Causality Along Subspaces: Theory[*]

Majid M. Al-Sadoon[†]

April 18, 2009

Abstract

This paper extends previous notions of causality to take into account the subspaces along which causality occurs as well as long run causality. The properties of these new notions of causality are extensively studied for a wide variety of time series processes. The paper then proves that the notions of stability, cointegration, and controllability can all be recast under the single framework of causality.

JEL Classification: C1, C32.

Keywords: Granger causality, indirect causality, long run causality, stability, controllability, VARMA.

1   Introduction

1.1   Summary

One of the most important concepts to have risen out of the econometric time series literature has been the concept of Granger causality, first suggested by Wiener (1956) and later developed by Granger (1969). The literature has grown considerably since then, with extensions to multivariate series, larger information sets, longer horizons,... etc. (see Geweke (1984), Hamilton (1994), or Lütkepohl (2006)). Yet problems of interpretation have plagued it since its inception (see e.g. Hamilton (1994)) and some have argued that it fails to capture what is actually meant by causality (see Hoover (2001) or Pearl (2000)). Against this backdrop, the purpose of this paper is to demonstrate that Granger causality is a much deeper concept than previously thought, going to the heart of many other concepts in time series analysis. We do this without taking any particular stance on the philosophical or empirical applicability of Granger causality per se; when "cause" or any other word to that effect occurs in this paper it is to be understood in the purely mathematical sense of Definition 3.2.

This paper proposes two extensions to Dufour & Renault (1998) – henceforth DR: (i) we take into account the subspaces of non–causality and (ii) we consider the long run properties of causality. To motivate the first extension, suppose that X and Y are vector processes and Y Granger–causes X. Now it may be that variations in X along some directions cannot be attributed to Y. Likewise, it may be that certain linear combinations of Y do not help predict X. Thus standard Granger causality tests may not give the full picture of the dependence structure. To motivate the second extension, suppose Y consists of nominal variables while X consists of real variables. Standard economic theory says that Y should have no long run effect on X. Existing time–domain theory allow us to check whether Y fails to cause X in the long run if they can be modeled by cointegrated VARMA models (see e.g. Bruneau & Jondeau (1999) and Yamamoto & Kurozumi (2006)); it would be useful to obtain criteria for long run non–causality for a wider class of processes.

Based on the aforementioned extensions we are able to show: (i) stability and cotrendedness (a generalization of cointegration) for a wider range of processes can be reformulated in terms of long run non–causality and (ii) controllability can be reformulated in terms of non–causality at all horizons.

Now causality has been known to be associated with cointegration and controllability at least since Granger (1988b) and Granger (1988a). However the association with cointegration was known to hold only in the context of bivariate models; on the other hand, the association with controllability was only shown in rather extreme forms of optimal control, where the policymaker puts infinite weight on a single variable in the model. The two extensions proposed in this paper allow us to flesh out and develop the association in its full generality. We find that subspace non–causality subsumes wider phenomena that stability and cointegration as well as the linear systems concept of controllability (see e.g. Kailath (1980)). Along the way we will extend various results by DR to full generality.

The theoretical framework of this study is based on linear projections on Hilbert spaces, which was introduced by Kolmogorov (1941). This framework, which is widely used in time series analysis, is particularly well–suited to the study of linear processes due to its simplicity and geometric appeal. However, other frameworks for studying causality are possible; Engle et al. (1983) study non–causality in terms of independence of probability distributions, while Florens & Mouchart (1982) study non–causality in terms of the orthogonality properties of σ–algebras. The results of this paper map easily to these other perspectives although, possibly, at a cost – for example, the condition in Theorem 4.1 is sufficient in the Florens & Mouchart (1982) framework but for necessity one needs stronger assumptions (e.g. normality).

A number of papers have recently built on DR. Eichler (2007) uses DR's results to conduct a graph–theoretic analysis in light of recent advances in the artificial intelligence literature on causality (see e.g. Pearl (2000)). Hill (2007) develops DR's results into a procedure for finding the exact horizon at which fluctuations in one variable anticipate changes in another variable when the model is trivariate. There is also a strand of literature which has considered dependence along subspaces in time series analysis. Brillinger (2001) considers the problem of approximating a time series X by a filter of Y where the filter is of reduced rank and both series are stationary; his analysis could be adapted to identify $\mathcal{U}_h^{XYH}$ with $H = \overline{\mathrm{sp}}\{1\}$ if we replace Y by X lagged h periods.[1] Velu et al. (1986) consider the problem of identifying $\mathcal{U}_1^{XXH}$ with H as before when X is a stationary VAR of finite order. Finally, Otter (1990) and Otter (1991) consider the use of canonical correlations in forecasting and causality analysis assuming normality, stationarity, and finite information sets; in particular, the results of Otter

---

[1]$\mathcal{U}_h^{XYH}$ is the subspace along which Y fails to cause X at horizon h given information set H – see Definition 3.4.

(1991) can be used to characterize $\mathcal{U}_1^{XYH}$. The results of this paper generalize the previous as they require neither stationarity, nor normality, nor finite information sets.

The paper proceeds as follows. Section 2 overviews the main ideas from Hilbert space theory that we will need. Section 3 develops the concept of non–causality along subspaces as an extension to DR, providing the basic definitions and results at the most general level of analysis. Section 4 specializes the theory to linear invertible processes. Section 5 specializes again to invertible VARMA processes. Necessary and sufficient conditions for non–causality are provided at each step of the specialization of the theory. Section 6 considers the connection to controllability. Section 7 concludes and section 8 is an appendix.

2   Some Concepts from Hilbert Space Theory

Here we lay out the main background from Hilbert space theory that we will need. Excellent overviews of the applications of Hilbert space theory to time series analysis can be found in Brockwell & Davis (1991) and Pourahmadi (2001).

Let L^2 be the Hilbert space of random variables on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ having finite second moments and let $\mathbb{E}$ be the expectations operator in this space. We define the inner product be $\langle X, Y \rangle = \mathbb{E}(X\overline{Y})$ for all $X, Y \in L^2$ and the norm to be $\|X\|^2 = \langle X, X \rangle$ for all $X \in L^2$. We will say that a random vector is in L^2 if all its elements are in L^2. If H and G are subspaces of L^2 then we define $H + G = \overline{\mathrm{sp}}\{H, G\}$, the closure of the span of all linear combinations of the elements of G and H; the subspace $H - G$ is defined as $\overline{\mathrm{sp}}\{H \cap G^\perp\}$.[2]

The time indexing set will be $(\omega, \infty) \subseteq \mathbb{Z}$ for $\omega \in \{-\infty\} \cup \mathbb{Z}$ for all processes in this paper; the case $\omega \in \mathbb{Z}$ will be necessary in order to take into account some non–stationary time series. The information or history at time $t \in \mathbb{Z}$ is denoted by $I(t)$; we consider it to be a closed subspace of L^2 satisfying the nesting property, $\omega < t \leq t' \Rightarrow I(t) \subseteq I(t')$. If X is an n dimensional stochastic process in L^2 then for $\omega < t < t'$ we define, $X(t, t'] = \overline{\mathrm{sp}}\{X_{is} : t < s \leq t', 1 \leq i \leq n\}$; for $\omega < t \leq t'$, $X[t, t']$ is defined in a similar fashion. Then $X(\omega, t]$ is the information collected about X up to time t and we will say that information set I is conformable with X if $X(\omega, t] \subseteq I(t)$ for all $t > \omega$. The most frequently encountered

---

[2]The statistical literature uses "+" to refer to the linear span. However, DR use "+" to signify the closed linear span and we follow their notation. The two are not equivalent as demonstrated in example 9.6 of Pourahmadi (2001).

information sets in this paper are of the form, $I(t) = H + X(\omega, t]$ for all $t > \omega$ for some L^2 random vector process X, where $H \subseteq L^2$ is the information available in every period, thus it contains deterministic term when H is the trivial subspace $\overline{\mathrm{sp}}\{1\}$ but it may be larger allowing for random initial conditions.

If $X \in L^2$ and H is a subspace of L^2 then the orthogonal projection of X onto H (or the best linear predictor of X given H) is denoted by $P(X|H)$. If X is vector of n variables in L^2 then $P(X|H) = (P(X_1|H), \ldots, P(X_n|H))'$.

3 Cartesian Causality and Subspace Causality

In this section we will operate under the following assumption.

Assumption 1. For $\omega \in \{-\infty\} \cup \mathbb{Z}$, $X = \{X(t) : \omega < t < \infty\}$ and $Y = \{Y(t) : \omega < t < \infty\}$ are discrete–time stochastic processes in L^2, of dimensions n_X and n_Y respectively. We also take I to be an information set. $\qquad\square$

We will be interested in studying the causal links between X and Y in the context of information set I. Typically, I is assumed to include all the variables that may be causally related to X including X and excluding Y; thus the totality of information in I and Y consists of everything that may be causally related to X – Hoover (2001) refers to this larger information set as the "causal field" of X. DR typically take I to include an auxiliary process Z through which there may be indirect effects of Y on X (see DR for further motivation and background). It is important to note that as far as Assumption 1 and the results derived from it are concerned, X and Y need not be distinct and in discussing the causal effects of a time series on its future evolution, we will be interested in the case $Y = X$.

The following definition, which appears in Granger (1980), is the main building block of Granger causality.

Definition 3.1 (Prediction Variation)**.** Under Assumption 1 with $h \geq 1$ we have,

$$\Delta_h^{XYI}(t) = P(X(t+h)|I(t) + Y(\omega, t]) - P(X(t+h)|I(t)), \qquad t > \omega$$

is the time–t prediction variation of X at horizon h due to Y when I is given. $\qquad\square$

The prediction variation $\Delta_h^{XYI}(t)$ is the modification to the h–period–ahead forecast of X based on information set $I(t)$, when the forecast is made on additional information on Y. By

Theorem 9.18(c) of Pourahmadi (2001), $\Delta_h^{XYI}(t) = P(X(t+h)|(I(t) + Y(\omega, t]) - I(t)).^3$ The

idea of Granger causality is that if Y causes X, Y should be helpful for predicting X over and

above the information in I. If not then $\Delta_h^{XYI}(t) = 0$ for all $t > \omega$ and the best linear predictor

of X at horizon h is independent of the history of Y when the information set I is specified;

in this case, the causal channels from I mitigate the influence of Y on X at horizon h.[4] Note

that by definition, $P(\Delta_h^{XYI}(t)|I(t)) = 0$ for all $t > \omega$; therefore the prediction variation is

linear in $Y(t), Y(t-1), \ldots$ and orthogonal to I.

Definition 3.2 (Cartesian Non–causality). Under Assumption 1 with $1 \le h < \infty$, we have

the following definitions,

(i) Y does not cause X given I at horizon h if $\Delta_h^{XYI}(t) = 0$ for all $t > \omega$. We denote this

by $Y \nrightarrow_h X \, [\, I \,]$.

(ii) Y does not cause X given I in the long run if $\Delta_j^{XYI}(t) \to 0$ in L^2 as $j \to \infty$ for all $t > \omega$.

We denote this by $Y \nrightarrow_\infty X \, [\, I \,]$.

(iii) Y does not cause X given I up to horizon h if $Y \nrightarrow_j X \, [\, I \,]$ for all $1 \le j \le h$. We denote

this by $Y \nrightarrow_{(h)} X \, [\, I \,]$.

(iv) Y does not cause X given I at any horizon if $Y \nrightarrow_j X \, [\, I \,]$ for all $j \ge 1$. We denote this

by $Y \nrightarrow_{(\infty)} X \, [\, I \,]$.

When it is clear from the context and there is no danger of confusion we drop the "given I"

phrase in the above definitions. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

When $h < \infty$ and $Y \nrightarrow_h X \, [\, I \,]$, $\Delta_h^{XYI}(t) = 0$ for all $t > \omega$ and there is no effect of

Y on X at horizon h. When $Y \nrightarrow_\infty X \, [\, I \,]$, the effect dissipates in the long run; this does

not, however, rule out the possible effect of Y on X in the short run.[5] (i), (iii), and (iv)

are due to DR although they require I to be conformable with X, which we do not. (ii)

generalizes Bruneau & Jondeau (1999) and Yamamoto & Kurozumi (2006) as they require

$\lim_{h\to\infty} P(X(t+h)|I(t) + Y(\omega, t]) = \lim_{h\to\infty} P(X(t+h)|I(t))$, where as we do not require

---

[3] Note that generally, $(I(t) + Y(\omega, t]) - I(t) \ne Y(\omega, t]$ although $(I(t) + Y(\omega, t]) - I(t) = Y(\omega, t] - I(t)$.

[4] This is similar to the idea of "screening off" that Hoover (2001) and Pearl (2000) utilize.

[5] We define the long run in terms of L^2 limits as this form of convergence is the most natural one for working in

L^2. In the Engle et al. (1983) framework, convergence in distribution seems more suitable; on the other hand, almost

sure or L^1 convergence would be more appropriate for generalizing the Florens & Mouchart (1982) framework.

these limits to exist. (iii) and (iv) are derived from (i) and describe non–causality over several periods and over all periods respectively; thus (iii) and (iv) will inherit some of the properties of (i). Being effectively the "primitives" of our definition, (i) and (ii) will capture most of our attention in this paper.

We refer to the notions of non–causality in Definition 3.2 as cartesian non–causality because they concern the cartesian components of W. Unfortunately, cartesian causality cannot capture the full range of dependence between X and Y. If X is causally related to Y, it may be that X varies only along limited directions in response to Y or that variations in Y along certain directions have no effect on X. In order to analyze these cases rigorously, we define some new concepts.

Definition 3.3 (Subspace Non–causality)**.** Under Assumption 1, with $1 \leq h < \infty$, subspaces $\mathcal{U} \subseteq \mathbb{R}^{n_X}$ and $\mathcal{V} \subseteq \mathbb{R}^{n_Y}$, and orthogonal projection matrices $P_{\mathcal{U}}$ and $P_{\mathcal{V}}$ (onto $\mathcal{U}$ and $\mathcal{V}$ respectively), we have the following definitions,

(i) Y along $\mathcal{V}$ does not cause X along $\mathcal{U}$ given I at horizon h if $P_{\mathcal{V}}Y \nrightarrow_h P_{\mathcal{U}}X$ $[I]$. We denote this by, $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}$ $[I]$.

(ii) Y along $\mathcal{V}$ does not cause X along $\mathcal{U}$ given I in the long run if $P_{\mathcal{V}}Y \nrightarrow_\infty P_{\mathcal{U}}X$ $[I]$. We denote this by, $Y|_{\mathcal{V}} \nrightarrow_\infty X|_{\mathcal{U}}$ $[I]$.

(iii) Y along $\mathcal{V}$ does not cause X along $\mathcal{U}$ given I up to horizon h if $P_{\mathcal{V}}Y \nrightarrow_{(h)} P_{\mathcal{U}}X$ $[I]$. We denote this by, $Y|_{\mathcal{V}} \nrightarrow_{(h)} X|_{\mathcal{U}}$ $[I]$.

(iv) Y along $\mathcal{V}$ does not cause X along $\mathcal{U}$ given I at all horizons if $P_{\mathcal{V}}Y \nrightarrow_{(\infty)} P_{\mathcal{U}}X$ $[I]$. We denote this by, $Y|_{\mathcal{V}} \nrightarrow_{(\infty)} X|_{\mathcal{U}}$ $[I]$.

