Short and Long Run Causality Measures: Theory and Inference*

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

First version: October 2004
Revised: November 2005, August 2006
This version: November 11, 2006
Compiled: November 11, 2006, 1:54pm

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

† Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 (514) 343 2400; FAX: 1 (514) 343 5831; e-mail: jean.marie.dufour@umontreal.ca . Web page: http://www.fas.umontreal.ca/SCECO/Dufour

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

The concept of causality introduced by Wiener (1956) and Granger (1969) is defined in terms of predictability one period ahead. Recently, Dufour and Renault (1998) generalized this concept by considering causality at a given horizon $h$, and causality up to any given horizon $h$. This generalization is motivated by the fact that, in the presence of an auxiliary variable vector $Z$, it is possible that the variable $Y$ does not cause variable $X$ at horizon 1, but causes it at horizon $h > 1$. In this case, there is an indirect causality transmitted by the auxiliary variable vector $Z$.

Another related problem consists in measuring the importance of causality between two variables. Existing causality measures have been defined only for the horizon 1 and fail to capture indirect causal effects. This paper proposes a generalization of such measures for any horizon $h$. We propose nonparametric and parametric measures for feedback and instantaneous effects at any horizon $h$. Parametric measures are defined in terms of impulse response coefficients of vector moving average representation ($VMA$). By analogy with Geweke (1982), we show it is always possible to define a measure of dependence at horizon $h$ which can be decomposed into a sum of causality measures from $X$ to $Y$, from $Y$ to $X$, and an instantaneous effect at horizon $h$. We propose a new approach to evaluate these measures based on the simulation of a large sample from the process of interest. We also propose a valid nonparametric confidence intervals, using the bootstrap technique. Finally, from an empirical application we find that nonborrowed reserves causes federal funds rate only at short-term, the effect of real gross domestic product on federal funds rate is significant for the first four horizons, the effect of federal funds rate on gross domestic product deflator is significant only at horizon 1, and finally federal funds rate causes the real gross domestic product until horizon 16.

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

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

1. Introduction 1
2. Motivation 2
3. Framework 4
4. Causality measures 6
5. Parametric causality measures 9
   5.1. Parametric causality measures in the context of a VARMA(p, q) process 9
   5.2. Characterization of causality measures for VMA(q) processes 16
6. Estimation 17
7. Evaluation by simulation of causality measures 21
8. Confidence intervals 24
9. Empirical illustration 29
10. Conclusion 33
A. Appendix: Proofs 38

List of Tables

1 Evaluation by simulation of causality at h=1, 2 23
2 Evaluation by simulation of causality at h=1, 2: Indirect causality 24
3 Dickey-Fuller tests: Variables in Logarithmic form 32
4 Dickey-Fuller tests: First difference 32
5 Stationarity test results 32
6 Summary of causality relations at various horizons for series in first difference 36
1. Introduction

The concept of causality introduced by Wiener (1956) and Granger (1969) is now a basic notion for studying dynamic relationships between time series. This concept is defined in terms of predictability at horizon one of a variable \( X \) from its own past, the past of another variable \( Y \), and possibly a vector \( Z \) of auxiliary variables. Following Granger (1969), we define causality from \( X \) to \( Y \) one period ahead as follows: \( Y \) causes \( X \) in the sense of Granger 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 precisely, we say that \( Y \) causes \( X \) in the sense of Granger 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 \( Y \).

The theory of causality has generated a considerable literature. In the context of bivariate ARMA models, Kang (1981), derived necessary and sufficient conditions for noncausality. Boudjellaba, Dufour and Roy (1992) and Boudjellaba, Dufour and Roy (1994) developed necessary and sufficient conditions of noncausality for multivariate ARMA models. Parallel to the literature on noncausality conditions, some authors developed tests for the presence of causality between time series. The first test is due to Sims (1972) in the context of bivariate time series. Other tests were developed for VAR models [see Pierce and Haugh (1977), Newbold (1982), Geweke (1984a)] and VARMA models [see Boudjellaba et al. (1992), and Boudjellaba et al. (1994)].

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 (1993b). Such generalization is motivated by the fact that, in the presence of auxiliary variables \( Z \), it is possible to have the variable \( Y \) not causing variable \( X \) at horizon one, but causing it at a longer horizon \( h > 1 \). In this case, we have an indirect causality transmitted by the auxiliary variables \( Z \). Necessary and sufficient conditions of noncausality between vectors of variables at any horizon \( h \) for stationary and nonstationary processes are also supplied.

The analysis of Wiener-Granger distinguishes among three types of causality: two unidirectional causalities (called feedbacks) from \( X \) to \( Y \) and from \( Y \) to \( X \) and an instantaneous causality associated with contemporaneous correlations. In practice, it is possible that these three types of causality coexist, hence the importance of finding means to measure their degree and determine the most important ones. Unfortunately, existing causality tests fail to accomplish this task, because they only inform us about the presence or the absence of causality. Geweke (1982, 1984b) extended the causality concept by defining measures of feedback and instantaneous effects, which can be decomposed in time and frequency domains. Gouriéroux, Monfort and Renault (1987) proposed causality measures based on the Kullback information. Polasek (1994) showed how causality measures can be calculated using the Akaike Information Criterion (AIC). Polasek (2000) 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 a one period horizon and fail to capture indirect causal effects. In this paper, we develop measures of causality at different horizons which can detect the well known indirect causality that appears at higher horizons. Specifically,
we propose generalizations to any horizon $h$ of the measures proposed by Geweke (1982) for the horizon one. Both nonparametric and parametric measures for feedback and instantaneous effects at any horizon $h$ are studied. Parametric measures are defined in terms of impulse response coefficients of the moving average (MA) representation of the process. By analogy with Geweke (1982, 1984b), we also define a measure of dependence at horizon $h$ which can be decomposed into the sum of feedback measures from $X$ to $Y$, from $Y$ to $X$, and an instantaneous effect at horizon $h$. To evaluate the measures associated with a given model – when analytical formulae are difficult to obtain – we propose a new approach based on a long simulation of the process of interest.

For empirical implementation, we propose consistent estimators as well as nonparametric confidence intervals, based on the bootstrap technique. The proposed causality measures can be applied in different contexts and may help to solve some puzzles of the economic and financial literature. They may improve the well known debate on long-term predictability of stock returns. In the present paper, they are applied to study causality relations 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, gross domestic product deflator, and real gross domestic product.

The plan of the paper is as follows. Section 2 provides the motivation behind an extension of causality measures at horizon $h > 1$. Section 3 presents the framework allowing the definition of causality at different horizons. In section 4, we propose nonparametric short-run and long-run causality measures. In section 5, we give a parametric equivalent, in the context of linear stationary invertible processes, of the causality measures suggested in section 4. Thereafter, we characterize our measures in the context of moving average models with finite order $q$. In section 6 we discuss different estimation approaches. In section 7 we suggest a new approach to estimate these measures based on the simulation. In section 8 we establish the asymptotic distribution of measures and the asymptotic validity of their nonparametric bootstrap confidence intervals. Section 9 is devoted to an empirical application and the conclusion relating to the results is given in section 10.

2. Motivation

The causality measures proposed in this paper constitute extensions of those developed by Geweke (1982) and Geweke (1984b) and others [see the introduction]. The existing causality measures quantify the effect of a vector of variables on another at 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 horizon $h$ strictly higher than 1 even if there is no causality at horizon 1. In this case, we speak about an indirect effect induced by the auxiliary variables $Z$. Clearly causality measures defined for horizon 1 are unable to quantify this indirect effect. This paper proposes causality measures at different horizons to quantify the degree of short and long run causality between vectors of random variables. Such causality measures detect and quantify the indirect effects due to auxiliary variables. To illustrate the importance of such causality measures, consider the following examples.

Example 2.1 Suppose we have information about two variables $X$ and $Y$. $(X, Y)'$ is a stationary
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 \( Z \). We consider a trivariate stationary process \((X,Y,Z)\) which follows a VAR(1) model:

\[
\begin{bmatrix}
  X(t + 1) \\
  Y(t + 1) \\
  Z(t + 1)
\end{bmatrix} =
\begin{bmatrix}
  0.60 & 0.00 & 0.80 \\
  0.00 & 0.40 & 0.00 \\
  0.00 & 0.60 & 0.10
\end{bmatrix}
\begin{bmatrix}
  X(t) \\
  Y(t) \\
  Z(t)
\end{bmatrix} +
\begin{bmatrix}
  \varepsilon_X(t + 1) \\
  \varepsilon_Y(t + 1) \\
  \varepsilon_Z(t + 1)
\end{bmatrix},
\]  

(4.2)

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

\[
\begin{bmatrix}
  X(t + 2) \\
  Y(t + 2) \\
  Z(t + 2)
\end{bmatrix} =
\begin{bmatrix}
  0.53 & 0.595 & 0.34 \\
  0.34 & 0.402 & 0.4 \\
  0.4 & 0.35 & 0.35
\end{bmatrix}
\begin{bmatrix}
  X(t) \\
  Y(t) \\
  Z(t)
\end{bmatrix} +
\begin{bmatrix}
  \varepsilon_X(t + 1) \\
  \varepsilon_Y(t + 1) \\
  \varepsilon_Z(t + 1)
\end{bmatrix}.
\]  

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

\[
X(t + 2) = 0.53 X(t) + 0.595 Y(t) + 0.5 \varepsilon_X(t + 1) + 0.7 \varepsilon_Y(t + 1) + \varepsilon_Z(t + 2).
\]  

(2.3)

The coefficient of \( Y(t) \) in equation (2.3) is equal to 0.595, so we can conclude that \( Y \) causes \( X \) at horizon 2. But, \textit{how can one measure the importance of this “long-run” causality? Existing measures do not answer this question.}

\textbf{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 \( Z \). We consider a trivariate stationary process \((X,Y,Z)\) which follows a VAR(1) model:

\[
\begin{bmatrix}
  X(t + 1) \\
  Y(t + 1) \\
  Z(t + 1)
\end{bmatrix} =
\begin{bmatrix}
  0.5 & 0.7 \\
  0.4 & 0.35 \\
  \varepsilon_X(t + 1) \\
  \varepsilon_Y(t + 1) \\
  \varepsilon_Z(t + 1)
\end{bmatrix}.
\]  

(2.2)
so that $X(t + 2)$ is given by

$$X(t + 2) = 0.36X(t) + 0.48Y(t) + 0.56Z(t) + 0.6\varepsilon X(t + 1) + 0.8\varepsilon Z(t + 1) + \varepsilon X(t + 2).$$

(2.8)

The coefficient of $Y(t)$ in equation (2.8) is equal to 0.48, which implies that $Y$ causes $X$ at horizon 2. This shows that the absence of causality at $h = 1$ does not exclude the possibility of a causality at horizon $h > 1$. This indirect effect is transmitted by the variable $Z$:

$$Y \rightarrow Z \rightarrow X,$$

$0.48 = 0.60 \times 0.80$, where 0.60 and 0.80 are the coefficients of the one period effect of $Y$ on $Z$ and the one period effect of $Z$ on $X$, respectively. So, how can one measure the importance of this indirect effect? Again, existing measures do not answer this question.

3. Framework

The notion of noncausality considered here is defined in terms of orthogonality conditions between subspaces of a Hilbert space of random variables with finite second moments. We denote $L^2 \equiv L^2(\mathcal{O}, \mathcal{A}, Q)$ the Hilbert space of real random variables defined on a common probability space $(\mathcal{O}, \mathcal{A}, Q)$, with covariance as inner product.

We consider three multivariate stochastic processes $\{X(t) : t \in \mathbb{Z}\}$, $\{Y(t) : t \in \mathbb{Z}\}$, and $\{Z(t) : t \in \mathbb{Z}\}$, with

$$X(t) = (x_1(t), \ldots, x_{m_1}(t))^\prime, \quad x_i(t) \in L^2, \quad i = 1, \ldots, m_1,$$

$$Y(t) = (y_1(t), \ldots, y_{m_2}(t))^\prime, \quad y_i(t) \in L^2, \quad i = 1, \ldots, m_2,$$

$$Z(t) = (z_1(t), \ldots, z_{m_3}(t))^\prime, \quad z_i(t) \in L^2, \quad i = 1, \ldots, m_3,$$

where $m_1 \geq 1$, $m_2 \geq 1$, $m_3 \geq 0$; and $m_1 + m_2 + m_3 = m$. We denote $X_t = \{X(s) : s \leq t\}$, $Y_t = \{Y(s) : s \leq t\}$ and $Z_t = \{Z(s) : s \leq t\}$ of $X$, $Y$ and $Z$, respectively. We denote $I_t$ the information set which contains $X_t$, $Y_t$ and $Z_t$. $I_t = A_t$, with $A_t = X_t, Y_t$ or $Z_t$, contains all the elements of $I_t$ except those of $A_t$. These information sets can be used to predict the value of $X$ at horizon $h$, denoted $X(t + h)$, for all $h \geq 1$.

For any information set $B_t$, let $P[x_i(t + h) \mid B_t]$ be the best linear forecast of $x_i(t + h)$ based on the information set $B_t$, the corresponding prediction error is

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

and $\sigma^2(x_i(t + h) \mid B_t)$ is the variance of this prediction error. Thus, the best linear forecast of $X(t + h)$ is

$$P(X(t + h) \mid B_t) = (P(x_1(t + h) \mid B_t), \ldots, P(x_{m_1}(t + h) \mid B_t))^\prime,$$
the corresponding vector of prediction error is

\[ U(X(t + h) \mid B_t) = (u(x_1(t + h) \mid B_t), \ldots, u(x_{m_1}(t + h) \mid B_t))', \]

and its covariance matrix is \( \Sigma(X(t + h) \mid B_t) \). Each component \( P[x_i(t + h) \mid B_t] \) of \( P[X(t + h) \mid B_t] \), for \( 1 \leq i \leq m_1 \), is then the orthogonal projection of \( x_i(t + h) \) on the subspace \( B_t \).

