Geometric and Long Run Aspects of Granger Causality

Majid M. Al-Sadoon

January 2013

Barcelona GSE Working Paper Series

Working Paper nº 682

Geometric and Long Run Aspects of Granger Causality*

Majid M. Al-Sadoon

Universitat Pompeu Fabra & Barcelona GSE

January 31, 2013

Abstract

This paper extends multivariate Granger causality to take into account the subspaces along which Granger causality occurs as well as long run Granger causality. The properties of these new notions of Granger causality, along with the requisite restrictions, are derived and extensively studied for a wide variety of time series processes including linear invertible process and VARMA. Using the proposed extensions, the paper demonstrates that: (i) mean reversion in L^2 is an instance of long run Granger non–causality, (ii) cointegration is a special case of long run Granger non–causality along a subspace, (iii) controllability is a special case of Granger causality, and finally (iv) linear rational expectations entail (possibly testable) Granger causality restriction along subspaces.

JEL Classification: C10, C32, C51.

Keywords: Granger causality, long run Granger causality, L^2–mean–reversion, ρ–mixing, cointegration, VARMA, controllability, Kalman Decomposition, linear rational expectations.

*An earlier draft of this paper was titled, "Causality Along Subspaces: Theory." My most sincere thanks and gratitude go to Sean Holly for his help, support, and patience throughout the writing of this paper, which was part of the author's Phd thesis at the University of Cambridge. Thanks also go to M. Hashem Pesaran, Hal White, Richard J. Smith, George Kapetanios, Rod McCrorie, Cheng Hsiao, two anonymous referees, and to seminar participants at Cambridge University and St. Andrews University. This paper is dedicated to the memory of Halbert L. White, Jr. He is profoundly missed. Financial support of the Yousef Jameel scholarship is gratefully acknowledged.

1 Introduction

First suggested by Wiener (1956) and later developed by Granger (1969), Granger causality (GC) and Granger non–causality (GNC) are two of the most important concepts of time series econometrics. Many extensions have been proposed throughout the years: multivariate time series (Tjøstheim, 1981), enlarged information sets (Hsiao, 1982), variable horizons (Dufour & Renault, 1998), etc.[1] Yet problems of interpretation have plagued it since its inception (see e.g. Hamilton (1994)) and some have argued that it may fail to capture what is actually meant by causality (see Hoover (2001) or Pearl (2009)). Against this backdrop, the purpose of this paper is to demonstrate that GC is a much deeper concept than previously thought, going to the heart of many other concepts in time series analysis. This is done without taking any particular stance on the philosophical or empirical applicability of GC per se. Suffice it to say that GC remains an important element of causal analysis in a dynamic setting and that it does capture structural causality under certain conditions (White & Lu (2010), White, Chalak & Lu (2010), White & Pettenuzzo (2011), White, Al-Sadoon & Chalak (2012)). In such instances it makes sense to use causal language such as "cause" and "effect" in referring to variables associated by GC and we will on occasion do so in this paper with the understanding that those conditions are met.

This paper proposes two extensions to Dufour & Renault (1998) (DR): (i) it takes into account the subspaces of GC and (ii) it considers long run GC. To motivate the first extension, suppose that X and Y are multivariate 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 multivariate GNC tests may not give the full picture of the dependence structure. To motivate the second extension, frequency–domain results are available for checking long run GNC (Hosoya, 1991, 2001). There are also time–domain results for cointegrated VAR models (Granger & Lin, 1995; Bruneau & Jondeau, 1999; Yamamoto & Kurozumi, 2006). It would be useful to obtain time–domain criteria for long run GNC for a wider class of processes.