When $\mathcal{U} = \mathbb{R}^{n_X}$ we will drop any reference to $\mathcal{U}$ (e.g. we will write $Y|_{\mathcal{V}} \nrightarrow_h X$ $[I]$ instead of $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathbb{R}^{n_X}}$ $[I]$). Similarly, when $\mathcal{V} = \mathbb{R}^{n_Y}$ we write $Y \nrightarrow_h X|_{\mathcal{U}}$ $[I]$ instead of $Y|_{\mathbb{R}^{n_Y}} \nrightarrow_h X|_{\mathcal{U}}$ $[I]$. Finally, as in Definition 3.2, we will drop the "given I" phrase in the above definitions when there is no danger of confusion . $\square$

Thus, subspace non–causality merely augments the definition of cartesian non–causality with projections of X and Y along certain subspaces. An alternative, and equivalent, way of defining subspace non–causality would have been to consider those linear combinations of X and Y that are not causally related as demonstrated in the following lemma.

Lemma 3.1 (The Matrix Characterization of Subspace Non–causality)**.** Under Assumption 1 with $1 \leq h \leq \infty$, $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\,[\,I\,]$ if and only if $V'Y \nrightarrow_h U'X\,[\,I\,]$, where the columns of U are an orthonormal basis for $\mathcal{U}$ and the columns of V are an orthonormal basis for $\mathcal{V}$.

Thus, $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\,[\,I\,]$ if and only if the linear combinations $V'Y$ fail to help forecast the linear combinations $U'X$ at horizon h. We are now ready to consider the properties of subspace non–causality.

Lemma 3.2. Under Assumption 1 with $1 \leq h \leq \infty$ and arbitrary indexing set J,

  (i) (Cause Monotonicity) $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\,[\,I\,]$ if and only if $Y|_{\mathcal{W}} \nrightarrow_h X|_{\mathcal{U}}\,[\,I\,]$ for all $\mathcal{W} \subseteq \mathcal{V}$.

 (ii) (Effect Monotonicity) $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\,[\,I\,]$ if and only if $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{W}}\,[\,I\,]$ for all $\mathcal{W} \subseteq \mathcal{U}$.

(iii) (Cause Additivity) If $Y|_{\mathcal{V}_j} \nrightarrow_h X|_{\mathcal{U}}\,[\,I\,]$ for all $j \in J$ then $Y|_{\sum_{j \in J} \mathcal{V}_j} \nrightarrow_h X|_{\mathcal{U}}\,[\,I\,]$.

(iv) (Effect Additivity) If $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}_j}\,[\,I\,]$ for all $j \in J$ then $Y|_{\mathcal{V}} \nrightarrow_h X|_{\sum_{j \in J} \mathcal{U}_j}\,[\,I\,]$.

An identical set of results hold for up–to–horizon–h non–causality.

Lemma 3.2 generalizes DR's Proposition 2.1 in three directions: first, it considers all subspaces along which X and Y vary where DR consider only the cartesian components; second, it considers long run non–causality where DR consider only finite horizons; third, DR require I to be conformable with $P_{\mathcal{U}}X$, which we do not . (i) and (ii) imply that if Y fails to cause X then the non–causality also exists along all linear combinations of the two vector processes; in other words, non–causality is invariant to linear transformations. (iii) and (iv) state that non–causal channels can be aggregated in any linear fashion; thus, non–causality is invariant to linear aggregation. It is crucial in Lemma 3.2 that J be arbitrary as we will require a countably infinite J to prove the existence part of Lemma 3.3.

Now in general if $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\,[\,I\,]$, the subspaces $\mathcal{U}$ and $\mathcal{V}$ may be parts of larger subspaces along which non–causality occurs. We would like to define what we mean by "the subspaces of non–causality at horizon h between X and Y." Unfortunately, the linear additivity properties in Lemma 3.2 hold only when keeping one side of the non–causality relationship fixed. So we can talk about "the subspace of $\mathbb{R}^{n_X}$ along which X fails to respond to $P_{\mathcal{V}}Y$ at horizon h" or we can talk about "the subspace of $\mathbb{R}^{n_Y}$ along which Y fails to affect $P_{\mathcal{U}}X$ at horizon h," but to leave both $\mathcal{U}$ and $\mathcal{V}$ unspecified risks running into inconsistencies. For a given $\mathcal{V}$ we could define the former to be the maximal subspace $\mathcal{U}$ along which $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\,[\,I\,]$ in the sense that

8

such a $\mathcal{U}$ is not properly contained in any other subspace along which non–causality occurs (and similarly when holding $\mathcal{U}$ fixed); however, we need to prove existence and uniqueness first.

Lemma 3.3. For $1 \leq h \leq \infty$ and subspace $\mathcal{V}$, the maximal subspace $\mathcal{U}$ along which $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\, [\,I\,]$ exists and is unique. Similarly, holding subspace $\mathcal{U}$ fixed, the maximal subspace $\mathcal{V}$ along which $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\, [\,I\,]$ also exists and is unique. The identical result holds as well for up–to–horizon–h non–causality.

To simplify notation, we will consider these maximal subspaces of non–causality either in the context of fixing $\mathcal{U} = \mathbb{R}^{n_X}$ or in the context of fixing $\mathcal{V} = \mathbb{R}^{n_Y}$. In fact, this involves no loss in generality as X and Y can always be linearly transformed to suite arbitrary $\mathcal{U}$ and $\mathcal{V}$.

Definition 3.4 (Subspace of Non–causality at Horizon h). The maximal subspace $\mathcal{U}$ such that $Y \nrightarrow_h X|_{\mathcal{U}}\, [\,I\,]$ (resp. $Y \nrightarrow_{(h)} X|_{\mathcal{U}}\, [\,I\,]$) is denoted by $\mathcal{U}_h^{XYI}$ (resp. $\mathcal{U}_{(h)}^{XYI}$); its orthogonal complement is denoted by $\mathcal{C}_h^{XYI}$ (resp. $\mathcal{C}_{(h)}^{XYI}$). We define, U_h^{XYI} (resp. $U_{(h)}^{XYI}$) to be a matrix of orthonormal columns which span $\mathcal{U}_h^{XYI}$ (resp. $\mathcal{U}_{(h)}^{XYI}$). Similarly, we define, C_h^{XYI} (resp. $C_{(h)}^{XYI}$) to be a matrix of orthonormal columns which span $\mathcal{C}_h^{XYI}$ (resp. $\mathcal{C}_{(h)}^{XYI}$).

Likewise, the maximal subspace $\mathcal{V}$ such that $Y|_{\mathcal{V}} \nrightarrow_h X\, [\,I\,]$ (resp. $Y|_{\mathcal{V}} \nrightarrow_{(h)} X\, [\,I\,]$) is denoted by $\mathcal{V}_h^{XYI}$ (resp. $\mathcal{V}_{(h)}^{XYI}$); its orthogonal complement is denoted by $\mathcal{D}_h^{XYI}$ (resp. $\mathcal{D}_{(h)}^{XYI}$). We define, V_h^{XYI} (resp. $V_{(h)}^{XYI}$) to be a matrix of orthonormal columns which span $\mathcal{V}_h^{XYI}$ (resp. $\mathcal{V}_{(h)}^{XYI}$). Finally, we define, D_h^{XYI} (resp. $D_{(h)}^{XYI}$) to be a matrix of orthonormal columns which span $\mathcal{D}_h^{XYI}$ (resp. $\mathcal{D}_{(h)}^{XYI}$). $\qquad\square$

The subspace $\mathcal{U}_h^{XYI}$ specifies along which directions variations in X at horizon h cannot be attributed to variations in Y; the subspace $\mathcal{C}_h^{XYI}$ then specifies the directions of variations in X attributable to variations in Y. Likewise, the subspace $\mathcal{V}_h^{XYI}$ specifies in what directions variations in Y produce no variations in X at horizon h; the subspace $\mathcal{D}_h^{XYI}$ then specifies the directions of variations in Y that have an effect on X. The columns of U_h^{XYI} are the linear combinations of the X's that are unaffected by Y at horizon h, while the columns of C_h^{XYI} are the linear combinations of the X's that are affected by Y. Likewise, the columns of V_h^{XYI} are the linear combinations of the Y's that have no effect on X, while the columns of D_h^{XYI} are the linear combinations of the Y's that have an effect on X. Note that these and the other matrices listed in Definition 3.4 are unique modulo left multiplication by orthogonal matrices.

The following proposition lists some additional useful properties of the above subspaces.

Proposition 3.1. Under Assumption 1, information set I, and $1 \leq h \leq \infty$,

(i) $\mathcal{U}_h^{XYI} = \sum_{\{\mathcal{U}:Y \nrightarrow_h X|_\mathcal{U} \ [I]\}} \mathcal{U}.$ (vii) $\mathcal{V}_h^{XYI} = \sum_{\{\mathcal{V}:Y \nrightarrow_h X|_\mathcal{V} \ [I]\}} \mathcal{V}.$

(ii) $\mathcal{U}_{(h)}^{XYI} = \sum_{\{\mathcal{U}:Y \nrightarrow_{(h)} X|_\mathcal{U} \ [I]\}} \mathcal{U}.$ (viii) $\mathcal{V}_{(h)}^{XYI} = \sum_{\{\mathcal{V}:Y \nrightarrow_{(h)} X \ [I]\}} \mathcal{V}.$

(iii) $\mathcal{U}_{(h)}^{XYI} = \bigcap_{j=1}^h \mathcal{U}_j^{XYI}.$ (ix) $\mathcal{V}_{(h)}^{XYI} = \bigcap_{j=1}^h \mathcal{V}_j^{XYI}.$

(iv) $\mathcal{U}_{(\infty)}^{XYI} \subseteq \mathcal{U}_\infty^{XYI}.$ (x) $\mathcal{V}_{(\infty)}^{XYI} \subseteq \mathcal{V}_\infty^{XYI}.$

(v) $\sum_{\{1 \leq j \leq h\}} \mathcal{C}_j^{XYI} = \mathcal{C}_{(h)}^{XYI}.$ (xi) $\sum_{\{1 \leq j \leq h\}} \mathcal{D}_j^{XYI} = \mathcal{D}_{(h)}^{XYI}.$

(vi) $\mathcal{C}_{(h)}^{XYI} \subseteq \mathcal{C}_{(h+1)}^{XYI}.$ (xii) $\mathcal{D}_{(h)}^{XYI} \subseteq \mathcal{D}_{(h+1)}^{XYI}.$

We will discuss only (i) – (vi) as similar, if not identical, observations can be made about (vii) – (xii). It follows from (i) (resp. (ii)) that there exists no subspace $\mathcal{W} \subseteq \mathcal{C}_h^{XYI}$ (resp. $\mathcal{W} \subseteq \mathcal{C}_{(h)}^{XYI}$) such that $Y \nrightarrow_h X|_\mathcal{W} \ [I]$ (resp. $Y \nrightarrow_{(h)} X|_\mathcal{W} \ [I]$). In other words, as far as Y is concerned $\mathcal{U}_h^{XYI}$ (resp. $\mathcal{U}_{(h)}^{XYI}$) accounts for all non–causal directions at (resp. up to) horizon h. This does not imply that there are no impediments to variations along $\mathcal{C}_h^{XYI}$ (resp. $\mathcal{C}_{(h)}^{XYI}$) as there may be non–linear ways of combining the X variables that make Y useless for prediction over and above I. This suggests, thinking of $\mathcal{C}_h^{XYI}$ (resp. $\mathcal{C}_{(h)}^{XYI}$) as the space reachable by X at (resp. up to) horizon h for suitable variations in Y when controlling for I; we discuss the relationship between reachability and causality in greater detail in section 6. (iii) and (iv) are trivial applications of Definitions 3.3 and 3.4. (v) says that what is reachable up to horizon h is reachable at some horizon between 1 and h. Finally, (vi) says that the reachable subspace grows across horizons.

Finally, we close this section with a discussion of the causal effects of a series on itself. Because nothing in our construction so far depends on X and Y being distinct, it is perfectly consistent to have $Y = X$ and so the causal properties of X on its future values is well defined. We will be particularly interested in this section in the long run effect of a series on itself. If the long run behavior of a series depends on its history at a particular point, any disturbances in its history never dissipate and the causal effects of this history are permanent. If on the other hand, the long run behavior of the series is independent of all its histories, the process is in a sense stable. This suggests the following notion of stability.

Definition 3.5 (L^2 Stability). Under Assumption 1, define $H_\omega(X) = \bigcap_{t > \omega} X(\omega, t]$ and $\mathcal{M}_\infty^X = \mathcal{U}_\infty^{XXH_\omega(X)}$. We say that X is L^2 stable if $\mathcal{M}_\infty^X = \mathbb{R}^{n_X}$, L^2 unstable if $\mathcal{M}_\infty^X = \{0\}$, and cotrending if $\{0\} \neq \mathcal{M}_\infty^X \neq \mathbb{R}^{n_X}$. The subspace $\mathcal{M}_\infty^X$ is referred to as the subspace of L^2

stability of X. Clearly, X is L^2 stable along any subspace $\mathcal{M} \subseteq \mathcal{M}_\infty^X$ and $\mathcal{M}_\infty^X$ is the maximal subspace along which X is L^2 stable. □

In general $H_\omega(X)$ consists of all the uncertainty surrounding X that is resolved at the "start" of the process; typically this consists of non–random trends, random initial conditions, or trends which depend on a random component that is constant through time. Definition 3.5 says that an L^2 process X is L^2 stable along some subspace if and only if its forecasts along that subspace revert to the "mean" in the L^2 norm in the long run. To illustrate what we mean by the "mean" suppose we have a second order stationary process X; if the deterministic component of its Wold decomposition (see e.g. Brockwell & Davis (1991), p. 187) is constant then $H_\omega(X) = \overline{\mathrm{sp}}\{1\}$ and so its mean is simply $\mathbb{E}(X(t))$; if instead the deterministic component is an L^2 random variable ξ then $H_\omega(X) = \overline{\mathrm{sp}}\{\xi\}$ and the mean is $P(X(t)|\overline{\mathrm{sp}}\{\xi\})$. Note that the Wold decomposition also shows that every second–order stationary process is L^2 stable.

Now it is clear that if any linear combination of X is long–run–caused by any other linear combination of X with respect to $H_\omega(X)$ then X cannot be L^2 stable. We may now decompose any L^2 process X uniquely into an L^2 stable process, $P_{\mathcal{M}_\infty^X} X$ and an L^2 unstable process, $(I_{n_X} - P_{\mathcal{M}_\infty^X})X$. If X is cotrending then neither component will be zero; $(C_\infty^{XXH_\omega(X)})'X$ can then be interpreted as common trends while $U_\infty^{XXH_\omega(X)}$ may be interpreted as equilibrium relationships between the X variables.[6]

Now Granger (1988b) shows that in a cointegrated bivariate model, at least one of the variables must cause the other. The generalization to multivariate processes in L^2 is that if X is cotrending at least one of its components must cause another of its components in the long run.

Theorem 3.1 (Long run Subspace Causality in Cotrending Time Series)**.** Under Assumption 1, if X is cotrending then there exists subspaces $\mathcal{M}_1 \subseteq \mathbb{R}^{n_X}$ and $\mathcal{M}_2 \subseteq (\mathcal{M}_\infty^X)^\perp$ such that $X|_{\mathcal{M}_1} \nrightarrow_\infty X|_{\mathcal{M}_2} [H_\omega(X)]$ fails to hold.