Following Dufour and Renault (1998), noncausality at horizon \( h \) and up to horizon \( h \), where \( h \) is a positive integer, are defined as follows.

**Definition 3.1** For \( h \geq 1 \),

(i) \( Y \) does not cause \( X \) at horizon \( h \) given \( I_t - Y_t \), denoted \( Y \not\rightarrow \bigwedge_{h} X \mid I_t - Y_t \) iff

\[
P[X(t + h) \mid I_t - Y_t] = P[X(t + h) \mid I_t], \ \forall t > w,
\]

where \( w \) represents a “starting point”;\(^1\)

(ii) \( Y \) does not cause \( X \) up to horizon \( h \) given \( I_t - Y_t \), denoted \( Y \not\rightarrow \bigwedge_{h} X \mid I_t - Y_t \) iff

\[
Y \not\rightarrow X \mid I_t - Y_t \ \text{for} \ k = 1, 2, \ldots, h;
\]

(iii) \( Y \) does not cause \( X \) at any horizon given \( I_t - Y_t \), denoted \( Y \not\rightarrow \bigwedge_{(\infty)} X \mid I_t - Y_t \) iff

\[
Y \not\rightarrow X \mid I_t - Y_t \ \text{for all} \ k = 1, 2, \ldots
\]

This definition corresponds to unidirectional 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 set \( I_t - Y_t \). An alternative definition can be expressed in terms of the covariance matrix of the forecast errors.

**Proposition 3.2** For \( h \geq 1 \),

(i) \( Y \) does not cause \( X \) at horizon \( h \) given \( I_t - Y_t \) iff

\[
\det \Sigma(X(t + h) \mid I_t - Y_t) = \det \Sigma(X(t + h) \mid I_t), \ \forall t > w;
\]

where \( \det \Sigma(X(t + h) \mid A_t) \), represents the determinant of the covariance matrix of the forecast error of \( X(t + h) \) given \( A_t = I_t \), \( I_t - Y_t \);

(ii) \( Y \) does not cause \( X \) up to horizon \( h \) given \( I_t - Y_t \) \( \forall \ t > w \) and \( k = 1, 2, \ldots, h \),

\[
\Sigma(X(t + k) \mid I_t - Y_t) = \Sigma(X(t + k) \mid I_t);
\]

---

\(^1\)The “starting point” \( w \) is not specified. In particular \( w \) may equal \(-\infty\) or 0 depending on whether we consider a stationary process on the integers \((t \in \mathbb{Z})\) or a process \( \{X(t) : t \geq 1\} \) on the positive integers given initial values preceding date 1.
(iii) $Y$ does not cause $X$ at any horizon given $I_t - Y_t$, if $\forall \ t > w$ and $k = 1, 2, \ldots$,

$$\Sigma(X(t + k) \mid I_t - Y_t) = \Sigma(X(t + k) \mid I_t).$$

4. Causality measures

In the remainder of this paper, we consider an information set $I_t$ which contains two vector valued random variables of interest $X$ and $Y$, and an auxiliary valued random variable $Z$. In other words, we suppose that $I_t = H \cup X_t \cup Y_t \cup Z_t$, where $H$ represents a subspace of the Hilbert space, possibly empty, containing time independent variables (e.g., the constant in a regression model).

The causality measures we consider are extensions of the measure introduced by Geweke (1982 and Geweke (1984b)). 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 $h \geq 1$.

**Definition 4.1** For $h \geq 1$, a causality measure from $Y$ to $X$ at horizon $h$, called the intensity of the causality from $Y$ to $X$ at horizon $h$ is given by

$$C(Y \to X \mid Z) = \ln \left[ \frac{\det \Sigma(X(t + h) \mid I_t - Y_t)}{\det \Sigma(X(t + h) \mid I_t)} \right].$$

**Remark 4.2** For $m_1 = m_2 = m_3 = 1$,

$$C(Y \to X \mid Z) = \ln \left[ \frac{\sigma^2(X(t + h) \mid I_t - Y_t - Z_t)}{\sigma^2(X(t + h) \mid I_t)} \right].$$

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

$$\frac{\sigma^2(X(t + h) \mid I_t - Y_t) - \sigma^2(X(t + h) \mid I_t)}{\sigma^2(X(t + h) \mid I_t)} = 1 - \exp[-C(Y \to X \mid Z)].$$

We can rewrite the conditional causality measures given by Definition 4.1 in terms of unconditional causality measures [see Geweke (1984b)]:

$$C(Y \to X \mid Z) = C(YZ \to X) - C(Z \to X)$$

where

$$C(YZ \to X) = \ln \left[ \frac{\det \Sigma(X(t + h) \mid I_t - Y_t - Z_t)}{\det \Sigma(X(t + h) \mid I_t)} \right].$$
\[
C(Z \to X) = \ln \left[ \frac{\det \Sigma(X(t+h) \mid I_t - Y_t - Z_t)}{\det \Sigma(X(t+h) \mid I_t - Y_t)} \right].
\]

\[C(YZ \to X)\] and \[C(Z \to X)\] represent the unconditional causality measures from \((Y', Z')'\) to \(X\) and from \(Z\) to \(X\), respectively. Similarly, we have:

\[C(X \to Y \mid Z) = C(XZ \to Y) - C(Z \to Y)\]

where

\[C(XZ \to Y) = \ln \left[ \frac{\det \Sigma(Y(t+h) \mid I_t - X_t - Z_t)}{\det \Sigma(Y(t+h) \mid I_t)} \right],\]
\[C(Z \to Y) = \ln \left[ \frac{\det \Sigma(Y(t+h) \mid I_t - X_t)}{\det \Sigma(Y(t+h) \mid I_t - X_t)} \right].\]

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

**Definition 4.3** For \(h \geq 1\), an instantaneous causality measure between \(Y\) and \(X\) at horizon \(h\), called the intensity of the instantaneous causality between \(Y\) and \(X\) at horizon \(h\), denoted \([C(X \to Y \mid Z)]\), is given by:

\[C(X \leftrightarrow Y \mid Z) = \ln \left[ \frac{\det \Sigma(X(t+h) \mid I_t) \det \Sigma(Y(t+h) \mid I_t)}{\det \Sigma(X(t+h), Y(t+h) \mid I_t)} \right].\]

where \(\det \Sigma(X(t+h), Y(t+h) \mid I_t)\) represents the determinant of the covariance matrix of the forecast error of the joint process \(\left(X', Y'\right)\) at horizon \(h\) given the information set \(I_t\).

**Remark 4.4** If we take \(m_1 = m_2 = m_3 = 1\), then

\[\det \Sigma((X(t+h), Y(t+h) \mid I_t) = \sigma^2(X(t+h) \mid I_t) \sigma^2(Y(t+h) \mid I_t) - \left(\text{cov}((X(t+h), Y(t+h) \mid I_t)\right)^2.\]  \hspace{1cm} (4.1)

So the instantaneous causality measure between \(X\) and \(Y\) at horizon \(h\) can be written as:

\[C(X \leftrightarrow Y \mid Z) = \ln \left[ \frac{1}{1 - \rho^2(X(t+h), Y(t+h) \mid I_t)} \right]\]

where

\[\rho(X(t+h), Y(t+h) \mid I_t) = \frac{\text{cov}(X(t+h), Y(t+h) \mid I_t)}{\sigma(X(t+h) \mid I_t)\sigma(Y(t+h) \mid I_t)}.\]  \hspace{1cm} (4.2)

is the correlation coefficient between \(X(t+h)\) and \(Y(t+h)\) given the information set \(I_t\). This instantaneous causality measure is higher when this coefficient becomes higher.

We can also define a measure of dependence between \(X\) and \(Y\) at horizon \(h\). This will enable
us to check if, at horizon $h$, the processes $X$ and $Y$ must be considered together or whether they can be treated separately given the information set $I_t - Y_t$.

**Definition 4.5** For $h \geq 1$, a measure of dependence between $X$ and $Y$ at horizon $h$, called the intensity of the dependence between $X$ and $Y$ at horizon $h$, denoted $C^{(h)}(X, Y \mid Z)$, is given by:

$$C^{(h)}(X, Y \mid Z) = \ln \left[ \frac{\det \Sigma (X(t + h) \mid I_t - Y_t) \det \Sigma (Y(t + h) \mid I_t - X_t)}{\det \Sigma (X(t + h), Y(t + h) \mid I_t)} \right].$$

We can easily show that the intensity of the dependence between $X$ and $Y$ at horizon $h$ is equal to the sum of feedbacks measures from $X$ to $Y$, from $Y$ to $X$, and the instantaneous measure at horizon $h$. We have:

$$C^{(h)}(X, Y \mid Z) = C(X \rightarrow h Y \mid Z) + C(Y \rightarrow h X \mid Z) + C(X \leftrightarrow h Y \mid Z). \quad (4.3)$$

It is possible to build a recursive formulation of causality measures. This one will depend on the predictability measure introduced by Diebold and Kilian (1998).

These authors proposed a predictability measure based on the ratio of expected losses of short and long run forecasts:

$$\tilde{P}(L, \Omega_t, j, k) = 1 - \frac{E(L(e_{t+j, t}))}{E(L(e_{t+k, t}))}$$

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

$$L(e_{t+s, t}) = e^2_{t+s, t}, \text{ for } s = j, k.$$

We have the following relationships.

**Proposition 4.6** Let $h_1, h_2$ be two different horizons. For $h_2 > h_1 \geq 1$ and $m_1 = m_2 = 1$,

$$C(Y \rightarrow h_1 X \mid Z) - C(Y \rightarrow h_2 X \mid Z) = \ln \left[ 1 - \tilde{P}_X(I_t - Y_t, h_1, h_2) \right] - \ln \left[ 1 - \tilde{P}_X(I_t, h_1, h_2) \right]$$

where $\tilde{P}_X(\cdot, h_1, h_2)$ represents the predictability measure for variable $X$,

$$\tilde{P}_X(I_t - Y_t, h_1, h_2) = 1 - \frac{\det \Sigma (X(t + h_1) \mid I_t - Y_t)}{\det \Sigma (X(t + h_2) \mid I_t - Y_t)}$$

$$\tilde{P}_X(I_t, h_1, h_2) = 1 - \frac{\det \Sigma (X(t + h_1) \mid I_t)}{\det \Sigma (X(t + h_2) \mid I_t)}.$$

Proofs appear in Appendix A. The following corollary follows immediately from the latter proposition.
Corollary 4.7 For $h \geq 2$ and $m_1 = m_2 = 1$,

\[ C(Y \rightarrow_h X | Z) = C(Y \rightarrow_1 X | Z) + \ln[1 - \bar{P}_X(I_t, 1, h)] - \ln[1 - \bar{P}_X(I_t - Y_t, 1, h)]. \]

For $h_2 \gg h_1$, the function $\bar{P}_k(\ldots, h_1, h_2)$, $k = X, Y$, represents the measure of short-run forecast relative to the long-run forecast, and $C(k \rightarrow_l l | Z) - C(k \rightarrow_l l | Z)$, for $l \neq k$ and $l, k = X, Y$, represents the difference between the degree of short run causality and that of long run causality. Further, $\bar{P}_k(\ldots, h_1, h_2) \gg 0$ means that the series is highly predictable at horizon $h_1$ relative to $h_2$, and $\bar{P}_k(\ldots, h_1, h_2) = 0$, means that the series is nearly unpredictable at horizon $h_1$ relative to $h_2$.

5. Parametric causality measures

We now consider a more specific set of linear invertible processes which includes VAR, VMA, and VARMA models of finite order as special cases. Under this set we show that it is possible to obtain parametric expressions for short-run and long-run causality measures in terms of impulse response coefficients of a VMA representation.

This section is divided into two subsections. In the first we calculate parametric measures of short-run and long-run causality in the context of an autoregressive moving average model. We assume that the process \{W(s) = (X'(s), Y'(s), Z'(s))' : s ≤ t\} is a VARMA(p,q) model (hereafter unconstrained model), where $p$ and $q$ can be infinite. The model of the process \{S(s) = (X'(s), Z'(s))' : s ≤ t\} (hereafter constrained model) can be deduced from the unconstrained model using Corollary 6.1.1 in Lütkepohl (1993b). This model follows a VARMA($\tilde{p}$, $\tilde{q}$) model with $\tilde{p} \leq mp$ and $\tilde{q} \leq (m - 1)p + q$. In the second subsection we provide a characterization of those parametric measures in the context of VMA(q) model, where $q$ is finite.