Based on the aforementioned extensions, it is shown that: (i) L^2–mean–reversion, a weaker form of weak dependence than ρ–mixing, is an instance of long run GNC, (ii) cointegration is a special case of long run GNC along a subspace, (iii) controllability is a special case of

[1] Excellent surveys can be found in Geweke (1984), Hamilton (1994), and Lütkepohl (2006).

subspace GC and Kalman's controllability decomposition is a partial converse of a result by DR, and finally (iv) linear rational expectations entail (possibly testable) GNC restriction along subspaces. Additionally, the paper presents extensions of various results by DR to subspace GNC in linear L^2 processes, including VARMA.

Now GC has been known to be associated with cointegration, controllability, and rational expectations equilibria for quite some time now. However these links have been established in rather restrictive contexts and do not span the full extents of the relationships. In particular, the association with cointegration was known to hold only in the context of bivariate models (Granger, 1988b), whereas we shall see that cointegration is a particular form of long run subspace GNC in any multivariate L^2 process. The association with controllability, on the other hand, was only shown in rather extreme forms of optimal control, where the policymaker cares only about a single variable in the model (Granger, 1988a). We will see that controllability in its most general form (see e.g. Kailath (1980)) is a particular instance of subspace GC. Finally, the association with rational expectations has been explored by Hansen & Sargent (1980) although in the highly specialized context of stochastic linear–quadratic control. They find that GC determines which variables ought to enter into the decision rule. In contrast, the result of this paper, which applies to a larger class of linear rational expectations model and any variable therein (whether or not it is a decision rule), is that the forward component of a rational expectations equilibrium lies within a particular subspace of GNC. It is important to emphasis that none of these results would have been possible without the two extensions proposed in this paper. The general theme of this paper, kindly noted by an anonymous referee, is therefore: "[Granger] causality is not invariant to linear projections onto alternative subspaces."

In addition to the above literature, various papers have considered time series dependence along subspaces. Velu et al. (1986) consider the problem of finding the subspaces along which a stationary VAR is forecastable. Otter (1990) considers the problem of finding the subspace of future variables predictable by past variables. Related to this work is Brillinger (2001), who 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. There are also a number of papers that 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 (Pearl, 2009). 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. Dufour & Taamouti (2010) develop measures of GC at finite horizons.

The paper proceeds as follows. Section 2 reviews some Hilbert space theory and sets the notation. Section 3 develops the main concepts of subspace and long run GNC as extensions to DR. Section 4 considers long run GNC in more detail. Section 5 specializes the theory to linear invertible processes. Section 6 specializes further to invertible VARMA processes. Section 7 considers the connection to controllability. Section 8 considers the connection to linear rational expectations equilibria. Section 9 concludes and Section 10 is an appendix.

2 Review of Hilbert Space Theory and Notation[2]

Throughout this paper, we work with a single probability space $(\Sigma, \mathscr{F}, \mathbb{P})$ with $\mathbb{E}$ as the expectation operator. We define L^2 to be the Hilbert space of random variables with finite second moments. The inner product is defined as $\langle X, Y\rangle = \mathbb{E}(XY)$ for all $X, Y \in L^2$ (we reserve $\|\cdot\|$ for the Euclidean vector norm and the norm it induces on matrices). We abuse notation by considering a random vector to be in L^2 if all its elements are in L^2. For a σ–algebra $\mathscr{X} \subseteq \mathscr{F}$, we take $L^2(\mathscr{X})$ to be the space of $\mathscr{X}$–measurable random variables in L^2.

If H and G are subspaces of L^2 then define $H + G = \overline{\mathrm{sp}}\{H, G\}$, the closure of the span of all linear combinations of the elements of G and H.[3] We set $H - G = H \cap G^{\perp}$, the part of H orthogonal to G. This subspace is closed whenever H is closed and is defined even when $G \cap H = \{0\}$, in which case $H - G = H$.