4 Subspace Causality in Linear Invertible Processes

We now change our notation slightly to suite the analysis of linear processes.

---

[6]Cotrending processes are defined analogously to cointegrating processes; in fact the concept of cointegration is subsumed by cotrendedness as we will see in greater detail in section 5.

Assumption 2. $W = \{W(t) = (X'(t), Y'(t), Z'(t))' : t \in \mathbb{Z}\}$ is a stochastic processes in L^2 of dimension n; the dimensions of the components X, Y, and Z are n_X, n_Y, and n_Z respectively. W has the autoregressive representation,

$$W(t) = \mu(t) + \sum_{j=1}^{\infty} \pi_j W(t-j) + a(t), \qquad t > \varpi, \tag{4.1}$$

$\mu(t) \in H_{-\infty}(W) = \bigcap_{t \in \mathbb{Z}} W(-\infty, t]$ for all $t > \varpi$. $\{a(t) : t > \varpi\}$ is a sequence of uncorrelated random vectors in L^2, with $\mathbb{E}(a(t)) = 0$ and $\mathbb{E}(a(t)a'(t)) = \Omega(t) > 0$ for all $t > \varpi$. Moreover $a(t)$ is uncorrelated with $W(-\infty, t-1]$ for all $t > \varpi$. The innovations process is partitioned conformably with W as, $a = (a_X', a_Y', a_Z')'$. We also assume that $\sum_{j=1}^{\infty} \pi_j W(t-j)$ converges in L^2 for all $t > \varpi$. If $\varpi = \omega = -\infty$, W has an autoregressive representation (4.1) for all $t \in \mathbb{Z}$; on the other hand, if $\varpi \in \mathbb{Z}$ we set $W(t)$ for $t \leq \varpi$ to any sequence of initial random vectors in $H_{-\infty}(W)$ that will guarantee convergence of (4.1); thus the process is assumed to start after time ϖ and all uncertainty in $H_{-\infty}(W)$ is resolved at time ϖ. We will be concerned with the following information sets:

(i) Causal channels between X and Y. Here we will assume that the subspaces, $\mathcal{U} \subseteq \mathbb{R}^{n_X}$ and $\mathcal{V} \subseteq \mathbb{R}^{n_Y}$ are given along with the information set, $I(t) = H_{-\infty}(W) + X(-\infty, t] + P_{\mathcal{V}^\perp} Y(-\infty, t] + Z(-\infty, t]$ for $t \in \mathbb{Z}$, which consists of all available information at time $t \in \mathbb{Z}$ excluding the contribution of variations in Y along the given $\mathcal{V}$; it may also be written as $I(t) = H_{-\infty}(W) + (W(-\infty, t] - P_{\mathcal{V}} Y(\varpi, t])$ for $t \in \mathbb{Z}$.[7]

(ii) Causal channels between W and itself. Here we will assume that the subspaces $\mathcal{U}, \mathcal{V} \subseteq \mathbb{R}^n$ are given and work with the information set $I(t) = H_{-\infty}(W) + P_{\mathcal{V}^\perp} W(-\infty, t]$ for $t \in \mathbb{Z}$. Thus $I(t)$ includes all available information excluding the variation of W along $\mathcal{V}$; it may also be written as $I(t) = H_{-\infty}(W) + (W(-\infty, t] - P_{\mathcal{V}} W(\varpi, t])$ for $t \in \mathbb{Z}$.

Finally, it will be convenient to consider the demeaned process of W, which we denote by $\widehat{W} = \{\widehat{W}(t) = W(t) - P(W(t)|H_{-\infty}(W)) : t \in \mathbb{Z}\}$. This will allow us to simplify the notation by eliminating $\mu(t)$ from equation (4.1),

$$\widehat{W}(t) = \begin{cases} \sum_{j=1}^{t-\varpi} \pi_j \widehat{W}(t-j) + a(t), & \text{for } t > \varpi, \\ 0, & \text{for } t \leq \varpi, \end{cases} \tag{4.2}$$

---

[7]Because the process the process (4.1) includes the deterministic term $\mu(t) \in H_{-\infty}(W)$ for $t > \varpi$, we are forced to include $H_{-\infty}(W)$ into the information set. We do this in the interest of maintaining continuity with previous literature despite the fact that excluding μ (i.e. setting $H_{-\infty}(W) = \{0\}$) makes for much more elegant theory.

Note that if $\overline{\mathrm{sp}}\{1\} \subseteq H_{-\infty}(W)$, then $\mathbb{E}\widehat{W}(t) = 0$ for all $t \in \mathbb{Z}$. The demeaned process is partitioned conformably with W as $\widehat{W} = (\widehat{X}', \widehat{Y}', \widehat{Z}')'$. $\qquad\qquad\square$

The class of processes in Assumption 2 includes invertible VARMA (see e.g. Lütkepohl (2006)) and long–memory processes (see e.g. section 13.2 of Brockwell & Davis (1991)); lemma 6.4 of Pourahmadi (2001) provides a full characterization of the stationary class of processes (4.1). The difference between this formulation and the class of processes considered by DR is that we require $\Omega(t)$ to be positive definite.

The working paper version of DR (Dufour & Renault, 1995) shows that under Assumption 2, the h–period forecasts of W are of the form,

$$P(W(t+h)|W(-\infty,t]) = \sum_{k=0}^{h-1} \pi_1^{(k)} \mu(t+h-k) + \sum_{j=1}^{\infty} \pi_j^{(h)} W(t+1-j), \qquad t > \varpi, \quad h \geq 1,$$

where the coefficients are defined by,

$$\pi_j^{(1)} = \pi_j, \qquad \pi_j^{(h+1)} = \pi_{j+h} + \sum_{l=1}^{h} \pi_{h-l+1} \pi_j^{(l)}, \qquad\qquad j,h \geq 1 \qquad (4.3)$$

$$= \pi_{j+1}^{(h)} + \pi_1^{(h)} \pi_j, \qquad\qquad j,h \geq 1 \qquad (4.4)$$

Equation (4.3) follows from direct substitution, while equation (4.4) is easily obtained from the VAR(1) representation of W.

Definition 4.1 (Projection Matrices and Impulse Responses). The matrices $\{\pi_j^{(h)}\}_{j=1}^{\infty}$ are termed the projection matrices at horizon h. If we set $\pi^{(h)}(z) = \sum_{j=1}^{\infty} \pi_j^{(h)} z^j$, with $\pi(z) = \pi^{(1)}(z)$, then the impulse response operator is defined by, $I_n + \psi(w) = (I_n - \pi(w))^{-1}$, where $\psi(w) = \sum_{h=1}^{\infty} \psi_h w^h$. $\qquad\qquad\square$

Dufour & Renault (1995) demonstrate that the impulse response operator $\psi(z)$ is retrievable from the projection matrices at horizon h via the formula,

$$\psi(w) = \sum_{j=1}^{\infty} \pi_1^{(h)} w^h, \qquad\qquad (4.5)$$

Assumption 3. The projection matrices are partitioned conformably with W as,

$$\pi_j^{(h)} = \begin{bmatrix} \pi_{XXj}^{(h)} & \pi_{XYj}^{(h)} & \pi_{XZj}^{(h)} \\ \pi_{YXj}^{(h)} & \pi_{YYj}^{(h)} & \pi_{YZj}^{(h)} \\ \pi_{ZXj}^{(h)} & \pi_{ZYj}^{(h)} & \pi_{ZZj}^{(h)} \end{bmatrix},$$

for all $j, h \geq 1$. The projection matrix operators $\pi^{(h)}(z)$ are partitioned similarly. $\qquad\square$

13

Given Assumptions 2 and 3, the projection variation for the effect of Y on X is given by,

$$\Delta_h^{P_{\mathcal{U}}XP_{\mathcal{V}}YI}(t) = \begin{cases} \sum_{j=1}^{t-\varpi} P_{\mathcal{U}}\pi_{XYj}^{(h)}P_{\mathcal{V}}\{Y(t+1-j) - P(Y(t+1-j)|I(t))\}, & t > \varpi \\ 0, & t \leq \varpi \end{cases} \quad (4.6)$$

Equation (4.6) makes clear that the existence of causal channels between X and Y will hinge on the properties of the matrices $\{P_{\mathcal{U}}\pi_{XYj}^{(h)}P_{\mathcal{V}}\}_{h,j\geq 1}$.

Theorem 4.1 (Characterization of Subspace Non-causality at Horizon $h < \infty$). Under Assumptions 2 and 3 and for $1 \leq h < \infty$, $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}} [I]$ if and only if, $P_{\mathcal{U}}\pi_{XYj}^{(h)}P_{\mathcal{V}} = 0$ for all $j \geq 1$.

Theorem 4.1 states that the generalization from cartesian non–causality to subspace non–causality involves nothing more than linear restrictions on the projection matrices $\{\pi_{XYj}^{(h)}\}_{j=1}^{\infty}$. When $\mathcal{U}$ and $\mathcal{V}$ are known, we simply test the restrictions,

$$U'\pi_{XYj}^{(h)}V = 0, \quad \text{for all} \quad j \geq 1, \quad (4.7)$$

where U and V are as in Lemma 3.1. If one of them is unknown – recall that we must specify at least one them – then we have a reduced rank regression *à la* Anderson (1951) and (4.7) can be imposed as a rank restriction. The case where we are interested in finding $\mathcal{V}_1^{XYI}$ by imposing rank restrictions of the form $\pi_{XYj}V = 0$ for all $j \geq 1$ can be seen as a variant of the problem considered by Sargent & Sims (1977), which is concerned with finding indices summarizing the information of a large set of variables Y; in this case, the indices are exactly $(D_1^{XYI})'Y$.

Now because of the linearity of the process, the subspaces of (non)causality are easily characterized in terms of the projection matrices as we see in the following corollary.

Corollary 4.1. Under Assumptions 2 and 3 and for $1 \leq h < \infty$,

    (i) $\mathcal{U}_h^{XYI} = \bigcap_{\{j\geq 1\}} \ker(\pi_{XYj}^{(h)}{}')$, for $h < \infty$.   (iii) $\mathcal{V}_h^{XYI} = \bigcap_{\{j\geq 1\}} \ker(\pi_{XYj}^{(h)})$, for $h < \infty$.

    (ii) $\mathcal{C}_h^{XYI} = \sum_{\{j\geq 1\}} \text{im}(\pi_{XYj}^{(h)})$, for $h < \infty$.   (iv) $\mathcal{D}_h^{XYI} = \sum_{\{j\geq 1\}} \text{im}(\pi_{XYj}^{(h)}{}')$, for $h < \infty$.

Long run non–causality is more subtle to deal with than its finite horizon counterpart. Assumptions 2 and 3 allow us to obtain necessary conditions for long run non–causality but sufficiency requires stronger assumptions.

Theorem 4.2 (Characterization of Long Run Subspace Non–causality). Under Assumptions 2 and 3, $Y|_{\mathcal{V}} \nrightarrow_{\infty} X|_{\mathcal{U}} [I]$ implies that $\lim_{h\to\infty} P_{\mathcal{U}}\pi_{XYj}^{(h)}P_{\mathcal{V}} = 0$ for all $j \geq 1$. Conversely, if

14

$\lim_{h\to\infty} \sum_{j=1}^{t-\varpi} \|P_{\mathcal{U}} \pi_{XYj}^{(h)} P_{\mathcal{V}}\| = 0$ and $\sup_{\varpi < s \leq t} \mathbb{E}\|P_{\mathcal{V}}\widehat{Y}(s)\|^2 < \infty$ for all $t \in \mathbb{Z}$ then $Y|_{\mathcal{V}} \nrightarrow_{\infty} X|_{\mathcal{U}}\,[\,I\,]$.

Thus when $\varpi \in \mathbb{Z}$, $Y|_{\mathcal{V}} \nrightarrow_{\infty} X|_{\mathcal{U}}\,[\,I\,]$ if and only if $\lim_{h\to\infty} P_{\mathcal{U}} \pi_{XYj}^{(h)} P_{\mathcal{V}} = 0$ for all $j \geq 1$. However, when $\varpi = -\infty$ stronger conditions are required; $P_{\mathcal{U}} \pi_{XYj}^{(h)} P_{\mathcal{V}}$ must converge to zero uniformly and the demeaned series $\{P_{\mathcal{V}}\widehat{Y}(s)\}_{\varpi < s \leq t}$ must be bounded in L^2.[8] Fortunately, however, the sufficiency conditions in Theorem 4.2 will be satisfied by most processes of interest as we will see below.[9] Note that $Y|_{\mathcal{V}} \nrightarrow_{\infty} X|_{\mathcal{U}}\,[\,I\,]$ implies that $\lim_{h\to\infty} P_{\mathcal{U}} \psi_{XYh} P_{\mathcal{V}} = 0$, thus the impulse response of X along $\mathcal{U}$ with respect to Y along $\mathcal{V}$ must diminish through time; however, this will not be sufficient for long run non–causality as we will see in the case of VARMA processes (see the proof of Theorem 5.1 (ii)).

Now equation (4.4) implies that,

$$P_{\mathcal{U}} \pi_{XYj}^{(h+1)} P_{\mathcal{V}} = P_{\mathcal{U}} \pi_{XYj+1}^{(h)} P_{\mathcal{V}} + P_{\mathcal{U}} \pi_{XX1}^{(h)} P_{\mathcal{U}} \pi_{XYj} P_{\mathcal{V}} + P_{\mathcal{U}} \pi_{XX1}^{(h)} P_{\mathcal{U}^\perp} \pi_{XYj} P_{\mathcal{V}} +$$
$$P_{\mathcal{U}} \pi_{XY1}^{(h)} P_{\mathcal{V}} \pi_{YYj} P_{\mathcal{V}} + P_{\mathcal{U}} \pi_{XY1}^{(h)} P_{\mathcal{V}^\perp} \pi_{YYj} P_{\mathcal{V}} + P_{\mathcal{U}} \pi_{XZ1}^{(h)} \pi_{ZYj} P_{\mathcal{V}},$$

and so we have that if $Y|_{\mathcal{V}} \nrightarrow_{(h)} X|_{\mathcal{U}}\,[\,I\,]$ with $h < \infty$ then $P_{\mathcal{U}} \pi_{XYj}^{(h+1)} P_{\mathcal{V}} = P_{\mathcal{U}} \pi_{XX1}^{(h)} P_{\mathcal{U}^\perp} \pi_{XYj} P_{\mathcal{V}} + P_{\mathcal{U}} \pi_{XY1}^{(h)} P_{\mathcal{V}^\perp} \pi_{YYj} P_{\mathcal{V}} + P_{\mathcal{U}} \pi_{XZ1}^{(h)} \pi_{ZYj} P_{\mathcal{V}}$. That is, if Y along $\mathcal{V}$ fails to cause X along $\mathcal{U}$ up to horizon h, it may still have an effect at horizon $h + 1$ through one of three indirect causal channels: either through X itself if the direct effect of $P_{\mathcal{V}}Y$ on $P_{\mathcal{U}^\perp}X$ persists for h periods, through Y if the direct effect of $P_{\mathcal{V}}Y$ on $P_{\mathcal{V}^\perp}Y$ persists for h periods, or through Z if Z causes X along $\mathcal{U}$ at horizon h. Following this line of reasons allows us to prove the following theorem characterizing subspace non–causality up to horizon h.

Theorem 4.3 (Characterization of Subspace Non-causality up to Horizon h)**.** Under Assumptions 2 and 3 and for $h \geq 2$,

(i) $P_{\mathcal{U}} \pi_{XYj} P_{\mathcal{V}} = 0$ for all $j \geq 1$.