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

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

\[
W(t) = \sum_{j=1}^{p} \pi_j W(t - j) + \sum_{j=1}^{q} \varphi_j u(t - j) + u(t)
\]

\[
= \sum_{j=1}^{p} \begin{bmatrix} \pi_{Xj} & \pi_{Yj} & \pi_{Zj} \\ \pi_{Yj} & \pi_{Yj} & \pi_{Zj} \\ \pi_{Zj} & \pi_{Zj} & \pi_{Zj} \end{bmatrix} \begin{bmatrix} X(t - j) \\ Y(t - j) \\ Z(t - j) \end{bmatrix} + \sum_{j=1}^{q} \begin{bmatrix} \varphi_{Xj} & \varphi_{Yj} & \varphi_{Zj} \\ \varphi_{Yj} & \varphi_{Yj} & \varphi_{Zj} \\ \varphi_{Zj} & \varphi_{Zj} & \varphi_{Zj} \end{bmatrix} \begin{bmatrix} u_X(t - j) \\ u_Y(t - j) \\ u_Z(t - j) \end{bmatrix} + \begin{bmatrix} u_X(t) \\ u_Y(t) \\ u_Z(t) \end{bmatrix}
\]

(5.1)
with
\[ E[u(t)] = 0, \quad E[u(t)u'(s)] = \begin{cases} \Sigma_u & \text{for } s = t \\ 0 & \text{for } s \neq t \end{cases}. \]
or, more compactly,
\[ \Pi(L)W(t) = \varphi(L)u(t) \]
where
\[ \Pi(L) = \begin{bmatrix} \pi_{XX}(L) & \pi_{XY}(L) & \pi_{XZ}(L) \\ \pi_{YX}(L) & \pi_{YY}(L) & \pi_{YZ}(L) \\ \pi_{ZX}(L) & \pi_{ZY}(L) & \pi_{ZZ}(L) \end{bmatrix}, \]
\[ \varphi(L) = \begin{bmatrix} \varphi_{XX}(L) & \varphi_{XY}(L) & \varphi_{XZ}(L) \\ \varphi_{YX}(L) & \varphi_{YY}(L) & \varphi_{YZ}(L) \\ \varphi_{ZX}(L) & \varphi_{ZY}(L) & \varphi_{ZZ}(L) \end{bmatrix}, \]
\[ \pi_{ii}(L) = I_{m_i} - \sum_{j=1}^{p} \pi_{ij}L^j, \quad \pi_{ik}(L) = - \sum_{j=1}^{p} \pi_{ikj}L^j, \]
\[ \varphi_{ii}(L) = I_{m_i} + \sum_{j=1}^{q} \varphi_{ij}L^j, \quad \varphi_{ik}(L) = \sum_{j=1}^{q} \varphi_{ikj}L^j, \text{ for } i \neq k, \ i, k = X, Y, Z. \]

We assume that \( u(t) \) is orthogonal to the Hilbert subspace \( \{W(s), \ s \leq (t - 1)\} \) and that \( \Sigma_u \) is a symmetric positive definite matrix. Under stationarity, \( W(t) \) has a VMA(\( \infty \)) representation,
\[ W(t) = \Psi(L)u(t) \quad (5.2) \]
where
\[ \Psi(L) = \Pi(L)^{-1}\varphi(L) = \sum_{j=0}^{\infty} \psi_jL^j = \sum_{j=0}^{\infty} \begin{bmatrix} \psi_{XXj} & \psi_{XYj} & \psi_{XZj} \\ \psi_{YXj} & \psi_{YYj} & \psi_{YZj} \\ \psi_{ZXj} & \psi_{ZYj} & \psi_{ZZj} \end{bmatrix} L^j, \quad \psi_0 = I_{m}. \]

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

**Proposition 5.1** (Linear transformation of a VMA(q) process). Let \( u(t) \) be a \( K \)-dimensional white noise process with nonsingular variance-covariance matrix \( \Sigma_u \) and let
\[ W(t) = \mu + \sum_{j=1}^{q} \Psi_ju(t - j) + u(t), \]
be a $K$-dimensional invertible VMA($q$) process. Furthermore, let $F$ be an $(M \times K)$ matrix of rank $M$. Then the $M$-dimensional process $S(t) = FW(t)$ has an invertible VMA($\bar{q}$) representation:

$$S(t) = F\mu + \sum_{j=1}^{\bar{q}} \theta_j \varepsilon(t - j) + \varepsilon(t),$$

where $\varepsilon(t)$ is $M$-dimensional white noise with nonsingular variance-covariance matrix $\Sigma_{\varepsilon}$, the $\theta_j$, $j = 1, \ldots, \bar{q}$, are $(M \times M)$ coefficient matrices and $\bar{q} \leq q$.

**Corollary 5.2** (Linear Transformation of a VARMA($p,q$) process). Let $W(t)$ be a $K$-dimensional, stable, invertible VARMA($p,q$) process and let $F$ be an $(M \times K)$ matrix of rank $M$. Then the process $S(t) = FW(t)$ has a VARMA($\bar{p}, \bar{q}$) representation with $\bar{p} \leq Kp$, $\bar{q} \leq (K-1)p + q$.

**Remark 5.3** If we assume that $W(t)$ follows a VAR($p$) = VARMA($p,0$) model, then its linear transformation $S(t) = FW(t)$ has a VARMA($\bar{p}, \bar{q}$) representation with $\bar{p} \leq Kp$ and $\bar{q} \leq (K-1)p$.

Assume that we are interested in measuring the causality from $Y$ to $X$ at a given horizon $h$. We need to apply Corollary (5.2) to define the structure of process $S(s) = (X(s)', Z(s)')'$. So, if we left-multiply equation (5.2) by the adjoint matrix of matrix $\Pi(L)$, denoted $\Pi(L)^*$, we get

$$\Pi(L)^* \Pi(L) W(t) = \Pi(L)^* \varphi(L) u(t)$$

(5.3)

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

$$\det [\Pi(L)] = 1 - \alpha_1 L - \cdots - \alpha_p L^p$$

where $\bar{p} \leq mp$. It is also easy to check that the degree of the operator $\Pi(L)^* \varphi(L)$ is at most $p(m-1) + q$. So equation (5.3) can be written as follows:

$$\det [\Pi(L)] W(t) = \Pi(L)^* \varphi(L) u(t).$$

(5.4)

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

$$F = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & 0 & I_{m_3} \end{bmatrix}.$$

On premultiplying (5.4) by $F$, we get:

$$\det [\Pi(L)] S(t) = F\Pi(L)^* \varphi(L) u(t).$$

(5.5)
The right-hand side of equation (5.5) is a linearly transformed finite order VMA process which, by Proposition 5.1, has a VMA($\bar{q}$) representation with $\bar{q} \leq p(m - 1) + q$. Thus, we have the following constrained model:

$$\text{det} \left[ \Pi(L) \right] S(t) = \theta(L) \varepsilon(t) = \left[ \begin{array}{cc} \theta_{XX}(L) & \theta_{XZ}(L) \\ \theta_{ZX}(L) & \theta_{ZZ}(L) \end{array} \right] \varepsilon(t)$$

(5.6)

where

$$E[\varepsilon(t)] = 0, \quad E[\varepsilon(t)\varepsilon'(s)] = \left\{ \begin{array}{ll} \Sigma_\varepsilon & \text{for } s = t \\ 0 & \text{for } s \neq t \end{array} \right.$$ 

$$\theta_{ii}(L) = I_m + \sum_{j=1}^{\bar{q}} \theta_{ij} L^j, \quad \theta_{ik}(L) = \sum_{j=1}^{\bar{q}} \theta_{kj} L^j, \quad \text{for } i \neq k, \ i, k = X, Z.$$ 

Note that, in theory, the coefficients $\theta_{ikj}$, $i, k = X, Z$; $j = 1, \ldots, \bar{q}$, and elements of the variance-covariance matrix $\Sigma_\varepsilon$, can be computed from coefficients $\pi_{ik}$, $i, k = X, Z, Y$; $j = 1, \ldots, p$; $l = 1, \ldots, q$, and elements of the variance-covariance matrix $\Sigma_u$. This is possible by solving the following system:

$$\gamma_\varepsilon(v) = \gamma_u(v), \ v = 0, 1, 2, \ldots$$

(5.7)

where $\gamma_\varepsilon(v)$ and $\gamma_u(v)$ are the autocovariance functions of the processes $\theta(L)\varepsilon(t)$ and $F\Pi(L)^{*}\varphi(L)u(t)$, respectively. For large numbers $m$, $p$, and $q$, system (5.7) can be solved by using optimization methods. In section 7 we discuss another approach to estimating the constrained model using the unconstrained model. The following example shows how one can calculate the theoretical parameters of the constrained model in terms of those of the unconstrained model in the context of a bivariate VAR(1) model.

**Example 5.4** Consider the following bivariate VAR(1) model:

$$\begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} = \begin{bmatrix} \pi_{XX} & \pi_{XY} \\ \pi_{YX} & \pi_{YY} \end{bmatrix} \begin{bmatrix} X(t-1) \\ Y(t-1) \end{bmatrix} + \begin{bmatrix} u_X(t) \\ u_Y(t) \end{bmatrix}$$

$$= \pi \begin{bmatrix} X(t-1) \\ Y(t-1) \end{bmatrix} + u(t).$$

(5.8)

We assume that all roots of $\text{det}[[\Pi(z)] = \text{det}(I_2 - \pi z)$ are outside of the unit circle. Under this assumption model (5.8) has the following VMA$(\infty)$ representation:

$$\begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} = \sum_{j=0}^{\infty} \psi_j \begin{bmatrix} u_X(t-j) \\ u_Y(t-j) \end{bmatrix} = \sum_{j=0}^{\infty} \begin{bmatrix} \psi_{XX,j} & \psi_{XY,j} \\ \psi_{YX,j} & \psi_{YY,j} \end{bmatrix} \begin{bmatrix} u_X(t-j) \\ u_Y(t-j) \end{bmatrix}$$

where

$$\psi_j = \pi^{j}, \quad j = 1, 2, \ldots, \psi_0 = I_2.$$

If we are interested in determining the model of marginal process $X(t)$, then by Corollary (5.2) and
for $F = [1, 0]$, we have,
\[
\det[\Pi(L)]X(t) = [1, 0] \Pi(L)^* u(t)
\]
where
\[
\Pi(L)^* = \begin{bmatrix}
1 - \pi_{YY} L & \pi_{XY} L \\
\pi_{YX} L & 1 - \pi_{XX} L
\end{bmatrix},
\]
\[
\det[\Pi(L)] = 1 - (\pi_{YY} + \pi_{XX})L - (\pi_{YX}\pi_{XY} - \pi_{XX}\pi_{YY})L^2.
\]
Thus,
\[
X(t) - (\pi_{YY} + \pi_{XX})X(t - 1) - (\pi_{YX}\pi_{XY} - \pi_{XX}\pi_{YY})X(t - 2)
= \pi_{XY}u_Y(t - 1) - \pi_{YY}u_X(t - 1) + u_X(t).
\]

The right-hand side, denoted $\varpi(t)$, is the sum of an MA(1) process and a white noise process. By Proposition 5.1, $\varpi(t)$ has an MA(1) representation, $\varpi(t) = \varepsilon_X(t) + \theta \varepsilon_X(t - 1)$. To determine parameters $\theta$ and $\text{Var}(\varepsilon_X(t)) = \sigma_{\varepsilon_X}^2$ in terms of the parameters of the unconstrained model, we have to solve system (5.7) for $v = 0$ and $v = 1$,
\[
\text{Var}[\varepsilon_X(t) + \theta \varepsilon_X(t - 1)] = \text{Var}[u_X(t) - \pi_{YY}u_X(t - 1) + \pi_{XY}u_Y(t - 1)],
\]
\[
E[(\varepsilon_X(t) + \theta \varepsilon_X(t - 1))(\varepsilon_X(t - 1) + \theta \varepsilon_X(t - 2))] = E[(u_X(t) - \pi_{YY}u_X(t - 1) + \pi_{XY}u_Y(t - 1))]
\times (u_X(t - 1) - \pi_{YY}u_X(t - 2) + \pi_{XY}u_Y(t - 2)),
\]
\[
\equiv (1 + \theta^2)\sigma_{\varepsilon_X}^2 = (1 + \pi_{YY}^2)\sigma_{u_X}^2 + \pi_{XY}^2\sigma_{u_Y}^2 - 2\pi_{YY}\pi_{XY}\sigma_{u_Y}u_X,
\]
\[
\theta \sigma_{\varepsilon_X}^2 = -\pi_{YY}\sigma_{u_X}^2.
\]

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

The VMA($\infty$) representation of model (5.6) is given by:
\[
S(t) = \text{det} [\Pi(L)]^{-1} \theta(L)\varepsilon(t) = \sum_{j=0}^{\infty} \Phi_j \varepsilon(t - j)
= \sum_{j=0}^{\infty} \begin{bmatrix}
\phi_{XX,j} & \phi_{XZ,j} \\
\phi_{ZX,j} & \phi_{ZZ,j}
\end{bmatrix}
\begin{bmatrix}
\varepsilon_X(t - j) \\
\varepsilon_Z(t - j)
\end{bmatrix},
\]

(5.9)

where $\Phi_0 = I_{(m_1 + m_2)}$. To quantify the degree of causality from $Y$ to $X$ at horizon $h$, we first consider the unconstrained and constrained models of process $X$. The unconstrained model is given
by the following equation:

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

whereas the constrained model is given by:

\[
X(t) = \sum_{j=1}^{\infty} \phi_{XXj} \varepsilon_X(t - j) + \sum_{j=1}^{\infty} \phi_{XZj} \varepsilon_Z(t - j) + \varepsilon_X(t).
\]

Second, we need to calculate the variance-covariance matrices of the unconstrained and constrained forecast errors of \(X(t + h)\). From equation (5.2), the forecast error of \(W(t + h)\) is given by:

\[
e_{nc}[W(t + h) \mid I_t] = \sum_{i=0}^{h-1} \psi_i u(t + h - i),
\]

associated with the variance-covariance matrix

\[
\Sigma(W(t + h) \mid I_t) = \sum_{i=0}^{h-1} \psi_i \text{Var}[u(t)] \psi_i^\prime = \sum_{i=0}^{h-1} \psi_i \Sigma_u \psi_i^\prime. \quad (5.10)
\]

Therefore, the unconstrained forecast error of \(X(t + h)\) is given by

\[
e_{nc}(X(t + h) \mid I_t) = \sum_{j=1}^{h-1} \psi_{XXj} u_X(t + h - j) + \sum_{j=1}^{h-1} \psi_{XYj} u_Y(t + h - j) + \sum_{j=1}^{h-1} \psi_{XZj} u_Z(t + h - j) + u_X(t),
\]

which is associated with the unconstrained variance-covariance matrix

\[
\Sigma(X(t + h) \mid I_t) = \sum_{i=0}^{h-1} [e_{nc}^{i} \Sigma_u \psi_i^\prime e_{nc}^{i^\prime}],
\]

where \(e_{nc}^{X} = \begin{bmatrix} I_m & 0 & 0 \end{bmatrix}\). Similarly, the forecast error of \(S(t + h)\) is given by

\[
e_c[S(t + h) \mid I_t - Y_t] = \sum_{i=0}^{h-1} \Phi_i \varepsilon_X(t + h - i)
\]
associated with the variance-covariance matrix

\[
\Sigma(S(t + h) | I_t - Y_t) = \sum_{i=0}^{h-1} \Phi_i \Sigma e \Phi_i'.
\]

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

\[
e_c(X(t + h) | I_t - Y_t) = \sum_{j=1}^{h-1} \phi_{Xj} \varepsilon_X(t + h - j) + \sum_{j=1}^{h-1} \phi_{XZj} \varepsilon_Z(t + h - j) + \varepsilon_X(t + h)
\]

associated with the constrained variance-covariance matrix

\[
\Sigma(X(t + h) | I_t - Y_t) = \sum_{i=0}^{h-1} e_c(X) \Phi_i \Sigma e \Phi_i' e_c(X)
\]

where \(e_c(X) = \begin{bmatrix} I_{m_1} & 0 \end{bmatrix}\). Thus, we can immediately deduce the following result by using the definition of a causality measure from \(Y\) to \(X\) [see Definition (4.1)].

