\Gamma}^{*-1}\big] = 0.$$

By the law of large numbers,

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

and $\operatorname{vec}\big(\hat{\Phi}^*\big) - \operatorname{vec}\big(\hat{\Phi}\big) \underset{T\to\infty}{\overset{p}{\to}} 0$. Now, to prove that $\operatorname{vech}\big(\hat{\Sigma}^*_u\big) \underset{T\to\infty}{\overset{p}{\to}} \operatorname{vech}\big(\hat{\Sigma}_u\big)$, we observe that

$$\operatorname{vech}\big(\hat{\Sigma}^*_u - \hat{\Sigma}_u\big) = \operatorname{vech}\Big\{\frac{1}{T-p}\sum_{t=p+1}^{T}\Big(u^*(t)u^*(t)' - \frac{1}{T-p}\sum_{s=p+1}^{T}\hat{u}(s)\hat{u}(s)'\Big)\Big\}.$$

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

$$\begin{aligned}
&\mathsf{E}^*\Big[u^*(s)u^*(s)' - \frac{1}{T-p}\sum_{s=p+1}^{T}\hat{u}(s)\hat{u}(s)'\Big] \\
&= \mathsf{E}^*\Big[\mathsf{E}^*\left[u^*(s)u^*(s)'\,|\,\Im^*_t\right] - \frac{1}{T-p}\sum_{s=p+1}^{T}\hat{u}(s)\hat{u}(s)'\Big] \\
&= \mathsf{E}^*\Big[\frac{1}{T-p}\sum_{s=p+1}^{T}\hat{u}(s)\hat{u}(s)' - \frac{1}{T-p}\sum_{s=p+1}^{T}\hat{u}(s)\hat{u}(s)'\Big] = 0.
\end{aligned}$$

Since

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

we get $\mathrm{vec}\big(\hat{\Sigma}_u^*\big) - \mathrm{vec}\big(\hat{\Sigma}_u\big) \underset{T\to\infty}{\overset{p}{\to}} 0$. Similarly, we can show that

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

Since $G(\cdot)$ and $H(\cdot)$ are 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

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

for $\delta \leq \frac{2}{3}$. Consequently,

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

where

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

We have shown that for $\delta \leq \frac{2}{3}$ [see the proof of Proposition 8.1],

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

Consequently,

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

Conditional on the sample, we can expand $\tilde{C}^*(Y \to X \mid I)$ around $\tilde{C}_L(Y \underset{h}{\longrightarrow} X \mid I)$:

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

hence

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

Conditional on the sample, we have:

$$T^{1/2}\begin{pmatrix} \mathrm{vec}(\hat{\Phi}^*) - \mathrm{vec}(\hat{\Phi}) \\ \mathrm{vech}(\hat{\Sigma}_u^*) - \mathrm{vech}(\hat{\Sigma}_u) \end{pmatrix} \overset{d}{\to} N[0, \ \Omega], \tag{A.9}$$

where Ω is given by equation (A.8); see Inoue and Kilian (2002). Thus,

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

where $\sigma_c(h)^2 = D_C \Omega D_C'$ and D_C is given by equation (A.7). □

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. and Blinder, A. (1992), 'The federal funds rate and the channels of monetary transmission', *American Economic Review* **LXXXII**, 901–21.

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., 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.

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.

Ireland, P. (2005), The monetary transmission mechanism, Technical report, Boston College and NBER, Boston, Massachusetts.

Kapetanios, G., Pagan, A. and Scott, A. (2007), 'Making a match: Combining theory and evidence in policy-oriented macroeconomic modeling', *Journal of Econometrics* **136**, 565–594.

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

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.

McCloskey, D. N. and Ziliak, S. T. (1996), 'The standard error of regressions', *The Journal of Economic Literature* **XXXIV**, 97–114.

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

Neyman, J. and Pearson, E. S. (1933), 'On the problem of the most efficient tests of statistical hypotheses', *Philosophical Transactions of the Royal Society A* **231**, 289–337.

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.