---

[8]The reason why we require stronger assumptions is evident from equation (4.6) once we recall that convergence in probability does not imply convergence in L^p and one must resort to assumptions of dominance, monotonicity, or uniform integrability to obtain L^p convergence (see e.g. section 4.5 of Chung (1974)).

[9]It is perhaps worth mentioning that the so–called "long run effect" of Y on X in the dynamic multiplier literature (see e.g. Lütkepohl (2006) p. 392) bears no relation to long run causality. Long run causality, as we have seen in the last section, cannot occur in L^2 stable processes, thus they cannot occur in stationary processes; on the other hand, the "long run effect" of Y on X is $\pi_{XY}(1)$, which may be non–zero in a stationary process.

(ii) $P_\mathcal{U}\pi_{XX1}^{(k)}P_{\mathcal{U}^\perp}\pi_{XYj}P_\mathcal{V} + P_\mathcal{U}\pi_{XY1}^{(k)}P_{\mathcal{V}^\perp}\pi_{YYj}P_\mathcal{V} + P_\mathcal{U}\pi_{XZ1}^{(k)}\pi_{ZYj}P_\mathcal{V} = 0$ for all $j \geq 1$, $1 \leq k < h$.

are necessary and sufficient for $Y|_\mathcal{V} \nrightarrow_{(h)} X|_\mathcal{U} [I]$.

Theorems 4.1 and 4.3 are extensions of results by DR. However, a judicious choice of coordinates and information set yields that they are in fact equivalent to DR's results; this is shown in section 8.1. In fact, a much more general formulation of Theorems 4.1–4.3 is possible as we summarize in the following theorem. The results are given without proof as they involves minimal modification of the proofs of DR along the same lines as the discussion in section 8.1.

Theorem 4.4 (The Causal Effects of a Series on Itself). Under Assumption 2 for $1 \leq h < \infty$,

(i) $W|_\mathcal{V} \nrightarrow_h W|_\mathcal{U} [I]$ if and only if, $P_\mathcal{U}\pi_j^{(h)}P_\mathcal{V} = 0$ for all $j \geq 1$.

(ii) $W|_\mathcal{V} \nrightarrow_\infty W|_\mathcal{U} [I]$ implies that $\lim_{h\to\infty} P_\mathcal{U}\pi_j^{(h)}P_\mathcal{V} = 0$ for all $j \geq 1$. Conversely, $W|_\mathcal{V} \nrightarrow_\infty$ $W|_\mathcal{U} [I]$ if $\lim_{h\to\infty} \sum_{j=1}^{t-\varpi} \|P_\mathcal{U}\pi_j^{(h)}P_\mathcal{V}\| = 0$ and $\sup_{\varpi < s \leq t} \mathbb{E}\|P_\mathcal{V}\widehat{W}(s)\|^2 < \infty$ for all $t \in \mathbb{Z}$.

(iii) The necessary and sufficient conditions for $W|_\mathcal{V} \nrightarrow_{(h)} W|_\mathcal{U} [I]$ are,

    (a) $P_\mathcal{U}\pi_j P_\mathcal{V} = 0$ for all $j \geq 1$.

    (b) $P_\mathcal{U}\pi_1^{(k)}P_{\mathcal{U}^\perp}\pi_j P_\mathcal{V} = 0$ for all $j \geq 1$, $1 \leq k < h$.

These results reduce to the earlier theorems by making the following substitutions,

$$\mathcal{U} \to \mathcal{U} \times \overbrace{\{0\} \times \cdots \times \{0\}}^{n_X + n_Z} \qquad \mathcal{V} \to \overbrace{\{0\} \times \cdots \times \{0\}}^{n_X} \times \mathcal{V} \times \overbrace{\{0\} \times \cdots \times \{0\}}^{n_Z}$$

The case $h = 1$ in Theorem 4.4(i) has been studied by Box & Tiao (1977) and Velu et al. (1986) in the context of stationary VARs for the purpose of model reduction and improving forecasts at horizon $h = 1$; here $(C_1^{WWI})'W$ is predictable by current and past values of W but $(U_1^{WWI})'W$ is not. The other results are straightforward generalizations following the same line of logic as before. The following corollary is immediate.

Corollary 4.2. Under Assumption 2 and for $1 \leq h < \infty$,

    (i) $\mathcal{U}_h^{WWI} = \bigcap_{\{j\geq 1\}} \ker(\pi_j^{(h)'})$, for $h < \infty$. (iii) $\mathcal{V}_h^{WWI} = \bigcap_{\{j\geq 1\}} \ker(\pi_j^{(h)})$, for $h < \infty$.

    (ii) $\mathcal{C}_h^{WWI} = \sum_{\{j\geq 1\}} \mathrm{im}(\pi_j^{(h)})$, for $h < \infty$. (iv) $\mathcal{D}_h^{WWI} = \sum_{\{j\geq 1\}} \mathrm{im}(\pi_j^{(h)'})$, for $h < \infty$.

5   Subspace Causality in VARMA Processes

To operationalize the theory of the last section, we must simplify the structure of the projection operator $\pi(z)$. One way to do this is to assume that $\pi(z)$ is rational; from linear system theory (see e.g. Sontag (1998)), this implies that the projection matrices are recursively and finitely generated, what will allow us to find truncation rules useful for empirical testing of non–causality along subspaces.

Assumption 4. $I_n - \pi(z) = \theta^{-1}(z)\phi(z)$, where $\phi(z) = I_n - \sum_{j=1}^p \phi_j z^j$ and $\theta(z) = I_n + \sum_{j=1}^q \theta_j z^j$ are assumed identified. We also assume that $\Omega(t) = \Omega$ for all $t > \varpi$. $\qquad\square$

Under Assumptions 2 and 4, $\widehat{W}$ is a VARMA process and the zeros of $\theta(z)$ lie outside the unit circle. We may now state the following truncation theorems.

Theorem 5.1 (Truncation Rules for Subspace Non–causality in VARMA Processes)**.** Under Assumptions 2-4 and for $1 \le h < \infty$,

(i) $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}} \, [\,I\,]$ if and only if $P_{\mathcal{U}}\pi_{XYj}^{(h)}P_{\mathcal{V}} = 0$ for all $1 \le j \le p + (n-1)q$.

(ii) $Y|_{\mathcal{V}} \nrightarrow_\infty X|_{\mathcal{U}} \, [\,I\,]$ if and only if $\lim_{h\to\infty} P_{\mathcal{U}}\pi_{XYj}^{(h)}P_{\mathcal{V}} = 0$ for all $1 \le j \le (n - \dim(\mathcal{V}))nq + \dim(\mathcal{V})(p + (n-1)q)$.

(iii) $Y|_{\mathcal{V}} \nrightarrow_{(\infty)} X|_{\mathcal{U}} \, [\,I\,]$ if and only if $P_{\mathcal{U}}\pi_{XYj}^{(h)}P_{\mathcal{V}} = 0$ for all $1 \le j \le p + (n-1)q$ and $1 \le h \le (p + (n-1)q)(n - \dim(\mathcal{U}) - \dim(\mathcal{V})) + 1$.

Theorem 5.1 can be used for empirical tests of subspace non–causality in VARMA processes. (i) and (iii) reduce to DR's Proposition 4.5 when $\mathcal{U} = \mathbb{R}^{n_X}$, $\mathcal{V} = \mathbb{R}^{n_Y}$, and $q = 0$ – i.e. when considering cartesian non–causality in VARs. (ii) specializes Theorem 4.2 to the VARMA case; note that it is equivalent to $\lim_{h\to\infty} P_{\mathcal{U}}\pi_{XYj}^{(h)}P_{\mathcal{V}} = 0$ for all $j \ge 1$.

The next result specializes Corollary 4.1 as well as Proposition 3.1 (v) and (xi) to VARMA's.

Corollary 5.1. Under Assumptions 2-4, for $1 \le h < \infty$ and $m_1 = n_X nq + n_Y(p + (n-1)q)$ and $m_2 = n_X np + n_Y((n-1)p + q)$,

(i) $\mathcal{U}_h^{XYI} = \bigcap_{\{1 \le j \le m_1\}} \ker(\pi_{XYj}^{(h)}{}')$.   (iv) $\mathcal{V}_h^{XYI} = \bigcap_{\{1 \le j \le m_1\}} \ker(\pi_{XYj}^{(h)})$, $h < \infty$.

(ii) $\mathcal{C}_h^{XYI} = \sum_{\{1 \le j \le m_1\}} \mathrm{im}(\pi_{XYj}^{(h)})$.   (v) $\mathcal{D}_h^{XYI} = \sum_{\{1 \le j \le m_1\}} \mathrm{im}(\pi_{XYj}^{(h)}{}')$, $h < \infty$.

(iii) $\mathcal{C}_{(\infty)}^{XYI} = \mathcal{C}_{(m_2)}^{XYI}$.   (vi) $\mathcal{D}_{(\infty)}^{XYI} = \mathcal{D}_{(m_2)}^{XYI}$.

It follows from Corollary 5.1 that in the context of VARMA processes we do not need an infinite number of projection matrices in order to construct the subspaces of (non)causality.

17

For example, $[C_h^{XYI} \, U_h^{XYI}]$ can be obtained as the orthogonal matrix in the QR decomposition of $[\pi_{XY1}^{(h)} \, \pi_{XY2}^{(h)} \, \cdots \, \pi_{XYm_1}^{(h)}]$; a similar construction gives us $[D_h^{XYI} \, V_h^{XYI}]$.

There are also truncation rules for the effect of W on itself. We summarize them in the next two result, which are – again – given without proof.

Theorem 5.2 (Truncation Rules for Subspace Non–causality in VARMA Processes)**.** Under Assumptions 2-4 and for $1 \leq h < \infty$,

(i) $W|_{\mathcal{V}} \nrightarrow_h W|_{\mathcal{U}} \, [\, I \,]$ if and only if $P_{\mathcal{U}} \pi_j^{(h)} P_{\mathcal{V}} = 0$ for all $1 \leq j \leq p + (n-1)q$.

(ii) $W|_{\mathcal{V}} \nrightarrow_{\infty} W|_{\mathcal{U}} \, [\, I \,]$ if and only if $\lim_{h \to \infty} P_{\mathcal{U}} \pi_j^{(h)} P_{\mathcal{V}} = 0$ for all $1 \leq j \leq (n - \dim(\mathcal{V}))nq + \dim(\mathcal{V})(p + (n-1)q)$.

(iii) $W|_{\mathcal{V}} \nrightarrow_{(\infty)} W|_{\mathcal{U}} \, [\, I \,]$ if and only if $P_{\mathcal{U}} \pi_j^{(h)} P_{\mathcal{V}} = 0$ for all $1 \leq j \leq p + (n-1)q$ and $1 \leq h \leq (p + (n-1)q)(n - \dim(\mathcal{U}) - \dim(\mathcal{V})) + 1$.

Corollary 5.2. Under Assumptions 2-4, for $1 \leq h < \infty$ and $m_1 = n(2nq + p + -q)$ and $m_2 = n(2np - p + q)$,

(i) $\mathcal{U}_h^{WWI} = \bigcap_{\{1 \leq j \leq m_1\}} \ker(\pi_j^{(h)\prime})$. (iv) $\mathcal{V}_h^{WWI} = \bigcap_{\{1 \leq j \leq m_1\}} \ker(\pi_j^{(h)})$, $h < \infty$.

(ii) $\mathcal{C}_h^{WWI} = \sum_{\{1 \leq j \leq m_1\}} \mathrm{im}(\pi_j^{(h)})$. (v) $\mathcal{D}_h^{WWI} = \sum_{\{1 \leq j \leq m_1\}} \mathrm{im}(\pi_j^{(h)\prime})$, $h < \infty$.

(iii) $\mathcal{C}_{(\infty)}^{WWI} = \mathcal{C}_{(m_2)}^{WWI}$. (vi) $\mathcal{D}_{(\infty)}^{WWI} = \mathcal{D}_{(m_2)}^{WWI}$.

Now a linearly transformed VARMA is itself a VARMA (see Lütkepohl (1984)); thus the projection of $\widehat{W}$ onto any non–zero subspace is itself a VARMA, which may be described as being stable or unstable depending on whether the roots of its autoregressive part lie outside the unit circle or not (see e.g. Lütkepohl (2006)). The next result proves that VARMA stability is equivalent to L^2 stability; in particular, $\mathcal{M}_\infty^W$ is the maximal subspace $\mathcal{M}$ such that $P_{\mathcal{M}} W$ is a stable VARMA.

Theorem 5.3 (Stability and Long run Subspace Non–causality in VARMA Processes)**.** Under Assumptions 2 and 4, with subspace $\mathcal{M} \subseteq \mathbb{R}^n$, the following are equivalent,

(i) $P_{\mathcal{M}} \widehat{W}$ is stable.

(ii) $W \nrightarrow_{\infty} W|_{\mathcal{M}} \, [\, J \,]$ for all information sets J satisfying $J(t) \subseteq W(-\infty, t]$ for all $t \in \mathbb{Z}$.

(iii) $\mathcal{M} \subseteq \mathcal{M}_\infty^W$.

The subspace $\mathcal{M}_\infty^W$ is easily derived from the projection matrices $\{\pi_j^{(h)}\}_{j,h \geq 1}$. By Theorem 5.2 (ii), $\mathcal{M}_\infty^W$ is the maximal subspace $\mathcal{M}$ satisfying $\lim_{h \to \infty} P_{\mathcal{M}} \pi_j^{(h)} = 0$ for all $j \geq 1$. As is

shown in the proof of Theorem 5.3 (ii), $\pi_j^{(h)}$ consists of terms of the form $h^k j^l \lambda^h \nu^j$, where λ and ν are eigenvalues associated with the state–space representation of the projection matrices and k and l are integers. It follows that $P_{\mathcal{M}_\infty^W}$ is precisely the projection matrix that annihilates all terms with $|\lambda| \geq 1$.

Clearly the cotrending space is the cointegration space for a cointegrated VARMA; however, the concept of a cotrending space includes the cointegration space as a special case because it is defined for any L^2 process including periodic or explosive VARMA processes. For a cointegrated VARMA, the decomposition $W = P_{\mathcal{M}_\infty^W} W + (I_n - P_{\mathcal{M}_\infty^W}) W$ is similar but not equal to the Beveridge & Nelson (1981) decomposition as the former is a geometric decomposition whereas the latter is an algebraic decomposition – see e.g. the proof of Theorem 4.2 of Johansen (1995) which starts off with a geometric decomposition but later shifts some of the stable processes back into the stable part of the decomposition.

6 Subspace Causality and Controllability

Controllability in the linear systems literature refers to the ability of the policymaker to hit any given target from any initial condition of the dynamic system. This issue arises in many important contexts of relevance to time series: linear systems (Kailath, 1980), Kalman filters (Anderson & Moore, 1979), and linear quadratic control (Bertsekas, 2001) among others and it has been variously considered in the economics literature as well; Pitchford & Turnovsky (1976) and Preston & Pagan (1982) is some of the earliest work on controllability in the context of the pure theory of policymaking; Hansen & Sargent (2005) provides a more recent consideration of controllability in the context of linear dynamic economic models.

Now consider the model most commonly encountered in the literature.