**Theorem 5.5** Under assumptions (5.1) and (5.2); and for \(h \geq 1\), where \(h\) is a positive integer,

\[
C(Y \rightarrow h X | Z) = \ln \left[ \frac{\det(\sum_{i=0}^{h-1} e_c(X) \Phi_i \Sigma e \Phi_i' e_c(X))}{\det(\sum_{i=0}^{h-1} e_c(X) \Phi_i \Sigma e \Phi_i' e_c(X))} \right],
\]

where \(e_c(X) = \begin{bmatrix} I_{m_1} & 0 \end{bmatrix}\), \(e_c(X) = \begin{bmatrix} I_{m_2} & 0 \end{bmatrix}\).

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

**Example 5.6** If we consider a bivariate VAR(1) process [see Example 5.4], we can calculate causality measures at any horizon \(h\) using only the unconstrained parameters. For example, the causality measures at horizons 1 and 2 are given by:

\[
C(Y \rightarrow 1 X) = \ln \left[ \frac{(1 + \pi_Y^2)\sigma_{uX}^2 + \pi_Y^2 \sigma_{uY}^2 + \sqrt{(1 + \pi_Y^2)\sigma_{uX}^2 + \pi_Y^2 \sigma_{uY}^2)^2 - 4\pi_Y^2 \sigma_{uX}^4}}{2\sigma_{uX}^2} \right],
\]

\[
C(Y \rightarrow 2 X) = \ln \left[ \frac{4\pi_Y^2 \sigma_{uX}^4 + (1 + \pi_Y^2)\sigma_{uX}^2 + \pi_Y^2 \sigma_{uY}^2 - \sqrt{((1 + \pi_Y^2)\sigma_{uX}^2 + \pi_Y^2 \sigma_{uY}^2)^2 - 4\pi_Y^2 \sigma_{uX}^4} - 2\pi_Y^2 \sigma_{uX}^2}{2((1 + \pi_Y^2)\sigma_{uX}^2 + \pi_Y^2 \sigma_{uY}^2)[(1 + \pi_Y^2)\sigma_{uX}^2 + \pi_Y^2 \sigma_{uY}^2 - \sqrt{((1 + \pi_Y^2)\sigma_{uX}^2 + \pi_Y^2 \sigma_{uY}^2)^2 - 4\pi_Y^2 \sigma_{uX}^4}]^2}} \right].
\]

The last two formulas are obtained under assumptions \(cov(u_X(t), u_Y(t)) = 0\) and

\[
((1 + \pi_Y^2)\sigma_{uX}^2 + \pi_Y^2 \sigma_{uY}^2)^2 - 4\pi_Y^2 \sigma_{uX}^4 \geq 0.
\]
Now we will determine the parametric measure of instantaneous causality 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. In this case, the variance-covariance matrix of the unconstrained forecast error of joint process \((X'(t+h), Y'(t+h))'\) is given by:

\[
\Sigma(X(t+h), Y(t+h) \mid I_t) = \sum_{i=0}^{h-1} G \psi_i \Sigma u \psi_i' G',
\]

where \( G = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & I_{m_2} & 0 \end{bmatrix} \). Thus, we have,

\[
\Sigma(X(t+h) \mid I_t) = \sum_{i=0}^{h-1} \begin{bmatrix} e^{nc} \psi_i \Sigma u \psi_i' e^{nc} \end{bmatrix},
\]

\[
\Sigma(Y(t+h) \mid I_t) = \sum_{i=0}^{h-1} \begin{bmatrix} e^{nc} \psi_i \Sigma u \psi_i' e^{nc} \end{bmatrix},
\]

where \( e^{nc} = \begin{bmatrix} 0 & I_{m_2} & 0 \end{bmatrix} \). Thus, we can immediately deduce the following result by using the definition of the instantaneous causality measure.

**Theorem 5.7** Under assumptions (5.1) and (5.2) and for \( h \geq 1 \),

\[
C(X_{h} \rightarrow Y \mid Z) = \ln \left[ \frac{\det(\sum_{i=0}^{h-1} e^{nc} \psi_i \Sigma u \psi_i' e^{nc}) \det(\sum_{i=0}^{h-1} e^{nc} \psi_i \Sigma u \psi_i' e^{nc})}{\det(\sum_{i=0}^{h-1} G \psi_i \Sigma u \psi_i' G')} \right]
\]

where \( G = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & I_{m_2} & 0 \end{bmatrix} \), \( e_X = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \), and \( e_Y = \begin{bmatrix} 0 & I_{m_2} & 0 \end{bmatrix} \).

The parametric measure of dependence at horizon \( h \) can be deduced from its decomposition given by equation (4.3).

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

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

\[
W(t) = \sum_{j=1}^{q} \Phi_j u(t-j) + u(t) = \sum_{j=1}^{q} \begin{bmatrix} \Phi_{Xj} & \Phi_{Yj} & \Phi_{Zj} \\ \Phi_{Xj} & \Phi_{Yj} & \Phi_{Zj} \\ \Phi_{Xj} & \Phi_{Yj} & \Phi_{Zj} \end{bmatrix} \begin{bmatrix} u_X(t-j) \\ u_Y(t-j) \\ u_Z(t-j) \end{bmatrix} + \begin{bmatrix} u_X(t) \\ u_Y(t) \\ u_Z(t) \end{bmatrix}.
\] (5.11)

16
More compactly,

\[ W(t) = \Phi(L)u(t) \]

where

\[
\Phi(L) = \begin{bmatrix}
\Phi_{XX}(L) & \Phi_{XY}(L) & \Phi_{XZ}(L) \\
\Phi_{YX}(L) & \Phi_{YY}(L) & \Phi_{YZ}(L) \\
\Phi_{ZX}(L) & \Phi_{ZY}(L) & \Phi_{ZZ}(L)
\end{bmatrix},
\]

\[
\Phi_{ii}(L) = I_{m_i} + \sum_{j=1}^{q} \Phi_{ij} j L^j, \Phi_{ik}(L) = \sum_{j=1}^{q} \Phi_{ik} j L^j, \quad \text{for } i \neq k, \ i, k = X, Z, Y.
\]

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

\[
S(t) = \theta(L)\varepsilon(t) = \sum_{j=0}^{q} \theta_j \varepsilon(t-j) = \sum_{j=0}^{q} \begin{bmatrix} \theta_{XX,j} & \theta_{XZ,j} \\ \theta_{ZX,j} & \theta_{ZZ,j} \end{bmatrix} \begin{bmatrix} \varepsilon_X(t-j) \\ \varepsilon_Z(t-j) \end{bmatrix}
\]

where

\[
E[\varepsilon(t)] = 0, \quad E[\varepsilon(t)\varepsilon'(s)] = \begin{cases} \Sigma_{\varepsilon} & \text{for } s = t \\ 0 & \text{for } s \neq t \end{cases}.
\]

We have the following result.

**Theorem 5.8** Let \( h_1 \) and \( h_2 \) be two different horizons. Under assumption (5.11) we have,

\[
C(Y \rightarrow X \mid Z) = C(Y \rightarrow X \mid Z), \quad \forall \ h_2 \geq h_1 \geq q.
\]

This result follows immediately from Proposition 4.6 and from the fact that for all \( h_2 \geq h_1 \geq q \) and \( m_1 = m_2 = 1 \),

\[
P_X(I_t - Y_t, h_1, h_2) = P_X(I_t, h_1, h_2).
\]

### 6. Estimation

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

We here illustrate three different approaches for estimating causality measures. The first, called the nonparametric approach, is the focus of this section. It assumes that the form of the parametric model appropriate for the process of interest is unknown and approximates it with a VAR(\( k \)) model, where \( k \) depends on the sample size [see Parzen (1974), Bhansali (1978), Lewis and Reinsel (1985)]. The second approach assumes that the process follows a finite order VARMA model. The standard methods for the estimation of VARMA models, such as maximum likelihood and nonlinear least squares, require nonlinear optimization. This might not be feasible because the number
of parameters can increase quickly. To circumvent this problem, several authors [see Hannan and Rissanen (1982), Hannan and Kavalieris (1984), Koreisha and Pukkila (1989), Dufour and Pelletier (2005), and Dufour and Jouini (2004)] have developed a relatively simple approach based only on linear regression. This approach enables estimation of VARMA models using a long VAR whose order depends on the sample size. The last and simplest approach assumes that the process follows a finite order VAR(p) model which can be estimated by OLS.

In practice, the precise form of the parametric model appropriate for a process is unknown. Parzen (1974), Bhansali (1978), Lewis and Reinsel (1985), and others have considered a nonparametric approach to predicting future values using an autoregressive model fitted to a series of $T$ observations. This approach is based on a very mild assumption of an infinite order autoregressive model for the process which includes finite-order stationary VARMA processes as a special case.

In this section, we describe the nonparametric approach to estimating the short- and long-run causality measures. First, we discuss estimation of the fitted autoregressive constrained and unconstrained models. We then point out some assumptions necessary for the convergence of the estimated parameters. Second, using Theorem 6 in Lewis and Reinsel (1985), we define approximations of variance-covariance matrices of the constrained and unconstrained forecast errors at horizon $h$. Last, we use these approximations to construct an asymptotic estimator of our short- and long-run causality measures.

In what follows we focus on the estimation of the unconstrained model. Let us consider a stationary vector process $\{W(s) = (X(s)', Y(s)', Z(s)', s \leq t)\}$. By Wold’s theorem, this process can be written in the form of a VMA($\infty$) model:

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

We assume that $\sum_{j=0}^{\infty} \|\varphi_j\| < \infty$ and $\det\{\varphi(z)\} \neq 0$ for $|z| \leq 1$, where $\|\varphi_j\| = tr(\varphi_j' \varphi_j)$ and $\varphi(z) = \sum_{j=0}^{\infty} \varphi_j z^j$, with $\varphi_0 = I_m$, an $(m \times m)$ identity matrix. Under these assumptions, $W(t)$ is invertible and can be written as an infinite autoregressive process:

$$W(t) = \sum_{j=1}^{\infty} \pi_j W(t - j) + u(t), \quad (6.1)$$

where $\sum_{j=1}^{\infty} \|\pi_j\| < \infty$ and $\pi(z) = I_m - \sum_{j=1}^{\infty} \pi_j z^j = \varphi(z)^{-1}$ satisfies $\det\{\pi(z)\} \neq 0$ for $|z| \leq 1$.

Let $\Pi(k) = (\pi_1, \pi_2, \ldots, \pi_k)$ denote the first $k$ autoregressive coefficients in the VAR($\infty$) representation. 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 $T$. The OLS estimators of the autoregressive coefficients of the fitted VAR($k$) model and variance-covariance matrix, $\hat{\Sigma}_u^k$, are given by the following equation:

$$\hat{\Pi}(k) = (\hat{\pi}_{1k}, \hat{\pi}_{2k}, \ldots, \hat{\pi}_{kk}) = \hat{\Gamma}_{k1}^{-1} \hat{\Gamma}_{kk}^{-1}, \quad \hat{\Sigma}_u^k = \sum_{t=k+1}^{T} \hat{u}_k(t) \hat{u}_k(t)' / (T - k),$$
Remark 6.1. The upper bound $K$ of the order $k$ in the fitted VAR($k$) model depends on the assumptions required to ensure convergence and the asymptotic distribution of the estimator. For convergence of the estimator, we need to assume that $k^2/T \to 0$, as $k$ and $T \to \infty$. Consequently, we can choose $K = CT^{1/2}$, where $C$ is a constant, as an upper bound. For the asymptotic distribution of the estimator we need to assume that $k^3/T \to 0$, as $k$ and $T \to \infty$, and thus we can choose as an upper bound $K = CT^{1/3}$.

Now we will calculate an asymptotic approximation of the variance-covariance matrix of the forecast error at horizon $h$. The forecast error of $W(t+h)$, based on the VAR($\infty$) model, is given by:

$$e_{nc}[W(t+h) \mid W(t), W(t-1), \ldots] = \sum_{j=0}^{h-1} \varphi_j u(t+h-j),$$

associated with the variance-covariance matrix

$$\Sigma[W(t+h) \mid W(t), W(t-1), \ldots] = \sum_{j=0}^{h-1} \varphi_j \Sigma_u \varphi_j'.$$

In the same way, the variance-covariance matrix of the forecast error of $W(t+h)$, based on the VAR($k$) model, is given by:

$$\Sigma_k[W(t+h) \mid W(t), \ldots, W(t-k+1)] = E[(W(t+h) - \sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t+1-j)) \times$$

$$(W(t+h) - \sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t+1-j))'],$$

where

$$\hat{\pi}_{jk}^{(h+1)} = \hat{\pi}_{jk}^{(h+1)} + \hat{\pi}_{jk}^{(h)} \hat{\pi}_{jk}^{(1)}, \quad \hat{\pi}_{jk}^{(1)} = \hat{\pi}_{jk}, \quad \hat{\pi}_{jk}^{(0)} = I_m, \quad \text{for } j \geq 1, \ h \geq 1;$$

and

$$\hat{\pi}_{jk} = (T-k)^{-1} \sum_{t=k+1}^{T} w(t)w(t)', \quad \text{for } w(t) = (W(t), \ldots, W(t-k+1)).$$

Theorem 1 in Lewis and Reinsel (1985) ensures convergence of $\hat{\pi}_{1k}, \hat{\pi}_{2k}, \ldots, \hat{\pi}_{kk}$ under three assumptions: (1) $E|u_i(t)u_j(t)uk(t)ui(t)| \leq \gamma_4 < \infty$, for $1 \leq i, j, k, l \leq m$; (2) $k$ is chosen as a function of $T$ such that $k^2/T \to 0$ as $k, T \to \infty$; and (3) $k$ is chosen as a function of $T$ such that $k^{1/2} \sum_{j=k+1}^{\infty} \|\pi_j\| \to \infty$ as $k, T \to \infty$. Theorem 4 from the same authors derives the asymptotic distribution for these estimators under 3 assumptions: (1) $E|u_i(t)u_j(t)uk(t)ui(t)| \leq \gamma_4 < \infty$, $1 \leq i, j, k, l \leq m$; (2) $k$ is chosen as a function of $T$ such that $k^3/T \to 0$ as $k, T \to \infty$; and (3) there exists $\{l(k)\}$ a sequence of $(km^2 \times 1)$ vectors such that $0 < M_1 \leq l(k) \|l(k)\|^2 = l(k)'l(k) \leq M_2 < \infty$, for $k = 1, 2, \ldots$. We also note that $\hat{\Sigma}_u^k$ converges to $\Sigma_u$, as $k$ and $T \to \infty$ [see Lütkepohl (1993, pages 308-309)].
see Dufour and Renault (1998). Moreover,

\[
W(t + h) - \sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t + 1 - j) = (W(t + h) - \sum_{j=1}^{\infty} \hat{\pi}_{j}^{(h)} W(t + 1 - j)) \\
- (\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t + 1 - j) - \sum_{j=1}^{\infty} \hat{\pi}_{j}^{(h)} W(t + 1 - j)) \\
= \sum_{j=0}^{h-1} \varphi_j u(t + h - j) - (\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t + 1 - j) - \sum_{j=1}^{\infty} \hat{\pi}_{j}^{(h)} W(t + 1 - j)).
\]  

(6.2)

where

\[
\hat{\pi}_{j}^{(h+1)} = \hat{\pi}_{j+1}^{(h)} + \pi_1^{(1)} \pi_j, \quad \hat{\pi}_{j}^{(1)} = \pi_j, \quad \hat{\pi}_{j}^{(0)} = I_m, \text{ for } j \geq 1 \text{ and } h \geq 1.
\]

Since the error terms \(u(t + h - j)\), for \(0 \leq j \leq (h - 1)\), are independent of \(W(t), W(t - 1), \ldots\) and \((\hat{\pi}_{1k}, \hat{\pi}_{2k}, \ldots, \hat{\pi}_{kk})\), the two terms on the right-hand side of equation (6.2) are independent. Thus,

\[
\Sigma_k [W(t + h) | W(t), W(t - 1), \ldots, W(t - k + 1)] = \Sigma_k [W(t + h) | W(t), W(t - 1), \ldots] \\
+ E[\left(\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t + 1 - j) - \sum_{j=1}^{\infty} \hat{\pi}_{j}^{(h)} W(t + 1 - j)\right) \left(\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t + 1 - j) - \sum_{j=1}^{\infty} \hat{\pi}_{j}^{(h)} W(t + 1 - j)\right)']
\]

(6.3)

As \(k \) and \(T \to \infty\), an asymptotic approximation of the last term in equation (6.3) is given by Theorem 6 in Lewis and Reinsel (1985):

\[
E[\left(\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t + 1 - j) - \sum_{j=1}^{\infty} \hat{\pi}_{j}^{(h)} W(t + 1 - j)\right) \left(\sum_{j=1}^{k} \hat{\pi}_{jk}^{(h)} W(t + 1 - j) - \sum_{j=1}^{\infty} \hat{\pi}_{j}^{(h)} W(t + 1 - j)\right)'] \\
\approx \frac{km}{T} \sum_{j=0}^{h-1} \varphi_j \Sigma_u \varphi_j'.
\]

Consequently, an asymptotic approximation of the variance-covariance matrix of the forecast error is equal to:

\[
\Sigma_k [W(t + h) | W(t), W(t - 1), \ldots, W(t - k + 1)] \approx (1 + \frac{km}{T}) \sum_{j=0}^{h-1} \varphi_j \Sigma_u \varphi_j'.
\]

(6.4)

An estimator of this quantity is obtained by replacing the parameters \(\varphi_j\) and \(\Sigma_u\) by their estimators
\( \hat{\varphi}_k \) and \( \hat{\Sigma}_u \), respectively.

We can also obtain an asymptotic approximation of the variance-covariance matrix of the constrained forecast error at horizon \( h \) following the same steps as before. We denote this variance-covariance matrix by:

\[
\Sigma_k[S(t+h) \mid S(t), S(t-1), \ldots, S(t-k+1)] \approx (1 + \frac{k(m_1 + m_3)}{T}) \sum_{j=0}^{h-1} \Phi_j \Sigma \epsilon \Phi_j',
\]

where \( \Phi_j \), for \( j = 1, \ldots, h-1 \), represents the coefficient of a VMA representation of the process constrained \( S \), and \( \Sigma \epsilon \) is the variance-covariance matrix of \( \epsilon(t) = (\epsilon_X(t)', \epsilon_Z(t)')' \). Based on these results, an asymptotic approximation of the causality measure from \( Y \) to \( X \) is given by:

\[
C^a(Y \rightarrow X/Z) = \ln \left[ \frac{\det[\sum_{j=0}^{h-1} e_X^c \Phi_j \Sigma \epsilon \Phi_j' e_X^c]}{\det[\sum_{j=0}^{h-1} e_X^c \hat{\Phi}_j \hat{\Sigma} \epsilon \hat{\Phi}_j' e_X^c]} \right] + \ln \left[ 1 - \frac{km_2}{T + km} \right],
\]

where \( e_X^c = \begin{bmatrix} I_{m_1} & 0 \\ 0 & 0 \end{bmatrix} \) and \( e_X^c = \begin{bmatrix} I_{m_1} & 0 \end{bmatrix} \). An estimator of this quantity will be obtained by replacing the unknown parameters by their estimators,

\[
\hat{C}^a(Y \rightarrow X/Z) = \ln \left[ \frac{\det[\sum_{j=0}^{h-1} e_X^c \hat{\Phi}_j \hat{\Sigma} \epsilon \hat{\Phi}_j' e_X^c]}{\det[\sum_{j=0}^{h-1} e_X^c \hat{\Phi}_j \hat{\Sigma} \epsilon \hat{\Phi}_j' e_X^c]} \right] + \ln \left[ 1 - \frac{km_2}{T + km} \right].
\]

7. Evaluation by simulation of causality measures

In this section, we propose a simple simulation-based technique to calculate causality measures at any horizon \( h \), for \( h \geq 1 \). To illustrate this technique we consider the same examples we used in section 1 and limit ourselves to horizons 1 and 2.

Since one source of bias in autoregressive coefficients is sample size, our 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. In what follows we describe an algorithm to calculate the causality measure at given horizon \( h \) using a large simulation technique:

1. given the parameters of the unconstrained model and its initial values, simulate a large sample of \( T \) observations, where \( T \) can be equal to 1000000, under the assumption that the probability distribution of the disturbance is completely specified;\(^2\)

2. estimate the constrained model using a large simulation;

3. calculate the constrained and unconstrained variance-covariance matrices of the forecast errors at horizon \( h \);

\(^2\)The form of the probability distribution of the disturbances does not affect the value of causality measures.
4. calculate the causality measure at horizon $h$ using the constrained and unconstrained variance-covariance matrices from step 3.

Now let us reconsider Example 1 from section 1:

\[
\begin{bmatrix}
X(t + 1) \\
Y(t + 1)
\end{bmatrix}
= \Pi \begin{bmatrix}
X(t) \\
Y(t)
\end{bmatrix} + u(t)
= \begin{bmatrix}
0.5 & 0.7 \\
0.4 & 0.35
\end{bmatrix} \begin{bmatrix}
X(t) \\
Y(t)
\end{bmatrix} + \begin{bmatrix}
u_X(t + 1) \\
u_Y(t + 1)
\end{bmatrix},
\]

(7.1)

Our illustration involves two steps. First, we calculate the theoretical values of the causality measures at horizons 1 and 2. We know from example (5.6) that for a bivariate VAR(1) model it is more easy to calculate the causality measure at any horizon $h$ using only the unconstrained parameters. This is not the case for models with more than one lag and two variables [see section (5)]. Second, we evaluate our causality measures using a large simulation technique and we compare them with theoretical values from step 1.

The theoretical values of causality measures at horizons 1 and 2 are recovered as follows.

1. We calculate the variances of the forecast errors of $X$ at horizons 1 and 2 using its own past and the past of $Y$. Note that

\[
\Sigma[(X(t + h), Y(t + h))' | X_t, Y_t] = \sum_{i=0}^{h-1} \Pi^i \Pi^i'.
\]

From this variance-covariance matrix we have $Var[X(t + 1) | X_t, Y_t] = 1$ and

\[
Var[X(t + 2) | X_t, Y_t] = \sum_{i=0}^{1} e \Pi^i \Pi^i' e' = 1.74,
\]

where $e = (1, 0)'$.

2. We calculate the variances of the forecast errors of $X$ at horizons 1 and 2 using only its own past. In this case we need to determine the structure of the constrained model. This is given by the following equation [see Example 3]:

\[
X(t + 1) = (\pi_{YY} + \pi_{XX})X(t) + (\pi_{YY} - \pi_{XX})X(t - 1) + \varepsilon_X(t + 1) + \theta \varepsilon_X(t),
\]

where $\pi_{YY} + \pi_{XX} = 0.85$ and $\pi_{YY} - \pi_{XX} = 0.015$. The parameters $\theta$ and $Var(\varepsilon_X(t)) = \sigma_{\varepsilon_X}^2$ are the solutions to the following system:

\[
(1 + \theta^2)\sigma_{\varepsilon_X}^2 = 1.6125, \quad \theta \sigma_{\varepsilon_X}^2 = -0.35.
\]

The set of possible solutions is \{$(\theta, \sigma_{\varepsilon_X}^2) = (-4.378, 0.08), (-0.2285, 1.53)$\}. To get an
invertible solution we must choose the combination which satisfies the condition $|\theta| < 1$, 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 given by: $\Sigma[X(t + 1) | X_t] = 1.53$, and the variance of the forecast error of $X$ at horizon 2 is $\Sigma[X(t + 2) | X_t] = 2.12$. Thus, we have:

$$C(Y \rightarrow X) = 0.425, \quad C(Y \rightarrow X) = 0.197.$$  

In a second step we use the algorithm described at the beginning of this section to evaluate the causality measures using a large simulation technique. In Table 1, we give results that we obtain for different lag orders $p$ in the constrained model with $T = 600000$. These confirm the convergence ensured by the law of large numbers.

Now consider Example 2 of section 1:

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

(7.2)

where

$$E[u(t)] = 0, \quad E[u(t)u(s)'] = \begin{cases} I_3 & \text{if } s = t \\ 0 & \text{if } s \neq t. \end{cases}$$

In Example 2, analytical calculation of the causality measures at horizons 1 and 2 is not easy. As we saw in section 1, in this example the variable $Y$ does not cause the variable $X$ at horizon 1, but causes it at horizon 2 (indirect causality). So, 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 for a sample size $T = 600000$ and for different lag orders $p$ in the constrained model, we get the results in Table 2. These results show clearly the presence of an indirect causality from $Y$ to $X$.  