Assumption 5. Let $Y = \{Y(t) \in \mathbb{R}^{n_Y} : t \geq 0\} \subset L^2$ consist of policy variables which are chosen by the policymaker. Let $X = \{X(t) \in \mathbb{R}^{n_X} : t \geq 0\}$ consist of target variables of interest which evolve according to,

$$Z(t) = AZ(t-1) + BY(t-1) + \varepsilon(t), \qquad t > 0 \tag{6.1}$$

$$X(t) = CZ(t) + \eta(t) \tag{6.2}$$

We assume that $\xi = \{\xi(t) = (\varepsilon'(t), \eta'(t))' : t \geq 0\} \subset L^2$ is a white noise process consisting of

unobserved shocks to the system. Z is an n_Z–dimensional vector processes describing system–wide dynamics, of which we observe only partial information through X; we assume that $Z(0) \in L^2$ and $\mathbb{E}(Z(0)) = 0$. The previous assumptions imply that X is in L^2. The purpose of the policymaker is to choose the sequence of Y's to pursue some objective, whatever it may be. Note that the trajectory of X is given by,

$$X(t) = CA^t Z(0) + \sum_{j=0}^{t-1} CA^j (BY(t-j-1) + \varepsilon(t-j)) + \eta(t), \qquad t > 0 \qquad (6.3)$$

Since X is determined by Y, $Z(0)$ and ξ and the latter two are unobservable, to study the effect of variations in Y along $\mathcal{V} \subseteq \mathbb{R}^{n_Y}$ on the variations in X along $\mathcal{U} \subseteq \mathbb{R}^{n_X}$, we will work with the information set $I(t) = P_{\mathcal{V}^\perp} Y[0, t]$ for all $t \geq 0$. Finally, denote by $\mathscr{T}$ the class of L^2 processes Y which are orthogonal to $Z(0)$ and ξ. $\qquad\qquad\square$

Now given this model, we would like to measure the effect of Y on X over and above the influence of all other factors. The engineering literature has solved this by looking at the effect of a deterministic process Y on $\mathbb{E}(X)$. Clearly, $\mathbb{E}(X)$ lies in the image of the sequence of matrices $\{CA^j B\}_{j=0}^\infty$; by the Cayley–Hamilton theorem (theorem 2.4.2 of Horn & Johnson (1985)) this is exactly the image of the matrix $[CB\ CAB\ \cdots\ CA^{n_Z-1}B]$, which is called the output controllability matrix. Thus the image of the output controllability matrix is precisely the range of values of X that are reachable in expectation by some choice of Y and the system is completely controllable (in the sense that any target is reachable in expectation) if and only if the output controllability matrix is of full rank.[10]

In contrast, the theory of causality allows us to approach the problem from a different point of view. For a given Y, the prediction variation $\Delta_h^{P_{\mathcal{U}} X P_{\mathcal{V}} Y I}(t)$ gives us some information about the causal effect of Y on X; therefore, to measure the independent effect of Y on X (i.e. in the absence of feedback) we will consider the causal effect of an arbitrary $Y \in \mathscr{T}$ on X. To keep things simple, let $Y \in \mathscr{T}$ be a white noise process with variance matrix I_{n_Y} and compute the prediction variation,

$$\Delta_h^{P_{\mathcal{U}} X P_{\mathcal{V}} Y I}(t) = \begin{cases} 0, & t = 0 \\ \sum_{j=h-1}^{t+h-1} P_{\mathcal{U}} CA^j B P_{\mathcal{V}} Y(t+h-j-1), & t > 0 \end{cases} \qquad (6.4)$$

---

[10]See Kailath (1980) or Sontag (1998) for more details. Preston & Pagan (1982) provide a fascinating interpretation of controllability in terms of Tinbergen's counting principle.

where we have used the fact that $P(Y(s)|I(t)) = P(Y(s)|P_{\mathcal{V}^\perp}Y(s)) = P_{\mathcal{V}^\perp}Y(s)$ for $0 \leq s \leq t$. It is now clear that $Y|_{\mathcal{V}} \not\rightarrow_h X|_{\mathcal{U}} [I]$ if and only if $P_{\mathcal{U}}CA^jBP_{\mathcal{V}} = 0$ for $j \geq h - 1$.[11] Note in particular that if $Y|_{\mathcal{V}} \not\rightarrow_h X|_{\mathcal{U}} [I]$ then $Y|_{\mathcal{V}} \not\rightarrow_j X|_{\mathcal{U}} [I]$ for all $j \geq h$ so that $Y|_{\mathcal{V}} \not\rightarrow_{(\infty)} X|_{\mathcal{U}} [I]$ if and only if $Y|_{\mathcal{V}} \not\rightarrow_1 X|_{\mathcal{U}} [I]$. In the special case where $h = 1$ and $\mathcal{V} = \mathbb{R}^{n_Y}$, we see that the reachable subspace is precisely $\mathcal{C}_1^{XYI}$. We prove a slightly stronger results in the following theorem.

Theorem 6.1. Under Assumption 5 with $\mathcal{V} = \mathbb{R}^{n_Y}$, the subspace $\mathcal{U} \subseteq \mathbb{R}^{n_X}$ is unreachable if and only if $\mathcal{U} \subseteq \mathcal{U}_1^{XYI}$ for all $Y \in \mathscr{T}$.

The relationship between causality and controllability is still more intimate. We know from DR's Separation Theorem that if $(Y', X'P_{\mathcal{U}^\perp})' \not\rightarrow_1 X|_{\mathcal{U}} [I_{P_{\mathcal{U}}X}]$ then $(Y', X'P_{\mathcal{U}^\perp})' \not\rightarrow_{(\infty)} X|_{\mathcal{U}} [I_{P_{\mathcal{U}}X}]$, where $I_{P_{\mathcal{U}}X}(t) = P_{\mathcal{U}}X(\omega, t]$ for $t > \omega$ and X and Y are as in Assumption 1; that is, if Y has neither a direct nor an indirect effect on X along $\mathcal{U}$ then Y has no effect at all on X. The next result shows that under Assumption 5 and when Z is perfectly observable the converse of the Separation Theorem holds and is precisely Kalman's controllability decomposition.

Theorem 6.2 (Partial Converse of the Separation Theorem). Suppose Assumptions 5 holds with $\mathcal{V} = \mathbb{R}^{n_Y}$, $C = I_{n_X}$, $\eta = 0$, and $I_{P_{\mathcal{U}}X}(t) = P_{\mathcal{U}}X[0, t]$ for $t \geq 0$. If $\mathcal{U} = \mathcal{U}_{(\infty)}^{XYI}$, then $X|_{\mathcal{U}^\perp} \not\rightarrow_{(\infty)} X|_{\mathcal{U}} [I_{P_{\mathcal{U}}X}]$.

We find in the proof of Theorem 6.2 that $P_{\mathcal{U}}AP_{\mathcal{U}^\perp} = 0$; thus if we set $U = U_{(\infty)}^{XYI}$ and $C = C_{(\infty)}^{XYI}$ then X decomposes as, $X = U\widetilde{X}_U + C\widetilde{X}_C$, where $\widetilde{X}_U = U'X$, $\widetilde{X}_C = C'X$ the system can be expressed as,

$$\begin{bmatrix} \widetilde{X}_U(t) \\ \widetilde{X}_C(t) \end{bmatrix} = \begin{bmatrix} U'AU & 0 \\ C'AU & C'AC \end{bmatrix} \begin{bmatrix} \widetilde{X}_U(t-1) \\ \widetilde{X}_C(t-1) \end{bmatrix} + \begin{bmatrix} 0 \\ C'B \end{bmatrix} Y(t-1) + \begin{bmatrix} U'\varepsilon(t) \\ C'\varepsilon(t) \end{bmatrix}$$

Thus the uncontrollable part $\widetilde{X}_U$ is a VAR(1) which is not causally related to Y, while $\widetilde{X}_C$ is related to Y and is characterized by a VARX(1,1). This is precisely Kalman's controllability decomposition, which can now be considered a partial converse to the Separation Theorem.

Finally, it has long been recognized that Granger–causality is directly relevant to optimal control (see e.g. Granger (1988a) and the references therein); however the full extent of the

---

[11]The "if" part follows from equation (6.4), while the "only if" part follows from the fact that if $\Delta_h^{P_{\mathcal{U}}XP_{\mathcal{V}}YI}(t) = 0$ for $t \geq 0$ then $0 = \mathbb{E}\Delta_h^{P_{\mathcal{U}}XP_{\mathcal{V}}YI}(t)Y'(t+h-j-1) = P_{\mathcal{U}}CA^jBP_{\mathcal{V}}$ for $h - 1 \leq j \leq t + h - 1$.

relationship has not been completely characterized as Granger only considers extreme forms of control where the policymaker gives zero weight to all variables except for one. The following result completely characterizes the solution to the linear quadratic optimal control problem in econometric terms.

Theorem 6.3. Suppose Assumption 5 holds and let $Q \in \mathbb{R}^{n_X \times n_X}$ and $R \in \mathbb{R}^{n_Y \times n_Y}$ be positive definite, with $\mathcal{L} = \mathbb{E}\{\sum_{t=0}^{\infty} \beta^t(X'(t)QX(t) + Y'(t)RY(t))\}$ and $0 < \beta < 1$. If $\mathcal{C}_{(\infty)}^{XYI} = \mathbb{R}^{n_X}$ for all $Y \in \mathscr{T}$ then the L^2 process Y that minimizes $\mathcal{L}$ exists and is unique.

7 Conclusion

This paper has demonstrated that the subspace perspective of causality encompasses existing notions of causality, stability, cointegration, and controllability. We have shown how to extend cartesian causality to take into account the subspaces along which causal links may reside. We have demonstrated that L^2 stability, a weaker form that second–order stationarity, can be viewed as a form of non–causality. We then specialized the theory to linear invertible process and derived the parametric restrictions for non–causality. The theory was then specialized even further to VARMA processes where we showed how cointegration can be seen as a special case of cotrendedness. Finally, we showed that the linear systems concept of controllability is also a special case of causality, providing purely econometric statements of two celebrated theorems in linear systems theory: the Kalman controllability decomposition and the existence and uniqueness theorem for optimal policies in linear quadratic control. For the rest of this section, therefore, we will focus on elaborating certain themes in the paper and suggest further extensions to the results.

First, the paper has relied heavily on the notion of maximality of subspaces with respect to a given property (in our case, the property of being a subspace along which there is non–causality). The existence of these subspaces follows from Zorn's lemma (see e.g. Artin (1991)) if the property is invariant to subspace summation; uniqueness then follows from maximality and additivity again. It is interesting to note the extent of analytic tractability that this method has afforded us. For example Theorem 3.1 is almost tautological and provides Granger's result in full generality where as the original Granger (1988b) result relies heavily on the representation theory of bivariate $I(1)$ time series. It would be fruitful to see this methodology

applied to other problems in multivariate time series analysis.

Second, we have completely ignored the relationship between reduced rank regression (i.e. the results of Section 4) and canonical correlations analysis (see e.g. Reinsel & Velu (1998)). Although the two points of view are practically equivalent in the case of finite information sets; the situation is drastically complicated when the information set is infinite dimensional. Certain results are available for canonical correlations analysis in infinite dimensions (see e.g. Jewell & Bloomfield (1983)); however these concern stationary processes and it would be interesting to see how they extend to our setting; in particular, one would expect that the subspaces of non–causality are precisely those pertaining to canonical correlations equal to zero.

Third, the paper introduced a new concept of long run causality, which encompasses the concepts of Bruneau & Jondeau (1999) and Yamamoto & Kurozumi (2006). There is, however, a frequency–domain concept of long run causality (Hosoya (1991) and Hosoya (2001)) and it was not clear at the time of writing this paper, whether or in what way the two concepts overlap. It would seem reasonable to expect that they are equivalent; however, an extension in that direction was beyond the scope of this paper and is left to further research.

Fourth, the linear theory we have studied in this paper can be seen as a first step towards a non–linear theory of Granger causality, which extracts causally related non–linear components from multivariate time series. In particular, we know from Lemma 3.1 that $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}} \, [\, I \,]$ if and only if $U'X(t+h)$ is not linearly related to past and present values of Y. The non–linear extension of this theory would consider the set of all Borel measureable functions g on $\mathbb{R}^{n_X}$ such that $\mathbb{E}(g(X(t+h))|X(t), Y(t), X(t-1), Y(t-1), \ldots) = \mathbb{E}(g(X(t+h))|X(t), X(t-1), \ldots)$.

Finally, subspace causality was demonstrated to be a generalization of model reduction techniques such as Sargent & Sims (1977) and Velu et al. (1986). It would be interesting to see how the more general kinds of subspace non–causality can be applied for model reduction. In the same vain, it would be interesting to see how Bayesian analysis can be conducted using subspace non–causality priors. These are all interesting questions, which will hopefully be addressed by future research.

8   Appendix

8.1   Relationships Between Cartesian and Subspace Non–Causality

Fortunately, very simple relationships exists between many of the results in the cartesian non–causality literature and the proposed subspace non–causality of this paper. We will focus on the case when $W = (X', Y', Z')$ is an L^2 process under investigation. From Lemma 3.1 we know that $Y|_\mathcal{V} \nrightarrow_h X|_\mathcal{U} \; [\, I \,]$ if and only if $\widetilde{Y} \nrightarrow_h \widetilde{X} \; [\, I \,]$, where $\widetilde{Y} = V'Y$ and $\widetilde{X} = U'X$. It would seem therefore that in order to use results about cartesian non–causality all that is required is to make the following "translation,"

$$X \mapsto \widetilde{X} = U'X$$

$$Y \mapsto \widetilde{Y} = V'Y$$

$$Z \mapsto \widetilde{Z} = (Z', X'U_\perp, Y'V_\perp)'$$

Note that such transformations involves no loss of information as it amounts to nothing more than multiplication of W by the unitary matrix,

$$\begin{bmatrix} U' & 0 & 0 \\ 0 & V' & 0 \\ 0 & 0 & I_{n_Z} \\ U'_\perp & 0 & 0 \\ 0 & V'_\perp & 0 \end{bmatrix}$$

Some cartesian non–causality results require assumptions about the information set I; these assumptions translate easily to the subspace setting. If, for example, I is required to be conformable with X, we work with an information set $\widetilde{I}$ that must now by conformable with $\widetilde{X}$. Some of DR's results require that $I(t) = H + X(\omega, t] + Z(\omega, t]$ for $t > \omega$, where H may include constants and initial conditions, in that case we require the information set to satisfy, $\widetilde{I}(t) = H + \widetilde{X}(\omega, t] + \widetilde{Z}(\omega, t] = H + X(\omega, t] + V'_\perp Y(\omega, t] + Z(\omega, t]$ for $t > \omega$.

The above correspondences can be used to translate any results about cartesian non–causality to the subspace perspective. Indeed we prove all of the new results below for the cartesian non–causality case as it is notationaly much more convenient.