<table>
<thead>
<tr>
<th>$p$</th>
<th>$C(Y \rightarrow X)$</th>
<th>$C(Y \rightarrow X)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.519</td>
<td>0.567</td>
</tr>
<tr>
<td>2</td>
<td>0.430</td>
<td>0.220</td>
</tr>
<tr>
<td>3</td>
<td>0.427</td>
<td>0.200</td>
</tr>
<tr>
<td>4</td>
<td>0.425</td>
<td>0.199</td>
</tr>
<tr>
<td>5</td>
<td>0.426</td>
<td>0.198</td>
</tr>
<tr>
<td>10</td>
<td>0.425</td>
<td>0.197</td>
</tr>
<tr>
<td>15</td>
<td>0.426</td>
<td>0.199</td>
</tr>
<tr>
<td>20</td>
<td>0.425</td>
<td>0.197</td>
</tr>
<tr>
<td>25</td>
<td>0.425</td>
<td>0.199</td>
</tr>
<tr>
<td>30</td>
<td>0.426</td>
<td>0.198</td>
</tr>
<tr>
<td>35</td>
<td>0.425</td>
<td>0.198</td>
</tr>
</tbody>
</table>
Table 2. Evaluation by simulation of $C(Y \rightarrow X \mid Z)$ and $C(Y \rightarrow X \mid Z)$ for Model 7.2

<table>
<thead>
<tr>
<th>$p$</th>
<th>$C(Y \rightarrow X \mid Z)$</th>
<th>$C(Y \rightarrow X \mid Z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.121</td>
</tr>
<tr>
<td>2</td>
<td>0.000</td>
<td>0.123</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>0.122</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.123</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.124</td>
</tr>
<tr>
<td>10</td>
<td>0.000</td>
<td>0.122</td>
</tr>
<tr>
<td>15</td>
<td>0.000</td>
<td>0.122</td>
</tr>
<tr>
<td>20</td>
<td>0.000</td>
<td>0.122</td>
</tr>
<tr>
<td>25</td>
<td>0.000</td>
<td>0.124</td>
</tr>
<tr>
<td>30</td>
<td>0.000</td>
<td>0.122</td>
</tr>
<tr>
<td>35</td>
<td>0.000</td>
<td>0.122</td>
</tr>
</tbody>
</table>

8. Confidence intervals

In this section, we assume that the process of interest $W \equiv \{W(s) = (X(s), Y(s), Z(s))': s \leq t\}$ follows a VAR($p$) model

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

or equivalently,

$$(I_3 - \sum_{j=1}^{p} \pi_j L^j)W(t) = u(t),$$

where the polynomial $\Pi(z) = I_3 - \sum_{j=1}^{p} \pi_j z^j$ satisfies $\det[\Pi(z)] \neq 0$, for $z \in \mathbb{C}$ with $|z| \leq 1$, and $\{u(t)\}_{t=0}^{\infty}$ is a sequence of i.i.d. random variables.

Remark 8.1 If we suppose that process $W$ follows a VAR($\infty$) model, then we can use Inoue and Kilian’s (2002) approach to get results that are similar to those developed in this section.

For a realization $W \equiv \{W(1), \ldots, W(T)\}$ of process (8.1), estimates of $\Pi = (\pi_1, \ldots, \pi_p)$ and $\Sigma_u$ are given by the following equations:

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

3We assume that $X$, $Y$, and $Z$ are univariate variables. However, it is easy to generalize the results of this section to the multivariate case.
From the above notations, the theoretical value of the causality measure from \( \hat{Y} \) may be defined as follows:

\[
\sum \text{estimators of the autoregressive coefficients of the VAR model (8.4)} \quad (8.3)
\]

Now assume we are interested in measuring causality from \( Y \) to \( X \) at horizon \( h \). In this case we need to know the structure of the marginal process \( \{ S(s) = (X(s), Z(s))', s \leq t \} \). This one has a VARMA(\( \bar{p}, \bar{q} \)) representation with \( \bar{p} \leq 3p \) and \( \bar{q} \leq 2p \),

\[
S(t) = \sum_{j=1}^{\bar{p}} \phi_j S(t-j) + \sum_{i=1}^{\bar{q}} \theta_i \varepsilon(t-i) + \varepsilon(t), \tag{8.3}
\]

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

\[
E[\varepsilon(t)] = 0, \quad E[\varepsilon(t)\varepsilon'(s)] = \left\{ \begin{array}{ll} \Sigma_\varepsilon & \text{if } s = t \\ 0 & \text{if } s \neq t \end{array} \right.,
\]

and \( \Sigma_\varepsilon \) is a positive definite matrix. Equation (8.3) can be written in the following reduced form,

\[
\phi(L)S(t) = \theta(L)\varepsilon(t),
\]

where \( \phi(L) = I_2 - \phi_1 L - \ldots - \phi_p L^p \) and \( \theta(L) = I_2 + \theta_1 L + \ldots + \theta_q L^q \). We assume that \( \theta(z) = I_2 + \sum_{j=1}^{q} \theta_j z^j \) satisfies \( \det[\theta(z)] \neq 0 \) for \( z \in \mathbb{C} \) and \( |z| \leq 1 \). Under this assumption, the VARMA(\( \bar{p}, \bar{q} \)) process is invertible and has a VAR(\( \infty \)) representation:

\[
S(t) - \sum_{j=1}^{\infty} \pi_j S(t-j) = \theta(L)^{-1} \phi(L)S(t) = \varepsilon(t). \tag{8.4}
\]

Now let \( \Pi^c = (\pi^c_1, \pi^c_2, \ldots) \) denote the matrix of all autoregressive coefficients in model (8.4) and \( \Pi^c(k) = (\pi^c_1, \pi^c_2, \ldots, \pi^c_k) \) denote its first \( k \) autoregressive coefficients. Suppose that we approximate (8.4) by a finite order VAR(\( k \)) model, where \( k \) depends on sample size \( T \). In this case, estimators of the autoregressive coefficients of the VAR(\( k \)) model and variance-covariance matrix, \( \Sigma_\varepsilon \), are given by:

\[
\hat{\Pi}^c(k) = (\hat{\pi}^c_{1k}, \hat{\pi}^c_{2k}, \ldots, \hat{\pi}^c_{kk}) = \hat{\Pi}^c_1 \hat{\Pi}^c_1^{-1}, \quad \hat{\Sigma}_\varepsilon = \sum_{t=k+1}^{T} \hat{\varepsilon}_k(t)\hat{\varepsilon}_k(t)'/ (T-k),
\]

where \( \hat{\Pi}^c_1 = (T-k)^{-1} \sum_{t=k+1}^{T} S_k(t)S_k(t)', \) for \( S_k(t) = (S'(t), \ldots, S'(t-k+1))' \), \( \hat{\Pi}^c_{kl} = (T-k)^{-1} \sum_{t=k+1}^{T} S_k(t)S_k(t+1)', \) and \( \hat{\varepsilon}_k(t) = S(t) - \sum_{j=1}^{k} \hat{\pi}_j S(t-j) \).

From the above notations, the theoretical value of the causality measure from \( Y \) to \( X \) at horizon \( h \) may be defined as follows:

\[
C(Y \rightarrow_h X | Z) = \ln \left( \frac{G(\text{vec}(\Pi^c), \text{vech}(\Sigma_\varepsilon))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_\varepsilon))} \right)
\]

25
where
\[
G(\text{vec}(\Pi^c), \text{vech}(\Sigma_e)) = \sum_{j=0}^{h-1} e'_c \pi_1^c \Sigma_e \pi_1^c e_c, \ e_c = (1, 1)',
\]
\[
H(\text{vec}(\Pi), \text{vech}(\Sigma_u)) = \sum_{j=0}^{h-1} e'_c \pi_1 \Sigma_u \pi_1 e_c, \ e_c = (1, 1)',
\]
with \(\pi_1^c(j) = \pi_1^c(j-1) + \pi_1^c(j-1) \pi_1^c\), for \(j \geq 2\), \(\pi_1^c(0) = I_2\), and \(\pi_1^c(1) = \pi_1^c\); see Dufour and Renault (1998). vec denotes the column stacking operator and vech is the column stacking operator that stacks the elements on and below the diagonal only.

By Corollary 5.2, it exists a function \(f(\cdot) : \mathbb{R}^{9(p+1)} \to \mathbb{R}^{4(k+1)}\) that associates the constrained parameters \((\text{vec}(\Pi^c), \text{vech}(\Sigma_e))\) to the unconstrained parameters \((\text{vec}(\Pi), \text{vech}(\Sigma_u))\) such that [see example (5.6)]
\[
(\text{vec}(\Pi^c), \text{vech}(\Sigma_e))' = f((\text{vec}(\Pi), \text{vech}(\Sigma_u))'),
\]
and
\[
C(Y \xrightarrow{h} X | Z) = \ln \left( \frac{G(f((\text{vec}(\Pi), \text{vech}(\Sigma_u))')))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))} \right),
\]
An estimate of the quantity \(C(Y \xrightarrow{h} X | Z)\) is given by:
\[
\hat{C}(Y \xrightarrow{h} X | Z) = \ln \left( \frac{G(\text{vec}(\hat{\Pi}^c(k))), \text{vech}(\hat{\Sigma}_{e,k}))}{H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u))} \right), \tag{8.5}
\]
where \(G(\text{vec}(\hat{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_{e,k}))\) and \(H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u))\) are estimates of the corresponding population quantities.

Now let consider the following assumptions.

**Assumption 8.2** The following conditions are satisfied [see Lewis and Reinsel (1985)]:

1. \(E | \varepsilon_i(t) \varepsilon_j(t) | \leq \gamma_1 < \infty\), for \(1 \leq h, i, j, l \leq 2\);
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 \(T^{1/2} \sum_{j=k+1}^{\infty} \parallel \pi_j^c \parallel \to 0\) 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.

**Assumption 8.3** \(f(\cdot)\) is a continuous and differentiable function.

We have the following results [see proofs in the appendix].
Proposition 8.4 (Consistency of \( \hat{C}(Y \rightarrow X | Z) \)) Under assumption (8.2), \( \hat{C}(Y \rightarrow X | Z) \) is a consistent estimator of \( C(Y \rightarrow X | Z) \).

To establish the asymptotic distribution of \( \hat{C}(Y \rightarrow X | Z) \), let us start by recalling the following result [see Lütkepohl (1990, page 118-119) and Kilian (1998a, page 221)]:

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

(8.6)

where

\[
\Omega = \begin{pmatrix}
\Gamma^{-1} \otimes \Sigma_u & 0 \\
0 & 2(D_3' D_3)^{-1} D_3' (\Sigma_u \otimes \Sigma_u) D_3 (D_3' D_3)^{-1}
\end{pmatrix},
\]

\( D_3 \) is the duplication matrix, defined such that \( \text{vech}(F) = D_3 \text{vech}(F) \) for any symmetric \( 3 \times 3 \) matrix \( F \).

Proposition 8.5 (Asymptotic distribution of \( \hat{C}(Y \rightarrow X | Z) \)) Under assumptions (8.2) and (8.3), we have:

\[
(T - p)^{1/2} [\hat{C}(Y \rightarrow X|Z) - C(Y \rightarrow X|Z)] \xrightarrow{d} N(0, \Sigma_C)
\]

where \( \Sigma_C = D_C \Omega D_C' \) and

\[
D_C = \frac{\partial C(Y \rightarrow X|Z)}{\partial (\text{vec}(\Pi)', \text{vech}(\Sigma_u)')},
\]

\[
\Omega = \begin{pmatrix}
\Gamma^{-1} \otimes \Sigma_u & 0 \\
0 & 2(D_3' D_3)^{-1} D_3' (\Sigma_u \otimes \Sigma_u) D_3 (D_3' D_3)^{-1}
\end{pmatrix}.
\]

Analytically differentiating the causality measure with respect to the vector \((\text{vec}(\Pi)', \text{vech}(\Sigma_u)')\)' is not feasible. One way to build confidence intervals for our causality measures is to use a large simulation technique [see section 7] to calculate the derivative numerically. Another way is by 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. The 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 by construction preserves 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 by OLS and save the residuals

\[
\tilde{u}(t) = W(t) - \sum_{j=1}^{p} \hat{\pi}_j W(t-j), \text{ for } t = p + 1, \ldots, T,
\]
where \( \hat{\pi}_j \), for \( j = 1, \ldots, p \), are given by equation (8.2).

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

3. Generate a random draw for the vector of \( p \) initial observations \( w(0) = (W'(1), \ldots, W'(p))' \).

4. Given \( \hat{\Pi} = (\hat{\pi}_1, \ldots, \hat{\pi}_p) \), \( \tilde{u}^*(t) \), and \( w(0) \), generate bootstrap data for the dependent variable \( W^*(t) \) from equation:

\[
W^*(t) = \sum_{j=1}^{p} \hat{\pi}_j W^*(t - j) + \tilde{u}^*(t), \quad \text{for } t = p + 1, \ldots, T.
\]

5. Calculate the bootstrap OLS regression estimates

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

where \( \hat{\Gamma}^* = (T - p)^{-1} \sum_{t=p+1}^{T} w^*(t)w^*(t)' \), for \( w^*(t) = (W'(t), \ldots, W'(t - p + 1))' \), \( \hat{\Gamma}_1^* = (T - p)^{-1} \sum_{t=p+1}^{T} w^*(t)W^*(t + 1)' \), and \( \tilde{u}^*(t) = W^*(t) - \sum_{j=1}^{p} \hat{\pi}_j W^*(t - j) \).

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 \( h \), denoted \( \hat{C}^{(j)}(Y \rightarrow X | Z) \), using equation (8.5).

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

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

\[
(T - p)^{1/2} \left( \begin{array}{c} \text{vec}(\hat{\Pi}^*) - \text{vec}(\hat{\Pi}) \\ \text{vech}(\hat{\Sigma}_u^*) - \text{vech}(\hat{\Sigma}_u) \end{array} \right) \xrightarrow{d} N(0, \Omega), \quad (8.7)
\]

where

\[
\Omega = \begin{pmatrix} \Gamma^{-1} \otimes \Sigma_u & 0 \\ 0 & 2(D_3'D_3)^{-1}D_3'(\Sigma_u \otimes \Sigma_u)D_3(D_3'D_3)^{-1} \end{pmatrix},
\]

\( D_3 \) is the duplication matrix defined such that \( \text{vech}(F) = D_3 \text{vech}(F) \) for any symmetric \( 3 \times 3 \) matrix \( F \).