8.2 Proofs

Proof of Lemma 3.1. Recall that $P_{\mathcal{U}} = UU'$ and $P_{\mathcal{V}} = VV'$ (see e.g. Theorem 2.5.1 of Brockwell & Davis (1991) and the subsequent remark). This implies that $P_{\mathcal{V}}Y(\omega, t] = V'Y(\omega, t]$. Now for $h < \infty$, $\Delta_h^{P_{\mathcal{U}}XP_{\mathcal{V}}YI}(t) = P_{\mathcal{U}}\Delta_h^{XP_{\mathcal{V}}YI}(t) = UU'\Delta_h^{XV'YI}(t) = U\Delta_h^{U'XV'YI}(t)$, which is zero if and only if $\Delta_h^{U'XV'YI}(t) = 0$. As for the long run case simply note that,

$$\mathbb{E}\|\Delta_h^{P_{\mathcal{U}}XP_{\mathcal{V}}YI}(t)\|^2 = \mathbb{E}\|U\Delta_h^{U'XV'YI}(t)\|^2 = \mathbb{E}\|\Delta_h^{U'XV'YI}(t)\|^2. \qquad \square$$

Proof of Lemma 3.2. We prove the case of non–causality at horizon h; the case of non–causality up to horizon h is almost identical and is omitted.

(i) Since $\mathcal{W} \subseteq \mathcal{V}$, $P_{\mathcal{W}}Y(\omega, t] \subseteq P_{\mathcal{V}}Y(\omega, t]$ and we have,

$$\begin{aligned}
\Delta_h^{P_{\mathcal{U}}XP_{\mathcal{W}}YI}(t) &= P(P_{\mathcal{U}}X(t+h)|I(t) + P_{\mathcal{W}}Y(\omega, t]) - P(P_{\mathcal{U}}X(t+h)|I(t)) \\
&= P(P_{\mathcal{U}}X(t+h) - P(P_{\mathcal{U}}X(t+h)|I(t))|I(t) + P_{\mathcal{W}}Y(\omega, t]) \\
&= P(P(P_{\mathcal{U}}X(t+h) - P(P_{\mathcal{U}}X(t+h)|I(t))|I(t) + P_{\mathcal{V}}Y(\omega, t])|I(t) + P_{\mathcal{W}}Y(\omega, t]) \\
&= P(\Delta_h^{P_{\mathcal{U}}XP_{\mathcal{V}}YI}(t)|I(t) + P_{\mathcal{W}}Y(\omega, t]),
\end{aligned}$$

by the law of iterated projections. Now if $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\ [I]$ and $h < \infty$ then the term inside the projection is zero and the result follows; if on the other hand, $h = \infty$, then the term inside the projection goes to zero in L^2 and the result follows from the continuity of the projection operator (see e.g. Proposition 2.3.2 (iv) of Brockwell & Davis (1991)). The converse for each case follows by taking $\mathcal{W} = \mathcal{V}$.

(ii) If $\mathcal{W} \subseteq \mathcal{U}$ then by the law of iterated projections $P_{\mathcal{W}}P_{\mathcal{U}} = P_{\mathcal{W}}$ and from the properties of matrix norms,

$$\|\Delta_h^{P_{\mathcal{W}}XP_{\mathcal{V}}YI}(t)\| = \|P_{\mathcal{W}}\Delta_h^{P_{\mathcal{U}}XP_{\mathcal{V}}YI}(t)\| \le \|P_{\mathcal{W}}\|\|\Delta_h^{P_{\mathcal{U}}XP_{\mathcal{V}}YI}(t)\|$$

If $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\ [I]$ and $h < \infty$ then the right hand side is zero; on the other hand if $h = \infty$ then the right hand side goes to zero in L^2. The converse follows by taking $\mathcal{W} = \mathcal{U}$.

(iii) $Y|_{\mathcal{V}_j} \nrightarrow_h X|_{\mathcal{U}}\ [I]$ for $j \in J$, implies that $P_{\mathcal{U}}X(t+h) - P(P_{\mathcal{U}}X(t+h)|I(t))$ is orthogonal (resp. asymptotically orthogonal) to the Hilbert spaces $I(t) + P_{\mathcal{V}_j}Y(\omega, t]$, $j \in J$ when $h < \infty$ (resp. $h = \infty$). The result then follows if we can prove that the spaces $\{I(t) + P_{\mathcal{V}_j}Y(\omega, t]\}_{j \in J}$ generate $I(t) + P_{\sum_{j \in J}\mathcal{V}_j}Y(\omega, t]$ because then $P_{\mathcal{U}}X(t+h) - P(P_{\mathcal{U}}X(t+h)|I(t))$ is orthogonal (resp. asymptotically orthogonal) to $I(t) + P_{\sum_{j \in J}\mathcal{V}_j}Y(\omega, t]$ for $h < \infty$ (resp. $h = \infty$). Thus

we claim that $\overline{\mathrm{sp}}\{I(t) + P_{\mathcal{V}_j}Y(\omega, t] : j \in J\} = I(t) + P_{\sum_{j \in J}\mathcal{V}_j}Y(\omega, t]$; we prove this using a Gram–Schmidt decomposition of the subspace $\sum_{j \in J}\mathcal{V}_j$.

Since $P_{\mathcal{V}_j} = P_{\mathcal{V}_j}P_{\sum_{j \in J}\mathcal{V}_j}$ for all $j \in J$, $I(t) + P_{\mathcal{V}_j}Y(\omega, t] \subseteq I(t) + P_{\sum_{j \in J}\mathcal{V}_j}Y(\omega, t]$ for all $j \in J$; therefore, $\overline{\mathrm{sp}}\{I(t) + P_{\mathcal{V}_j}Y(\omega, t] : j \in J\} \subseteq I(t) + P_{\sum_{j \in J}\mathcal{V}_j}Y(\omega, t]$. On the other hand, since we are in finite Euclidean space, $\sum_{j \in J}\mathcal{V}_j = \sum_{j \in J'}\mathcal{V}_j$, where $J' \subseteq J$ is finite; we relabel the elements of this set to consist of integers in $\{1, 2, \ldots\}$. Now partition the latter subspace as follows.

$$\mathcal{W}_1 = \mathcal{V}_1, \quad \mathcal{W}_{j+1} = \mathcal{V}_{j+1} \cap \mathcal{W}_j^\perp, \quad j = 1, \ldots, |J'| - 1,$$

and reorder the sets if necessary to put all the null spaces at the end of the list with the set $J'' \subseteq J'$ consisting of the non–null spaces. Then, $\sum_{j \in J}\mathcal{V}_j = \sum_{j \in J''}\mathcal{W}_j$ and $P_{\sum_{j \in J}\mathcal{V}_j} = \sum_{j \in J''}P_{\mathcal{W}_j}$. Since $\mathcal{W}_j \subseteq \mathcal{V}_j$ for all $j \in J''$ it follows that, $I(t) + P_{\sum_{j \in J}\mathcal{V}_j}Y(\omega, t] = I(t) + P_{\mathcal{W}_1}Y(\omega, t] + \cdots P_{\mathcal{W}_{|J''|}}Y(\omega, t] \subseteq I(t) + P_{\mathcal{V}_1}Y(\omega, t] + \cdots P_{\mathcal{V}_{|J''|}}Y(\omega, t] \subseteq \overline{\mathrm{sp}}\{I(t) + P_{\mathcal{V}_j}Y(\omega, t] : j \in J\}$.

(iv) As we did in (iii), let $\{\mathcal{W}_j\}_{j \in J''}$ be a finite collection of mutually orthogonal spaces such that, $\sum_{j \in J}\mathcal{U}_j = \sum_{j \in J''}\mathcal{W}_j$ and $\mathcal{W}_j \subseteq \mathcal{U}_j$ for all $j \in J''$. Then $P_{\sum_{j \in J}\mathcal{U}_j} = \sum_{j \in J''}P_{\mathcal{W}_j}$. Since each $\mathcal{W}_j$ is a subspace along which non–causality occurs, by (ii) we have, $P(P_{\mathcal{W}_j}X(t+h)|I(t) + P_{\mathcal{V}}Y(\omega, t]) = P(P_{\mathcal{W}_j}X(t+h)|I(t))$ for $h < \infty$. The result then follows on summing across j. If on the other hand $h = \infty$, then $P(P_{\mathcal{W}_j}X(t+h)|I(t) + P_{\mathcal{V}}Y(\omega, t]) - P(P_{\mathcal{W}_j}X(t+h)|I(t)) \to 0$ in L^2 as $h \to \infty$; summing again across j, we arrive at the desired result. $\qquad\square$

Proof of Lemma 3.3. We prove only the case of non–causality at horizon h; the case of up to horizon h non–causality follows a similar argument. To prove existence consider the collection of all subspaces $\mathcal{U}$ such that $Y|_{\mathcal{V}} \not\rightarrow_h X|_{\mathcal{U}} \, [\, I \,]$ and order them by inclusion. Now any linearly ordered subset of these subspaces will have an upper bound namely its sum; this follows from Lemma 3.2 (iv). Therefore by Zorn's lemma a maximal element exists.[12] Uniqueness is proven by noting that if $\mathcal{U}_1$ and $\mathcal{U}_2$ are maximal then by Lemma 3.2 (iv) again $Y|_{\mathcal{V}} \not\rightarrow_h X|_{\mathcal{U}_1 + \mathcal{U}_2} \, [\, I \,]$; maximality then gives us that $\mathcal{U}_1 + \mathcal{U}_2$ is equal to both $\mathcal{U}_1$ and $\mathcal{U}_2$. The opposite case, fixing $\mathcal{U}$ instead of $\mathcal{V}$, follows a similar argument. $\qquad\square$

Proof of Proposition 3.1. We prove only (i) – (vi) as (vii) – (xii) follow similar arguments. Since $\mathcal{U}_h^{XYI}$ is maximal, $\mathcal{U} \subseteq \mathcal{U}_h^{XYI}$ for every $\mathcal{U}$ such that $Y \not\rightarrow_h X|_{\mathcal{U}} \, [\, I \,]$. By Lemma 3.2, $\sum_{\{\mathcal{U}:Y \not\rightarrow_h X|_{\mathcal{U}} \, [I]\}}\mathcal{U} \subseteq \mathcal{U}_h^{XYI}$. On the other the other hand, $\mathcal{U}_h^{XYI} \in \{\mathcal{U} : Y \not\rightarrow_h X|_{\mathcal{U}} \, [\, I \,]\}$ so

---

[12]Artin (1991) gives a clear and concise exposition on the uses of Zorn's lemma in algebra.

that $\mathcal{U}_h^{XYI} \subseteq \sum_{\{\mathcal{U}:Y \nrightarrow_h X|_{\mathcal{U}}[I]\}} \mathcal{U}$. This proves (i) and (ii) follows the same line of argument. (iii) follows from Definition 3.3. To prove (iv) note that $P_{\mathcal{U}_{(\infty)}^{XYI}} \Delta_h^{XYI}(t) = 0$ for all $h \geq 1$ and $t > \omega$ implies that $P_{\mathcal{U}_{(\infty)}^{XYI}} \Delta_h^{XYI}(t) \to 0$ in L^2 as $h \to \infty$ for all $t > \omega$. (v) and (vi) follow from the fact that $\sum_{i=1}^h \mathcal{W}_i^{\perp} = (\bigcap_{i=1}^h \mathcal{W}_i)^{\perp}$ and $(\bigcap_{i=1}^h \mathcal{W}_i)^{\perp} \subseteq (\bigcap_{i=1}^{h+1} \mathcal{W}_i)^{\perp}$ respectively for any collection of subspaces $\{\mathcal{W}_i\}_{i=1}^{h+1}$ of $\mathbb{R}^{n_X}$ (see exercise 15 p. 254 of Artin (1991)). $\square$

Proof of Theorem 3.1. Follows directly from the maximality of $\mathcal{M}_{\infty}^X$. A more constructive proof is the following: suppose to the contrary that for all $\mathcal{M}_1 \subseteq \mathbb{R}^{n_X}$ and $\mathcal{M}_2 \subseteq (\mathcal{M}_{\infty}^X)^{\perp}$, $X|_{\mathcal{M}_1} \nrightarrow_{\infty} X|_{\mathcal{M}_2} [H_{\omega}(X)]$. Then the choice $\mathcal{M}_1 = \mathbb{R}^{n_X}$, $\mathcal{M}_2 = (\mathcal{M}_{\infty}^X)^{\perp}$ leads to a contradiction as it implies, by Lemma 3.2 (iv), that $\mathcal{M}_{\infty}^X = \mathbb{R}^{n_X}$. $\square$

Proof of Theorem 4.1. Follows from DR's Theorem 3.1 and subsection 8.1. $\square$

Proof of Corollary 4.1. $\mathcal{C}_h^{XYI}$ is the orthogonal complement of $\mathcal{U}_h^{XYI}$, which is the space orthogonal to the span of the columns of $\{\pi_{XYj}^{(h)}\}_{j=1}^{\infty}$ by Theorem 4.1; this proves (i). (ii) follows from the fact that $\mathrm{im}(\pi_{XYj}^{(h)})^{\perp} = \ker(\pi_{XYj}^{(h)}{}')$ and the fact that $\sum_{i=1}^h \mathcal{W}_i^{\perp} = (\bigcap_{i=1}^h \mathcal{W}_i)^{\perp}$ for any collection of subspaces $\{\mathcal{W}_i\}_{i=1}^{h+1}$ of $\mathbb{R}^{n_X}$ (see exercise 15 p. 254 of Artin (1991)). (iii) and (iv) follow similarly. $\square$

Proof of Theorem 4.2. We will prove the cartesian causality version of the theorem (i.e. the case $\mathcal{U} = \mathbb{R}^{n_X}$ and $\mathcal{V} = \mathbb{R}^{n_Y}$); the general case then follows from subsection 8.1.

The first part is proven similarly to DR's Theorem 3.1. Suppose that $\Delta_h^{XYI}(t) = (\pi_{X.}^{(h)}(L) - \phi_{X.}^{(h)}(L))W(t+1)$, where $\phi_{X.}^{(h)}(L) = [\phi_{XX}^{(h)}(L) \ 0 \ \phi_{XZ}^{(h)}(L)]$ is a power series in the lag operator L and $\pi_{X.}^{(h)}(L) = [\pi_{XX}^{(h)}(L) \ \pi_{XY}^{(h)}(L) \ \pi_{XZ}^{(h)}(L)]$. If $\Delta_h^{XYI}(t) \to 0$ in L^2 then from the properties of the dot product, $\mathbb{E}(\Delta_h^{XYI}(t)a'(t)) \to 0$. Therefore, $\sum_{j=1}^{\infty}[\pi_{XXj}^{(h)} - \phi_{XXj}^{(h)} \ \pi_{XYj}^{(h)} \ \pi_{XZj}^{(h)} - \phi_{XZj}^{(h)}]\mathbb{E}(W(t-j)a'(t)) \to 0$. Since $\mathbb{E}(W(t-j)a'(t)) = \Omega(t) > 0$ for $j = 0$ and is zero otherwise, this implies that $[\pi_{XX1}^{(h)} - \phi_{XX1}^{(h)} \ \pi_{XY1}^{(h)} \ \pi_{XZ1}^{(h)} - \phi_{XZ1}^{(h)}] \to 0$ and so $\pi_{XY1}^{(h)} \to 0$. Now since the first summand of $\Delta_h^{XYI}(t)$ converges to zero the entire process can be repeated again, first noting that $\mathbb{E}(\Delta_h^{XYI}(t)a'(t-1)) \to 0$, then factoring out $\Omega(t-1)$ and finally isolating $[\pi_{XX2}^{(h)} - \phi_{XX2}^{(h)} \ \pi_{XY2}^{(h)} \ \pi_{XZ2}^{(h)} - \phi_{XZ2}^{(h)}] \to 0$. Continuing on with this process proves that, $\lim_{h \to \infty} \pi_{XYj}^{(h)} = 0$ for all $j \geq 1$.