We have the following result which establish the validity of the percentile bootstrap technique.
Proposition 8.6 (Asymptotic validity of the residual-based bootstrap) Under assumptions (8.2) and (8.3), we have:

\[(T - p)^{1/2}(\hat{C}^*(Y \rightarrow h X|Z) - \hat{C}(Y \rightarrow h X|Z)) \overset{d}{\rightarrow} N(0, \Sigma_C),\]

where \(\Sigma_C = D_C \Omega D_C'\) and

\[D_C = \frac{\partial C(Y \rightarrow h X|Z)}{\partial (\text{vec}(\Pi)', \text{vech}(\Sigma_u)'},\]

\[\Omega = \begin{pmatrix} \Gamma^{-1} \otimes \Sigma_u & 0 \\ 0 & 2(D_3'D_3)^{-1}D_3'(\Sigma_u \otimes \Sigma_u)D_3(D_3'D_3)^{-1} \end{pmatrix}.\]

Remark 8.7 Proposition 8.6 may be used to justify the application of bias-corrected bootstrap confidence intervals, proposed by Kilian (1998), to improve the performance of the small sample percentile bootstrap intervals of our causality measures.

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. However, if one assume that ordinary least-squares (OLS) small-sample bias can be responsible for bias in the estimated impulse response function, replacing the biased slope coefficient estimates by bias-corrected slope coefficient estimates may help to reduce the bias in the impulse response function. He also shows that the additional modifications proposed in the bias-corrected bootstrap confidence intervals method, for improving the performance of the percentile bootstrap intervals in small samples, do not alter its asymptotic validity. The reason is that the effect of bias corrections is negligible asymptotically.

9. Empirical illustration

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

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

We performed an Augmented Dickey-Fuller test (hereafter ADF-test) for nonstationarity of the four variables of interest and their first differences. The values of the test statistics, as well as the critical values corresponding to a 5% significance level, are given in tables 3 and 4.
Figure 1: NBR in logarithmic form

Figure 2: \( r \) in logarithmic form

Figure 3: \( P \) in logarithmic form

Figure 4: GDP in logarithmic form
Figure 5: The first differentiation of ln(NBR)

Figure 6: The first differentiation of ln(r)

Figure 7: The first differentiation of ln(P)

Figure 8: The first differentiation of ln(GDP)
Table 3. Augmented Dickey-Fuller tests for the variables in level

<table>
<thead>
<tr>
<th></th>
<th>ADF test statistic</th>
<th>5% Critical Value</th>
<th></th>
<th>ADF test statistic</th>
<th>5% Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NBR</strong></td>
<td>−0.510587</td>
<td>−2.8694</td>
<td><strong>With Intercept</strong></td>
<td>−1.916428</td>
<td>−3.4234</td>
</tr>
<tr>
<td><strong>R</strong></td>
<td>−2.386082</td>
<td>−2.8694</td>
<td><strong>With Intercept</strong></td>
<td>−2.393276</td>
<td>−3.4234</td>
</tr>
<tr>
<td><strong>P</strong></td>
<td>−1.829982</td>
<td>−2.8694</td>
<td></td>
<td>−0.071649</td>
<td>−3.4234</td>
</tr>
<tr>
<td><strong>GDP</strong></td>
<td>−1.142940</td>
<td>−2.8694</td>
<td></td>
<td>−3.409215</td>
<td>−3.4234</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>ADF test statistic</th>
<th>5% Critical Value</th>
<th></th>
<th>ADF test statistic</th>
<th>5% Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NBR</strong></td>
<td>−5.956394</td>
<td>−2.8694</td>
<td><strong>With Intercept</strong></td>
<td>−5.937564</td>
<td>−3.9864</td>
</tr>
<tr>
<td><strong>r</strong></td>
<td>−7.782581</td>
<td>−2.8694</td>
<td><strong>With Intercept</strong></td>
<td>−7.817214</td>
<td>−3.9864</td>
</tr>
<tr>
<td><strong>P</strong></td>
<td>−2.690660</td>
<td>−2.8694</td>
<td></td>
<td>−3.217793</td>
<td>−3.9864</td>
</tr>
<tr>
<td><strong>GDP</strong></td>
<td>−5.922453</td>
<td>−2.8694</td>
<td></td>
<td>−5.966043</td>
<td>−3.9864</td>
</tr>
</tbody>
</table>

Table 5, below, summarizes the results of the stationarity tests for all variables.

As we can read from Table 5, all variables in logarithmic form are nonstationary. However, their first differences are stationary except for the GDP deflator, P. We performed a nonstationarity test for the second difference of variable P. The test statistic values are equal to −11.04826 and −11.07160 for the ADF-test with only an intercept and with both intercept and trend, respectively. The critical values in both cases are equal to −2.8695 and −3.4235. Thus, the second difference of variable P is stationary.

Once the data is made stationary, we use a nonparametric approach for the estimation and Akaike’s information criterion to specify the orders of the VAR(κ) models which describe the constrained and unconstrained processes. To choose the upper bound on the admissible lag orders, K, we apply the results in Lewis and Reinsel (1985). Using Akaike’s criterion for the unconstrained VAR model, which corresponds to four variables, we observe that it is minimized at κ = 16. We use same criterion to specify the orders of the constrained VAR models, which correspond to different combinations of three variables, and we find that the orders are all less than or equal to 16. To

Table 5. Unit root test results

<table>
<thead>
<tr>
<th></th>
<th>Variables in logarithmic form</th>
<th>First difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NBR</strong></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>r</strong></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>P</strong></td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>GDP</strong></td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
compare the determinants of the variance-covariance matrices of the constrained and unconstrained forecast errors at horizon \( h \), we take the same order \( k = 16 \) for the constrained and unconstrained models.

We calculate different causality measures for horizons \( h = 1, \ldots, 40 \) [see Figures 9–14 in the appendix]. Higher values of measures indicate greater causality. We also calculate the corresponding nominal 95% bootstrap confidence intervals as described in the previous section.

Based on the results, we have the following conclusions. From Figure 9 we observe that nonborrowed reserves (NBR) have a considerable effect on the federal funds rate (\( r \)) at horizon 1 comparatively with other variables [see Figures 10 and 11]. We conclude that nonborrowed reserves cause the federal funds rate only in the short term. This effect is well known in the literature and can be explained by the theory of supply and demand for money. Furthermore, we note that nonborrowed reserves have a short-term effect on GDP and can cause the GDP deflator until horizon 4. Figure 14 shows the effect of GDP on the federal funds rate is significant for the first four horizons. However, the effect of the federal funds rate on the GDP deflator is significant only at horizon 1; see Figure 12. Other significant results concern the causality from \( r \) to GDP. Figure 13 shows the effect of the interest rate on GDP is significant until horizon 16. These results are consistent with conclusions obtained by Dufour et al. (2005).

Table 6 represents results of other causality directions until horizon 20. As we can read from this table, there is no causality in these directions. Finally, note that the above results do not change when we consider the second, rather than first, difference of variable \( P \).

### 10. Conclusion

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

Another related problem arises when measuring the importance of the causality between two variables. Existing causality measures have been established only for horizon 1 and fail to capture indirect causal effects. This paper proposes a generalization of such measures for any horizon \( h \). We propose parametric and nonparametric measures for feedback and instantaneous effects at any horizon \( h \). Parametric measures are defined in terms of impulse response coefficients in the VMA representation. By analogy with Geweke (1982), we show that it is always possible to define a measure of dependence at horizon \( h \) which can be decomposed into a sum of feedback measures from \( X \) to \( Y \), from \( Y \) to \( X \), and an instantaneous effect at horizon \( h \). We also show how these causality measures can be related to the predictability measures developed in Diebold and Kilian (1998).

We propose a new approach to estimating these measures based on simulating a large sample from the process of interest. We also propose a valid nonparametric confidence interval, using the bootstrap technique.
Figure 9: Causality measures from Nonborrowed reserves to Federal funds rate

Figure 10: Causality measures from Nonborrowed reserves to GDP Deflator

Figure 11: Causality measures from Nonborrowed reserves to Real GDP

Figure 12: Causality measures from Federal funds rate to GDP Deflator
Figure 13: Causality measures from Federal funds rate to Real GDP

Figure 14: Causality measures from Real GDP to Federal funds rate
Table 6. Summary of causality relations at various horizons for series in first difference

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBR → R</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NBR → P</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NBR → GDP</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>R → NBR</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R → P</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>R → GDP</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>P → NBR</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>P → R</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>P → GDP</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>GDP → NBR</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>GDP → R</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>GDP → P</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
From an empirical application we find that nonborrowed reserves cause the federal funds rate only in the short term, the effect of real gross domestic product on the federal funds rate is significant for the first four horizons, the effect of the federal funds rate on the gross domestic product deflator is significant only at horizon 1, and finally the federal funds rate causes the real gross domestic product until horizon 16.
A. Appendix: Proofs

PROOF OF PROPOSITION 4.6. From definition 4.1,
\[
C(Y \rightarrow X \mid Z) = \ln \left[ \frac{\det(\Sigma(t + h_1) \mid I_t - Y_t)}{\det(\Sigma(t + h_1) \mid I_t)} \right] + \ln \left[ \frac{\det(\Sigma(I_t) \mid I_t - Y_t)}{\det(\Sigma(I_t) \mid I_t)} \right]
\]
\[
= C(Y \rightarrow X \mid Z) + \ln \left[ \frac{\det(\Sigma(t + h_1) \mid I_t)}{\det(\Sigma(t + h_2) \mid I_t)} \right] - \ln \left[ \frac{\det(\Sigma(t + h_1) \mid I_t - Y_t)}{\det(\Sigma(t + h_2) \mid I_t - Y_t)} \right].
\]

According to Diebold and Kilian (1998), the predictability measure of vector \( X \) under the information sets \( I_t - Y_t \) and \( I_t \) are, respectively, defined as follows:
\[
\tilde{P}_X(I_t - Y_t, h_1, h_2) = 1 - \frac{\det(\Sigma(t + h_1) \mid I_t - Y_t)}{\det(\Sigma(t + h_2) \mid I_t - Y_t)},
\]
\[
\tilde{P}_X(I_t, h_1, h_2) = 1 - \frac{\det(\Sigma(t + h_1) \mid I_t - Y_t)}{\det(\Sigma(t + h_2) \mid I_t - Y_t)}.
\]

Hence the result to be proved:
\[
C(Y \rightarrow X \mid Z) - C(Y \rightarrow X \mid Z) = \ln \left[ 1 - \tilde{P}_X(I_t - Y_t, h_1, h_2) \right] - \ln \left[ 1 - \tilde{P}_X(I_t, h_1, h_2) \right].
\]

PROOF (CONSISTENCY OF \( \tilde{C}(Y \rightarrow X \mid Z) \)). Under assumption 1 and using Theorem 6 in Lewis and Reinsel (1985),
\[
G(\text{vec}(\tilde{\Pi}^c(k)), \text{vech}(\tilde{\Sigma}_c, k)) = (1 + \frac{2k}{T})G(\text{vec}(\Pi^c), \text{vech}(\Sigma_c))
\]
\[
= (1 + O(T^{-\delta}))G(\text{vec}(\Pi^c), \text{vech}(\Sigma_c)), \text{ for } \frac{2}{3} < \delta < 1.
\]

The second equality follows from condition 2 of assumption 1. If we consider that \( k = T^\alpha \) for \( \alpha > 0 \), then condition 2 implies that \( k^3/T = T^{3\alpha-1} \) with \( 0 < \alpha < \frac{1}{3} \). Similarly, we have \( \frac{2k}{T} = 2T^{\alpha-1} \) and \( T^\delta(2T^{\alpha-1}) \rightarrow 2 \) for \( \frac{2}{3} < \delta < 1 \). Thus, for \( \frac{2}{3} < \delta < 1 \),
\[
\ln \left( G(\text{vec}(\tilde{\Pi}^c(k)), \text{vech}(\tilde{\Sigma}_c, k)) \right) = \ln \left( G(\text{vec}(\Pi^c), \text{vech}(\Sigma_c)) \right) + \ln (1 + O(T^{-\delta}))
\]
\[
= \ln \left( G(\text{vec}(\Pi^c), \text{vech}(\Sigma_c)) \right) + O(T^{-\delta}), \quad \text{(A.1)}
\]

38
For $H(.)$ a continuous function in $(\text{vec}(\Pi), \text{vech}(\Sigma_u))$ and because $\hat{\Pi} \rightarrow p \Pi$, $\hat{\Sigma}_u \rightarrow p \Sigma_u$, we have

$$\ln\left( \frac{H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))} \right) \rightarrow p \ln\left( \frac{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))} \right).$$  \hspace{1cm} (A.2)

Thus, from (A.1)-(A.2) and for $\frac{2}{3} < \delta < 1$, we get

$$\hat{C}(Y \rightarrow h X \mid Z) = \ln \left( \frac{G(\text{vec}(\hat{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_{\varepsilon,k}))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))} \right) + O(T^{-\delta}) + o_p(1).$$

Consequently,

$$\hat{C}(Y \rightarrow h X \mid Z) \rightarrow p C(Y \rightarrow h X \mid Z).$$

PROOF OF PROPOSITION 4. We have shown [see proof of consistency] that, for $\frac{2}{3} < \delta < 1$,

$$\ln \left( \frac{G(\text{vec}(\hat{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_{\varepsilon,k}))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))} \right) = \ln \left( \frac{G(f(\text{vec}(\Pi)), \text{vech}(\Sigma_u))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))} \right) + O(T^{-\delta}).$$  \hspace{1cm} (A.3)

Under Assumption 8.3, we have

$$\ln\left( \frac{G(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))}{G(\text{vec}(\Pi), \text{vech}(\Sigma_u)))} \right) \rightarrow p \ln\left( \frac{G(\text{vec}(\Pi), \text{vech}(\Sigma_u)))}{G(\text{vec}(\Pi), \text{vech}(\Sigma_u)))} \right).$$  \hspace{1cm} (A.4)