To prove the converse we use equation (4.6), setting $\xi(t+1-j) = Y(t+1-j) - P(Y(t+$

$1-j)|I(t))$ to simplify the notation,

$$\mathbb{E}\|\Delta_h^{XYI}(t)\|^2 = \mathbb{E}\left\|\sum_{j=1}^{t-\varpi} \pi_{XYj}^{(h)}\xi(t+1-j)\right\|^2$$

$$\le \mathbb{E}\left(\sum_{j=1}^{t-\varpi} \|\pi_{XYj}^{(h)}\xi(t+1-j)\|\right)^2$$

$$\le \mathbb{E}\left(\sum_{j=1}^{t-\varpi} \|\pi_{XYj}^{(h)}\|\|\xi(t+1-j)\|\right)^2,$$

where the last two inequalities follow from properties of the norm.

$$= \mathbb{E}\sum_{j=1}^{t-\varpi}\sum_{k=1}^{t-\varpi} \|\pi_{XYj}^{(h)}\|\|\pi_{XYk}^{(h)}\|\|\xi(t+1-j)\|\|\xi(t+1-k)\|$$

$$= \sum_{j=1}^{t-\varpi}\sum_{k=1}^{t-\varpi} \|\pi_{XYj}^{(h)}\|\|\pi_{XYk}^{(h)}\|\mathbb{E}\{\|\xi(t+1-j)\|\|\xi(t+1-k)\|\},$$

by the Fubini–Tonelli theorem.

$$\le \sum_{j=1}^{t-\varpi}\sum_{k=1}^{t-\varpi} \|\pi_{XYj}^{(h)}\|\|\pi_{XYk}^{(h)}\| \left(\mathbb{E}\|\xi(t+1-j)\|^2\right)^{\frac{1}{2}} \left(\mathbb{E}\|\xi(t+1-k)\|^2\right)^{\frac{1}{2}},$$

by the Cauchy–Schwartz theorem.

$$\le \sum_{j=1}^{t-\varpi}\sum_{k=1}^{t-\varpi} \|\pi_{XYj}^{(h)}\|\|\pi_{XYk}^{(h)}\| \sup_{\varpi < s \le t} \mathbb{E}\|\xi(s)\|^2$$

$$\le \sum_{j=1}^{t-\varpi}\sum_{k=1}^{t-\varpi} \|\pi_{XYj}^{(h)}\|\|\pi_{XYk}^{(h)}\| \sup_{\varpi < s \le t} \mathbb{E}\|Y(s) - P(Y(s)|H_{-\infty}(W))\|^2,$$

because projections onto H produce larger mean squre error than projections on $I(t)$.

$$= \left(\sum_{k=1}^{t-\varpi} \|\pi_{XYj}^{(h)}\|\right)^2 \sup_{\varpi < s \le t} \mathbb{E}\|\widehat{Y}(s)\|^2,$$

which goes to zero as $h \to 0$ by assumption. $\qquad\square$

Proof of Theorem 4.3. Follows from DR's Theorem 3.2 and subsection 8.1. $\qquad\square$

Proof of Theorem 5.1. We prove the theorem from the cartesian causality perspective, the subspace version then follows from subsection 8.1.

(i) The proof is in two steps. We will require the following result (DR's Lemma A.4), which is easily proven by applying the multiplication rule for power series (see e.g. p. 84 of Brockwell & Davis (1991) or Lütkepohl (2006) Proposition 2.4).

28

Lemma 8.1. Suppose $a(z) = \sum_{i=0}^{\infty} a_i z^i$ is a power series and $a(z) = b(z)c(z)$, where $b(z)$ is a power series with a non–zero radius of convergence and $c(z)$ is a polynomial of degree p. Then $\{a_i\}_{i=0}^{\infty} = \{0\}$ if and only if $\{a_i\}_{i=0}^{p} = \{0\}$.

Step 1: $Y \nrightarrow_1 X \,[\,I\,]$ if and only if $\pi_{XYj} = 0$ for all $1 \leq j \leq p + (n-1)q$.

$I_n - \pi(z) = \frac{\theta^*(z)\phi(z)}{\det(\theta(z))}$, where $\theta^*(z)$ is the adjoint of $\theta(z)$. The degree of $\theta^*(z)\phi(z)$ is at most $p + (n-1)q$, while the degree of $\det(\theta(z))$ is nq; thus the typical element of $I_n - \pi(z)$ is representable by a fraction with numerator of degree $p + (n-1)q$ and denominator $\det(\theta(z))$. It follows that the same holds true for $-\pi_{XY}(z)$, being an off diagonal submatrix of $I_n - \pi(z)$. By Lemma 8.1 now, $\pi_{XY}(z) = 0$ if and only if its first $p + (n-1)q$ coefficients are zero, that is, if and only if $\pi_{XYj} = 0$ for all $1 \leq j \leq p + (n-1)q$.

Step 2: $Y \nrightarrow_h X \,[\,I\,]$ if and only if $\pi_{XYj}^{(h)} = 0$ for all $1 \leq j \leq p + (n-1)q$.

We prove this by showing that $\pi_{XY}^{(h)}(z)$ is a ratio of a $p + (n-1)q$–order polynomial and the nq–order polynomial $\det(\theta(z))$. The proof is by induction. Suppose that the typical element of $\pi_{XY}^{(i)}(z)$ is representable by a ratio of a $p + (n-1)q$–order polynomial and the nq–order polynomial $\det(\theta(z))$ for $1 \leq i \leq h-1$. The case $h = 1$ was proven in step 1; if we can prove the general case then the statement of step 2 will follow as a corollary using Lemma 8.1. From equation (4.4),

$$\pi^{(h+1)}(z) = z^{-1}\pi^{(h)}(z) + \pi_1^{(h)}(\pi(z) - I_n), \qquad h \geq 1 \tag{8.1}$$

It follows that,

$$\pi_{XY}^{(h)}(z) = z^{-1}\pi_{XY}^{(h-1)}(z) + \pi_{XX1}^{(h-1)}\pi_{XY}(z) + \pi_{XY1}^{(h-1)}(\pi_{YY}(z) - I_{n_Y}) + \pi_{XZ1}^{(h-1)}\pi_{ZY}(z) \tag{8.2}$$

Each summand on the left hand side is representable by a ratio of a $p + (n-1)q$–order polynomial and the nq–order polynomial $\det(\theta(z))$ by the induction hypothesis and the discussion in step 1. In particular, since $\pi_{XY}^{(h-1)}(z)$ is representable by a ratio of a $p + (n-1)q$–order polynomial and the nq–order polynomial $\det(\theta(z))$ and it clearly has a zero at $z = 0$, $z^{-1}\pi_{XY}^{(h-1)}(z)$ is representable by a ratio with a numerator of degree $p + (n-1)q - 1$ and the denominator $\det(\theta(z))$.

(ii) Note that, $\mathbb{E}\|\widehat{Y}(s)\|^2 \leq \mathbb{E}\|\widehat{W}(s)\|^2 \leq \sup_{\varpi < s \leq t} \mathbb{E}\|\widehat{W}(s)\|^2$ for all $\varpi < s \leq t \in \mathbb{Z}$. By Theorem 4.2, since $\sup_{\varpi < s \leq t} \mathbb{E}\|\widehat{W}(s)\|^2 < \infty$ all that remains to be shown is uniform convergence. The proof is in two steps.

Step 1. $\lim_{h \to \infty} \sum_{j=1}^{t-\varpi} \|\pi_{XYj}^{(h)}\| = 0$ if $\lim_{h \to \infty} \pi_{XYj}^{(h)} = 0$ for all $j \geq 1$.

Section 11.5 of Lütkepohl (2006) gives the following formula for h–step forecasts of VARMA processes,

$$P(W(t+h)|W(-\infty,t]) = \sum_{k=0}^{h-1} \pi_1^{(k)} \mu(t+h-k) + C'A_1^h \widetilde{W}(t), \qquad t \geq \varpi + q$$

where,

$$A_1 = \left[
\begin{array}{cccccc|ccccc}
\phi_1 & \phi_2 & \cdots & & \cdots & \phi_p & \theta_1 & \theta_2 & \cdots & \cdots & \theta_q \\
I_n & 0 & \cdots & & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 \\
0 & I_n & \ddots & & & \vdots & \vdots & & & & \vdots \\
\vdots & \vdots & \ddots & & & & & & & & \\
\vdots & \vdots & & \ddots & \ddots & \vdots & \vdots & & & & \vdots \\
0 & 0 & \cdots & I_n & 0 & 0 & 0 & \cdots & & \cdots & 0 \\
\hline
0 & 0 & \cdots & & \cdots & 0 & 0 & 0 & \cdots & & \cdots & 0 \\
0 & 0 & \cdots & & \cdots & 0 & I_n & 0 & \cdots & & \cdots & 0 \\
\vdots & & & & & \vdots & 0 & I_n & \ddots & & & \vdots \\
& & & & & & & & \ddots & & \\
\vdots & & & & & \vdots & \vdots & & & \ddots & \ddots & \vdots \\
0 & \cdots & & & \cdots & 0 & 0 & \cdots & & & 0 & I_n & 0
\end{array}
\right],$$

$$\widetilde{W}(t) = \left[
\begin{array}{c}
W(t) \\
W(t-1) \\
\vdots \\
\\
\vdots \\
W(t-p+1) \\
\hline
a(t) \\
a(t-1) \\
\vdots \\
\\
\vdots \\
a(t-q+1)
\end{array}
\right], \quad C = \left[
\begin{array}{c}
I_n \\
0 \\
\vdots \\
\\
\vdots \\
0 \\
\hline
0 \\
0 \\
\vdots \\
\\
\vdots \\
0
\end{array}
\right]$$

If we now substituted for the shocks $\{a(t-j)\}_{j=0}^{q-1}$ we arrive at,

$$\widetilde{W}(t) = \left[\begin{array}{cccccc|ccc}
I_n & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots \\
0 & I_n & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots \\
0 & 0 & I_n & \cdots & \cdots & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots & & \vdots & \vdots & \vdots & \\
\vdots & \vdots & \vdots & & \ddots & \vdots & \vdots & \vdots & \\
0 & 0 & 0 & \cdots & \cdots & I_n & 0 & 0 & \cdots \\
\hline
I_n & -\pi_1 & -\pi_2 & \cdots & \cdots & -\pi_{p-1} & -\pi_p & -\pi_{p+1} & \cdots \\
0 & I_n & -\pi_1 & \cdots & \cdots & -\pi_{p-2} & -\pi_{p-1} & -\pi_p & \cdots \\
0 & 0 & I_n & \cdots & \cdots & -\pi_{p-3} & -\pi_{p-2} & -\pi_{p-1} & \cdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \cdots & I_n & -\pi_1 & -\pi_2 & -\pi_3 & \cdots
\end{array}\right] \left[\begin{array}{c}
W(t) \\
W(t-1) \\
W(t-2) \\
\vdots \\
\vdots \\
W(t-p+1) \\
\hline
W(t-p) \\
W(t-p-1) \\
\vdots
\end{array}\right],$$

Note that the above equation presumes that $q = p - 1$, although we make no such assumption (the form above is given for illustration only). Now setting,

$$F_j = \left[\begin{array}{c}
0 \\
0 \\
0 \\
\vdots \\
\vdots \\
0 \\
\hline
-\pi_j \\
-\pi_{j-1} \\
-\pi_{j-2} \\
\vdots \\
\vdots \\
-\pi_{j-q+1}
\end{array}\right],$$

which is an $n(p+q) \times n$ matrix, and matching coefficients we finally arrive at,

$$\pi_j^{(h)} = C'A_1^h F_j, \qquad j \geq \max(p,q)$$

On the other hand, following the analysis of section 11.3 in Lütkepohl (2006), the projection

matrices can be obtained as,

$$\pi_j = -C' A_2^j B,$$

where,

$$A_2 = \left[\begin{array}{ccccc|ccccc}
-\theta_1 & -\theta_2 & \cdots & \cdots & -\theta_q & -\phi_1 & -\phi_2 & \cdots & \cdots & -\phi_p \\
I_n & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 \\
0 & I_n & \ddots & & \vdots & \vdots & & & & \vdots \\
\vdots & \vdots & \ddots & & & & & & & \\
\vdots & \vdots & & \ddots & \ddots & \vdots & \vdots & & & \vdots \\
0 & 0 & \cdots & I_n & 0 & 0 & 0 & \cdots & \cdots & 0 \\
\hline
0 & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 \\
0 & 0 & \cdots & \cdots & 0 & I_n & 0 & \cdots & \cdots & 0 \\
\vdots & & & & \vdots & 0 & I_n & \ddots & & \vdots \\
& & & & & & & \ddots & & \\
\vdots & & & & \vdots & \vdots & & \ddots & \ddots & \vdots \\
0 & \cdots & & \cdots & 0 & 0 & \cdots & 0 & I_n & 0
\end{array}\right],$$

$$B = \left[\begin{array}{c}
I_n \\
0 \\
\vdots \\
\\
\vdots \\
0 \\
\hline
I_n \\
0 \\
\vdots \\
\\
\\
\vdots \\
0
\end{array}\right]$$

It follows from the properties of matrix powers that $\pi_{XYj}^{(h)}$ consists of linear combinations of the form $h^k j^l \lambda^h \nu^j$, where the λ's are eigenvalues of A_1, the ν's are eigenvalues of A_2, and the k's and l's are integers. If $\lim_{h\to\infty} \pi_{XYj}^{(h)} = \lim_{j\to\infty} \pi_{XYj}^{(h)} = 0$ then all such λ's and ν's must

lie strictly inside the unit circle. It follows that $\|\pi_{XYj}^{(h)}\| < c\rho^{j+h}$ for some constants $c > 0$ and $0 < \rho < 1$ and uniformity follows.

Step 2. $\lim_{h\to\infty} \pi_{XYj}^{(h)} = 0$ for all $j \geq 1$ if and only if $\lim_{h\to\infty} \pi_{XYj}^{(h)} = 0$ for all $1 \leq j \leq (n_X + n_Z)nq + n_Y(p + (n-1)q)$.

From equation (4.4), it is easy to check that for any integer $m \geq 2$, $\lim_{h\to\infty} \pi_{XYj}^{(h)} = 0$ for $1 \leq j \leq m$ if and only if $\lim_{h\to\infty} \pi_{XY1}^{(h)} = 0$ and $\lim_{h\to\infty} \pi_{XX1}^{(h)}\pi_{XYj} + \pi_{XZ1}^{(h)}\pi_{ZYj} = 0$ for all $1 \leq j \leq m-1$. Thus we will have proven our claim if we can show that $\lim_{h\to\infty} \pi_{XX1}^{(h)}\pi_{XYj} + \pi_{XZ1}^{(h)}\pi_{ZYj} = 0$ for all $j \geq 1$ if and only if the first $(n_X+n_Z)nq+n_Y(p+(n-1)q)-1$ equations hold. Recall from (i) that $\pi_{XY}(z)$ and $\pi_{ZY}(z)$ are representable as ratios of $(p + (n - 1)q)$–order matrix polynomials and an nq–order polynomial; thus, following the standard state space representation methodology, $[\pi'_{XY}(z) \quad \pi'_{ZY}(z)]'$ is expressible as a matrix power series of the form $C(I_m - Az)^{-1}Bz$, where $m = (n_X + n_Z)nq + n_Y(p + (n - 1)q)$ (see e.g. pp. 426–429 of Lütkepohl (2006)). It follows that $\pi_{XX1}^{(h)}\pi_{XY}(z) + \pi_{XZ1}^{(h)}\pi_{ZY}(z)$ is expressible as $D_hC(I_{m_1} - Az)^{-1}Bz$. Now the Cayley–Hamilton Theorem (see Theorem 2.4.2 in Horn & Johnson (1985)) implies that $\lim_{h\to\infty} D_hCA^jB = 0$ for all $j \geq 1$ if and only if the first $m-1$ terms are zero.