Thus, from (A.3)–(A.4) and for $\frac{2}{3} < \delta < 1$, we get:

$$\ln \left( \frac{G(\text{vec}(\hat{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_{\varepsilon,k}))}{G(\text{vec}(\Pi), \text{vech}(\Sigma_u)))} \right) = \ln \left( \frac{G(f(\text{vec}(\Pi)), \text{vech}(\Sigma_u)))}{G(\text{vec}(\Pi), \text{vech}(\Sigma_u)))} \right) + O(T^{-\delta}) + o_p(1).$$

Consequently, for $\frac{2}{3} < \delta < 1$,

$$\hat{C}(Y \rightarrow h X \mid Z) = \ln \left( \frac{G(f(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))} \right) + O(T^{-\delta}) + o_p(1)

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

where

$$\hat{C}(Y \rightarrow h X \mid Z) = \ln \left( \frac{G(f(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))} \right).$$

Since

$$\ln \left( \frac{G(f(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))}{H(\text{vec}(\Pi), \text{vech}(\Sigma_u))} \right) = O_p(1),$$

the asymptotic distribution of $\hat{C}(Y \rightarrow h X \mid Z)$ will be the same as that of $\hat{C}(Y \rightarrow h X \mid Z)$. 39
Furthermore, using Assumption 8.3 and a first-order Taylor expansion of \( \tilde{C}(Y \rightarrow X | Z) \), we have:

\[
\tilde{C}(Y \rightarrow X | Z) = C(Y \rightarrow X | Z) + D_C \left( \frac{\text{vec}(\hat{\Pi}) - \text{vec}(\Pi)}{\text{vech}(\hat{\Sigma}_u) - \text{vech}(\Sigma_u)} \right) + o_p(T^{-\frac{1}{2}}),
\]

where

\[
D_C = \frac{\partial C(Y \rightarrow X | Z)}{\partial \left( \text{vec}(\Pi)', \text{vech}(\Sigma_u)' \right)},
\]

hence

\[
(T - p)^{1/2} [\tilde{C}(Y \rightarrow X | Z) - C(Y \rightarrow X | Z)] \approx D_C \left( \frac{(T - p)^{1/2} \text{vec}(\hat{\Pi}) - \text{vec}(\Pi)}{(T - p)^{1/2} \text{vech}(\hat{\Sigma}_u) - \text{vech}(\Sigma_u)} \right).
\]

From (8.6), we have

\[
(T - p)^{1/2} [\tilde{C}(Y \rightarrow X | Z) - C(Y \rightarrow X | Z)] \rightarrow \mathcal{N}(0, \Sigma_C),
\]

hence

\[
(T - p)^{1/2} [\tilde{C}(Y \rightarrow X | Z) - C(Y \rightarrow X | Z)] \rightarrow \mathcal{N}(0, \Sigma_C)
\]

where

\[
\Sigma_C = D_C \Omega D_C',
\]

\[
\Omega = \left( \begin{array}{cc} \Gamma^{-1} \otimes \Sigma_u & 0 \\ 0 & 2(D_3'D_3)^{-1}D_3'(\Sigma_u \otimes \Sigma_u)D_3(D_3'D_3)^{-1} \end{array} \right).
\]

\(D_3\) is the duplication matrix, defined such that \( \text{vech}(F) = D_3\text{vech}(F) \) for any symmetric \( 3 \times 3 \) matrix \( F \). \(\square\)

**Proof (Asymptotic Validity of the Residual-Based Bootstrap).** We start by showing that

\[
\text{vec}(\hat{\Pi}^*) \rightarrow_p \text{vec}(\hat{\Pi}), \quad \text{vech}(\hat{\Sigma}_u^*) \rightarrow_p \text{vech}(\hat{\Sigma}_u), \quad \text{vec}(\hat{\Pi}^{*\epsilon}(k)) \rightarrow_p \text{vec}(\hat{\Pi}^{\epsilon}(k)), \quad \text{vech}(\hat{\Sigma}_{\epsilon,k}) \rightarrow \text{vech}(\hat{\Sigma}_{\epsilon,k}).
\]

We first note that

\[
\text{vec}(\hat{\Pi}^*) = \text{vec}(\hat{\Pi}_1^{*\epsilon-1}) = \text{vec}((T - p)^{-1} \sum_{t=p+1}^T W(t+1)^* w^*(t) \hat{\Pi}^{*\epsilon-1})
\]

40
\[\begin{align*}
&= \text{vec}((T - p)^{-1} \sum_{t=p+1}^{T} [\hat{H} w^*(t) + \check{u}^*(t + 1)] w^*(t)' \hat{I}^s - 1) \\
&= \text{vec}(\hat{H}((T - p)^{-1} \sum_{t=p+1}^{T} w^*(t) w^*(t)') \hat{I}^s - 1) \\
&+ \text{vec}((T - p)^{-1} \sum_{t=p+1}^{T} \check{u}^*(t + 1) w^*(t)' \hat{I}^s - 1) \\
&= \text{vec}(\hat{I}^* \hat{I}^s - 1) + \text{vec}((T - p)^{-1} \sum_{t=p+1}^{T} \check{u}^*(t + 1) w^*(t)' \hat{I}^s - 1).
\end{align*}\]

Let \(\mathcal{F}_t^* = \sigma(\check{u}^*(1), \ldots, \check{u}^*(t))\) denote the \(\sigma\)-algebra generated by \(\check{u}^*(1), \ldots, \check{u}^*(t)\). Then,

\[\begin{align*}
E^*[\check{u}^*(t + 1) w^*(t)' \hat{I}^s - 1] &= E^*[E^*[\check{u}^*(t + 1) w^*(t)' \hat{I}^s - 1 | \mathcal{F}_t^*]] \\
&= E^*[E^*[\check{u}^*(t + 1) | \mathcal{F}_t^*] w^*(t)' \hat{I}^s - 1] = 0.
\end{align*}\]

By the law of large numbers,

\[(T - p)^{-1} \sum_{t=p+1}^{T} \check{u}^*(t + 1) w^*(t)' \hat{I}^s - 1 = E^*[\check{u}^*(t + 1) w^*(t)' \hat{I}^s - 1] + o_p(1).
\]

Thus,

\[\text{vec}(\hat{I}^*) - \text{vec}(\hat{H}) \to 0.\]

Now, to prove that \(\text{vech}(\hat{\Sigma}_u^*) \to \text{vech}(\hat{\Sigma}_u)\), we observe that

\[\begin{align*}
\text{vech}(\hat{\Sigma}_u^* - \hat{\Sigma}_u) &= \text{vech}[(T - p)^{-1} \sum_{t=p+1}^{T} \check{u}^*(t) \check{u}^*(t)' - (T - p)^{-1} \sum_{t=p+1}^{T} \check{u}(t) \check{u}(t)'] \\
&= \text{vech}[(T - p)^{-1} \sum_{t=p+1}^{T} (\check{u}^*(t) \check{u}^*(t)' - (T - p)^{-1} \sum_{t=p+1}^{T} \check{u}(t) \check{u}(t)')]
\end{align*}\]

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

\[\begin{align*}
E^*[\check{u}^*(t) \check{u}^*(t)' - (T - p)^{-1} \sum_{t=p+1}^{T} \check{u}(t) \check{u}(t)'] \\
= E^*[E^*[\check{u}^*(t) \check{u}^*(t)' - (T - p)^{-1} \sum_{t=p+1}^{T} \check{u}(t) \check{u}(t)' | \mathcal{F}_t^*]]
\end{align*}\]

41
\[ E^*[E^*[\tilde{u}^*(t)\tilde{u}^*(t)'] | \mathcal{F}_t^*] - (T - p)^{-1} \sum_{t=p+1}^{T} \tilde{u}(t)\tilde{u}(t)' \].

Because
\[ E^*[E^*[\tilde{u}^*(t)\tilde{u}^*(t)'] | \mathcal{F}_{t-1}^*] = (T - p)^{-1} \sum_{t=p+1}^{T} E^*[\tilde{u}^*(t)\tilde{u}^*(t)'], \]
then
\[ E^*[\tilde{u}^*(t)\tilde{u}^*(t)'] - (T - p)^{-1} \sum_{t=p+1}^{T} \tilde{u}(t)\tilde{u}(t)' = 0. \]

Since
\[ (T - p)^{-1} \sum_{t=p+1}^{T} (\tilde{u}^*(t)\tilde{u}^*(t)' - (T - p)^{-1} \sum_{t=p+1}^{T} \tilde{u}(t)\tilde{u}(t)') \]
\[ = E^*[\tilde{u}^*(t)\tilde{u}^*(t)'] - (T - p)^{-1} \sum_{t=p+1}^{T} \tilde{u}(t)\tilde{u}(t)'] + o_p(1), \]
we get
\[ \text{vec}(\hat{\Sigma}_u^*) - \text{vec}(\hat{\Sigma}_u) \to 0. \]

Similarly, we can show that
\[ \text{vec}(\hat{\Pi}^c(k)) \to \text{vec}(\hat{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_\varepsilon, k) \to \text{vech}(\hat{\Sigma}_\varepsilon, k). \]

For \( H(.) \) and \( G(.) \) continuous functions, we have
\[ \ln (H(\text{vec}(\hat{\Pi}^*), \text{vech}(\hat{\Sigma}_u^*))) = \ln (H(\text{vec}(\tilde{\Pi}), \text{vech}(\hat{\Sigma}_u))) + o_p(1), \]
\[ \ln (G(\text{vec}(\hat{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_\varepsilon, k))) = \ln (G(\text{vec}(\tilde{\Pi}^c(k)), \text{vech}(\hat{\Sigma}_\varepsilon))) + o_p(1). \]

By Theorems 2.1–3.4 in Paparoditis (1996) and Theorem 6 in Lewis and Reinsel (1985), we have, for \( \frac{3}{4} < \delta < 1, \)
\[ \ln (G(\text{vec}(\hat{\Pi}^{*c}(k)), \text{vech}(\hat{\Sigma}_\varepsilon^*, k))) = \ln (G(\text{vec}(\tilde{\Pi}^c), \text{vech}(\hat{\Sigma}_\varepsilon))) + O(T^{-\delta}). \]

Thus,
\[ \hat{C}^*(Y \to X \mid Z) = \ln \left( \frac{G(f(\text{vec}(\hat{\Pi}^*), \text{vech}(\hat{\Sigma}_u^*)))}{H(\text{vec}(\tilde{\Pi}^*), \text{vech}(\hat{\Sigma}_u))} \right) + O(T^{-\delta}) + o_p(1) \]
\[ = \hat{C}^*(Y \to X \mid ZZ) + O(T^{-\delta}) + o_p(1) \]

42
where
\[
\hat{C}^*(Y \rightarrow h X | Z) = \ln \left( \frac{G(f(\text{vec}(\hat{\Pi})), \text{vech}(\hat{\Sigma}^*_u)))}{H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))} \right).
\]

We have also shown [see the proof of Proposition A] that, for \( \frac{2}{3} < \delta < 1 \),
\[
\hat{C}(Y \rightarrow h X | Z) = \ln \left( \frac{G(f(\text{vec}(\hat{\Pi})), \text{vech}(\hat{\Sigma}^*_u)))}{H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))} \right) + O(T^{-\delta}) + o_p(1).
\]

Consequently
\[
\hat{C}^*(Y \rightarrow X | Z) = \ln \left( \frac{G(f(\text{vec}(\hat{\Pi})), \text{vech}(\hat{\Sigma}^*_u)))}{H(\text{vec}(\hat{\Pi}), \text{vech}(\hat{\Sigma}_u)))} \right) + O(T^{-\delta}) + o_p(1).
\]

Furthermore, by Assumption 8.3 and a first order Taylor expansion of \( \hat{C}^*(Y \rightarrow X | Z) \), we have
\[
\hat{C}^*(Y \rightarrow X | Z) = \hat{C}(Y \rightarrow X | Z) + D_C \left( \begin{array}{c}
\text{vec}(\hat{\Pi}^*) - \text{vec}(\hat{\Pi}) \\
\text{vech}(\hat{\Sigma}^*_u) - \text{vech}(\hat{\Sigma}_u)
\end{array} \right) + o_p(T^{1/2}),
\]
and
\[
(T - p)^{1/2} [\hat{C}^*(Y \rightarrow X | Z) - \hat{C}(Y \rightarrow X | Z)] \approx D_C \left( \begin{array}{c}
(T - p)^{1/2} (\text{vec}(\hat{\Pi}^*) - \text{vec}(\hat{\Pi})) \\
(T - p)^{1/2} (\text{vech}(\hat{\Sigma}^*_u) - \text{vech}(\hat{\Sigma}_u))
\end{array} \right).
\]

By (8.7),
\[
(T - p)^{1/2} [\hat{C}^*(Y \rightarrow X | Z) - \hat{C}(Y \rightarrow X | Z)] \overset{d}{\rightarrow} N(0, \Sigma_C),
\]
hence
\[
(T - p)^{1/2} [\hat{C}^*(Y \rightarrow X | Z) - \hat{C}(Y \rightarrow X | Z)] \overset{d}{\rightarrow} N(0, \Sigma_C)
\]
where
\[
\Sigma_C = D_C \Omega D_C',
\]
\[
\Omega = \left( \begin{array}{cc}
\Gamma^{-1} \otimes \Sigma_u & 0 \\
0 & 2(D_3D_3)^{-1} D_3' (\Sigma_u \otimes \Sigma_u) D_3 (D_3'D_3)^{-1}
\end{array} \right),
\]
\(D_3\) is the duplication matrix, defined such that \( \text{vech}(F) = D_3 \text{vech}(F) \) for any symmetric \( 3 \times 3 \) matrix \( F \).
References