(iii) The proof is in two steps.

Step 1: $Y \nrightarrow_{(\infty)} X\,[\,I\,]$ if and only if $\pi_{XYj}^{(h)} = 0$ for all $j \geq 1$ and $1 \leq h \leq n_Z(p+(n-1)q)+1$.

From DR's Lemma 3.2 and Lemma A.3, $Y \nrightarrow_{(\infty)} X\,[\,I\,]$ is equivalent to $\pi_{XY}(z) = 0$ and $\pi_{XZ}(z)(I_{n_Z} - \pi_{ZZ}(z))^{-1}\pi_{ZYj} = 0$ for all $j \geq 1$.[13] Now we know from step 1 of (i) that $I_{n_Z} - \pi_{ZZ}(z)$ is representable by a matrix polynomial of degree $p + (n - 1)q$ divided by a polynomial of degree nq. Thus, modulo $\det(\theta(z))$, the typical element of $(I_{n_Z} - \pi_{ZZ}(z))^{-1}$ is representable by a fraction with numerator of degree $(n_Z-1)(p+(n-1)q)$ and a denominator of degree $n_Z(p + (n - 1)q)$. Now the common factor, $\det(\theta(z))$, cancels out of $\pi_{XZ}(z)(I_{n_Z} - \pi_{ZZ}(z))^{-1}$ and so each of its elements is representable by fraction with numerator of degree $n_Z(p+(n-1)q)$ and a denominator of degree $n_Z(p+(n-1)q)$. It follows from Lemma 8.1 that

---

[13] A quicker proof of this than DR's proof is obtained by by noting from equation (8.2) that $Y \nrightarrow_{(\infty)} X\,[\,I\,]$ if and only if $\pi_{XY}(z) = 0$ and $\psi_{XZ}(w)\pi_{ZY}(z) = 0$. Writing out the XY block of the identity $(I_n + \psi(w))(I_n - \pi(w)) = I_n$ gives us that $\psi_{XY}(w) = 0$, whence the XZ block gives us that $\psi_{XZ}(w)\pi_{ZY}(z) = (I_{n_X} + \psi_{XX}(w))\pi_{XZ}(w)(I_{n_Z} - \pi_{ZZ}(w))^{-1}\pi_{ZY}(z)$. The reverse implication follows from the very same equations, first by showing that $\psi_{XY}(w) = 0$ and then concluding that $\psi_{XZ}(w)\pi_{ZY}(z) = 0$.

$\pi_{XZ}(z)(I_{n_Z} - \pi_{ZZ}(z))^{-1}\pi_{ZYj}$ is identically zero if and only if the first $n_Z(p + (n-1)q)$ terms are zero; but according to DR's Lemma 3.2 and Lemma A.3 that is equivalent to $\pi_{XYj}^{(h)} = 0$ for all $1 \le h \le n_Z(p + (n-1)q) + 1$ and all $j \ge 1$.

Step 2: $\pi_{XYj}^{(h)} = 0$ for all $j, h \ge 1$ if and only if it holds for $1 \le j \le p + (n-1)q$ and $1 \le h \le n_Z(p + (n-1)q) + 1$.

From step 1, $\pi_{XYj}^{(h)} = 0$ for all $j, h \ge 1$ if and only if it holds for $j \ge 1$ and $1 \le h \le n_Z(p + (n-1)q) + 1$. From (i) we know that for each h in the aforementioned range, the number of equations that must be solved is truncated at $p + (n-1)q$ and so the result follows. $\qquad\square$

Proof of Corollary 5.1. We prove only (i) – (iii); (iv) – (vi) follow similar arguments. (i) is equivalent to (ii), following the same line of argument as used in proving Corollary 4.1 (ii). To prove (ii) we must show that, $\sum_{\{j\ge1\}} \text{im}(\pi_{XYj}^{(h)}) = \sum_{\{1\le j\le m_1\}} \text{im}(\pi_{XYj}^{(h)})$ for all $h \ge 1$. Now it was shown in the proof of Theorem 5.1 that $\pi_{XY}^{(h)}(z)$ is representable as a ratio of a $(p+(n-1)q)$–order matrix polynomial and an nq–order polynomial. It follows by similar methods to those used in Lütkepohl (2006) pp. 426–429, that $\pi_{XY}^{(h)}(z) = C(I_{m_1} - Az)^{-1}Bz$ for some state–space representation (A, B, C) of the transfer function $\pi_{XY}^{(h)}(z)$ and $m_1 = n_X nq + n_Y(p + (n-1)q)$. Now the Cayley–Hamilton Theorem (see Theorem 2.4.2 in Horn & Johnson (1985)) implies that $\sum_{\{j\ge0\}} \text{im}(CA^j B) = \sum_{\{0\le j\le m_1-1\}} \text{im}(CA^j B)$. The result follows on noting that $\pi_{XYj}^{(h)} = CA^{j-1}B$ for $j \ge 1$.

To prove (iii) we will show that, $\sum_{\{h\ge1\}} \text{im}(\pi_{XYj}^{(h)}) = \sum_{\{1\le h\le m_2\}} \text{im}(\pi_{XYj}^{(h)})$ for all $j \ge 1$. In order to do that we define the operators, $\psi^{(j)}(w) = \sum_{h=1}^{\infty} \pi_j^{(h)} w^h$ for $j \ge 1$ and let each be partitioned as in Assumption 3. Then it follows from equation (4.5) that $\psi^{(1)}(w) = \psi(w)$. Moreover, from equation (4.4) we have that,

$$\psi^{(j+1)}(w) = w^{-1}\psi^{(j)}(w) - (I_n + \psi(w))\pi_j, \qquad j \ge 1,$$

which is similar in structure to equation (8.1). Therefore, following a similar line of argument to that used in the proof of Theorem 5.1 (i) we find that $\psi_{XY}^{(j)}(w)$ is representable by a ratio of polynomials, the numerator of degree $p(n-1)+q$ and the denominator (in this case, $\det(\phi(w))$) of degree pn. Now by a similar argument to that used in (i) we conclude that the first m_2 terms in $\{\pi_{XYj}^{(h)}\}_{h=1}^{\infty}$ span the space spanned by the entire collection for any $j \ge 1$. $\qquad\square$

Proof of Theorem 5.3. If $P_{\mathcal{M}}\widehat{W}$ is stable then all of the unstable roots of $\det(\phi(z))$ cancel out of the equation $\det(\phi(L))P_{\mathcal{M}}\widehat{W}(t) = P_{\mathcal{M}}\phi^*(L)\theta(L)a(t)$. Following a construction similar to that undertaken in the proof of Theorem 5.1 (ii), $P(P_{\mathcal{M}}\widehat{W}(t+h)|\widehat{W}(-\infty,t])$ can be expressed as a linear combination of a finite number of initial values of $P_{\mathcal{M}}\widehat{W}$'s and a's; stability is then equivalent to the asymptotic vanishing of the linear coefficients of the said expression. But this then is equivalent to $\mathbb{E}\|P(P_{\mathcal{M}}\widehat{W}(t+h)|\widehat{W}(-\infty,t])\|^2 \to 0$ as $h \to \infty$ for all $t \in \mathbb{Z}$. (ii) then follows from the law of iterated projections and the continuity of the projection operator. (iii) follows from (ii) by taking $J(t) = H_{-\infty}(W)$ for all $t \in \mathbb{Z}$. Finally, if $W \nrightarrow_\infty W|_{\mathcal{M}}\left[\,H_{-\infty}(W)\,\right]$ then $P(P_{\mathcal{M}}\widehat{W}(t+h)|\widehat{W}(-\infty,t]) \to 0$ in L^2 as $h \to \infty$ for all $t \in \mathbb{Z}$, which is equivalent – as we just saw – to the stability of $P_{\mathcal{M}}\widehat{W}$. $\qquad\square$

Proof of Theorem 6.1. We have already proven the "if" part; simply choose $Y \in \mathcal{T}$ to be a white noise process of positive definite variance matrix. The "only if" part follows from equation (6.4) and the fact that $\mathcal{U}$ must be orthogonal to $\sum_{\{j\geq 0\}}\mathrm{im}(CA^jB)$. $\qquad\square$

Proof of Theorem 6.2. By the Exhaustivity Theorem of DR, $X|_{\mathcal{U}^\perp} \nrightarrow_{(\infty)} X|_{\mathcal{U}}\left[\,I_{P_{\mathcal{U}}X}\,\right]$ is equivalent to $X|_{\mathcal{U}^\perp} \nrightarrow_1 X|_{\mathcal{U}}\left[\,I_{P_{\mathcal{U}}X}\,\right]$. Since $\mathcal{U}$ is the maximal subspace orthogonal to $\sum_{\{j\geq 0\}}\mathrm{im}(A^jB)$, Kalman's controllability decomposition (Lemma 3.3.3 of Sontag (1998)) implies that, $P_{\mathcal{U}}AP_{\mathcal{U}^\perp} = 0$. The result then follows from the fact that $\Delta_1^{P_{\mathcal{U}}XYI_{P_{\mathcal{U}}X}}(t) = P_{\mathcal{U}}AP_{\mathcal{U}^\perp}X(t)$. $\qquad\square$

Proof of Theorem 6.3. $\mathcal{C}_{(\infty)}^{XYI} = \mathbb{R}^{n_X}$ for all $Y \in \mathcal{T}$ if and only if X is controllable. The rest then follows by standard linear quadratic optimization methods (see e.g. Bertsekas (2001)). $\qquad\square$

References

Anderson, B. D. O. & Moore, J. B. (1979). *Optimal Filtering.* Englewood Cliffs, NJ, USA: Prentice-Hall, Inc.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics, 22,* 327351.

Artin, M. (1991). *Algebra.* New Jersey, US: Prentice Hall Inc.

Bertsekas, D. P. (2001). *Dynamic Programming and Optimal Control I and II, 3rd Edition.* Athena Scientific.

Beveridge, S. & Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics, 7*, 151–174.

Box, G. E. P. & Tiao, G. C. (1977). A canonical analysis of multiple time series. *Biometrika, 64*(2), 355–365.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory.* Classics in Applied Mathematics. Philadelphia, US: Society of Industrial and Applied Mathematics.

Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods, 2nd Edition.* Springer.

Bruneau, C. & Jondeau, E. (1999). Long-run causality, with an application to international links between long-term interest rates. *Oxford Bulletin of Economics and Statistics, 61*(4), 545–568.

Chung, K. L. (1974). *A Course in Probability Theory* (2 ed.). Probability and Mathematical Statistics: A Series of Monographs and Textbooks. San Diego: Academic Press.

Dufour, J.-M. & Renault, E. (1995). Short-run and long-run causality in time series: Theory. Cahiers de recherche 9538, Universite de Montreal, Departement de sciences economiques.

Dufour, J.-M. & Renault, E. (1998). Short run and long run causality in time series: Theory. *Econometrica, 66*(5), 1099–1125.

Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *Journal of Econometrics, 137*(2), 334–353.

Engle, R. F., Hendry, D. F., & Richard, J.-F. (1983). Exogeneity. *Econometrica, 51*(2), 277–304.

Florens, J. P. & Mouchart, M. (1982). A note on noncausality. *Econometrica, 50*(3), 583–91.

Geweke, J. (1984). Inference and causality in economic time series models. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of Econometrics*, volume 2 of *Handbook of Econometrics* chapter 19, (pp. 1101–1144). Elsevier.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*, 428–38.

Granger, C. W. J. (1980). Testing for causality : A personal viewpoint. *Journal of Economic Dynamics and Control, 2*(1), 329–352.

Granger, C. W. J. (1988a). Causality, cointegration, and control. *Journal of Economic Dynamics and Control, 12*(2-3), 551–559.

Granger, C. W. J. (1988b). Some recent development in a concept of causality. *Journal of Econometrics, 39*(1-2), 199–211.

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

Hansen, L. & Sargent, T. J. (2005). *Recursive Models of Dynamic Linear Economies.* Preprint. Available at: zia.stanford.edu/pub/sargent/webdocs/research/mbook2.pdf.

Hill, J. B. (2007). Efficient tests of long-run causation in trivariate var processes with a rolling window study of the money-income relationship. *Journal of Applied Econometrics, 22*(4), 747–765.

Hoover, K. (2001). *Causality in Macroeconomics.* Cambridge, UK: Cambridge University Press.

Horn, R. A. & Johnson, C. R. (1985). *Matrix Analysis.* Cambridge, United Kingdom: Cambridge University Press.

Hosoya, Y. (1991). The decomposition and measurement of the interdependency between second-order stationary processes. *Probability Theory and Related Fields,* (88), 429–444.

Hosoya, Y. (2001). Elimination of third-series effects and defining partial measures of causality. *Journal of Time Series Analysis, 22*(5), 537–554.

Jewell, N. P. & Bloomfield, P. (1983). Canonical correlations of past and future for time series: Definitions and theory. *The Annals of Statistics, 11*(3), 837–847.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

Kailath, T. (1980). *Linear Systems.* Englewood Cliffs, NJ: PrenticeHall.

Kolmogorov, A. N. (1941). Stationary sequences in a hilbert space. *Bull. Moscow State University, 2,* 1–40.

Lütkepohl, H. (1984). Linear transformations of vector arma processes. *Journal of Econometrics, 26*(3), 283–293.

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis.* Springer.

Otter, P. W. (1990). Canonical correlation in multivariate time series analysis with an application to one-year-ahead and multiyear-ahead macroeconomic forecasting. *Journal of Business and Economic Statistics, 8*(4), 453–457.

Otter, P. W. (1991). On wiener-granger causality, information and canonical correlation. *Economics Letters, 35,* 187–191.

Pearl, J. (2000). *Causality.* Cambridge, UK: Cambridge University Press.

Pitchford, J. D. & Turnovsky, S. J. (Eds.). (1976). *Applications of Control Theory to Economic Analysis.* Amsterdam: North-Holland.

Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory.* New York, USA: John Wiley and Sons Inc.

Preston, A. J. & Pagan, A. R. (1982). *The Theory of Economic Policy. Statics and Dynamics.* Cambridge: Cambridge University Press.

Reinsel, G. C. & Velu, R. P. (1998). *Multivariate Reduced-Rank Regression.* Lecture Notes in Statistics. New York, USA: Springer.

Sargent, T. J. & Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. Technical report.

Sontag, E. D. (1998). *Mathematical Control Theory: Deterministic Finite Dimensional Systems, 2nd Edition.* Texts in Applied Mathematics, Vol. 6. Springer Verlag. Available at: www.math.rutgers.edu/~sontag/.

Velu, R. P., Reinsel, G. C., & Wichern, D. W. (1986). Reduced rank models for multiple time series. *Biometrika*, *73*(1), 101–118.

Wiener, N. (1956). The theory of prediction. In E. F. Beckenback (Ed.), *Modern Mathematics for Engineers*, 1 chapter 8.

Yamamoto, T. & Kurozumi, E. (2006). Tests for long-run granger non-causality in cointegrated systems. *Journal of Time Series Analysis*, *27*(5), 703–723.