The time indexing set will be $(\omega, \infty) \subseteq \mathbb{Z}$ with $\omega \in \{-\infty\} \cup \mathbb{Z}$ for all processes in this paper. The information or history at time $t > \omega$ is denoted by $I(t)$, a closed subspace of L^2 satisfying $I(t) \subseteq I(t')$ whenever $\omega < t \leq t'$. $I = \{I(t) : t > \omega\}$ is an information set. If X is an n–dimensional stochastic process in L^2 then for $\omega \leq t < t'$ define, $X(t, t'] = \overline{\mathrm{sp}}\{X_{is} : t < s \leq t', 1 \leq i \leq n\}$. $X[t, t')$ is defined in similar fashion and for $\omega < t \leq t'$ and so is $X[t, t']$. The information set I is said to be conformable with X if $X(\omega, t] \subseteq I(t)$ for all $t > \omega$. The

[2]Excellent overviews of the applications of Hilbert space theory to time series analysis can be found in Brockwell & Davis (1991) and Pourahmadi (2001). This paper closely follows the notation of DR.

[3]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).

most frequently encountered information sets in this paper take 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 a closed subspace. When $I(t) = H$ for all $t > \omega$ we will refer to I as H. The remote information set of X is defined as $\bigcap_{t>\omega} X(\omega, t]$. We will also require $\mathscr{X}_T \subseteq \mathscr{F}$, the σ–algebra generated by $\{X(t) : t \in T\}$, where T is a subset of the time indexing set (ω, ∞).

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

3 Cartesian and Subspace Granger Causality

First, we consider the basic idea behind subspace and long run GC as an extension to Cartesian GC. This study is conducted within a large class of time series processes, namely L^2 processes.

We take the following to be our most basic assumption.

Assumption 1. Let $\omega, \varpi \in \{-\infty\} \cup \mathbb{Z}$ and $\omega \leq \varpi$. $X = \{X(t) : \omega < t < \infty\}$ and $Y = \{Y(t) : \omega < t < \infty\}$ are L^2 processes, of finite dimensions n_X and n_Y respectively. We also take I to be an information set. Here ω specifies the start of the process and ϖ specifies the start of the prediction period. $\square$

We will be interested in studying the predictability of X in terms of Y in the context of information set I. Because prediction sometimes requires initial conditions to be specified, this predictability is assessed over a range of periods (ϖ, ∞), which may be a proper subset of the time indexing set, (ω, ∞).[4] Typically, I is assumed to include all the variables that may help predict X, including X and excluding Y, thus the totality of information in I and Y consists of everything that may help predict X.[5] DR take I to include an auxiliary process Z through which Y may indirectly help predict X (see DR for further motivation and background). However, 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 processes.

[4] DR and Al-Sadoon (2009b) did not make this distinction clear. They derive their general L^2 results for the case $\omega = \varpi$ but when discussing linear invertible models they allow for $\omega < \varpi$. This then begs the question of whether their general L^2 results continue to hold for linear invertible processes with initial conditions.

[5] This larger information set is related to Hoover's (2001) idea of a "causal field" of X.

The following concept, due to Granger (1980), is the building block of GC.

Definition 3.1 (Predictive Effect)**.** Under Assumption 1, for $t + h > \omega$ and $t > \varpi$, define,

$$\Delta_h^{XYI}(t) = P(X(t+h)|I(t) + Y(\omega, t]) - P(X(t+h)|I(t)), \tag{3.1}$$

the horizon–h predictive effect of Y on X at time t when I is given. $\square$

The predictive effect is the change in the time–t prediction of $X(t+h)$ based on I when we include Y as a predictor. GC occurs when there is a difference between the two predictions for $h \geq 1$, otherwise we have GNC where Y does not help forecast X at horizon h given I.[6]

Definition 3.2 (Cartesian GNC)**.** Under Assumption 1, we have the following definitions:

(i) Y does not Granger–cause X given I at horizon h if $\Delta_h^{XYI}(t) = 0$ for all $t > \varpi$. We denote this by $Y \nrightarrow_h X\ [\,I\,]$.

(ii) Y does not Granger–cause X given I in the long run if $\Delta_h^{XYI}(t) \to 0$ in L^2 as $h \to \infty$ for all $t > \varpi$. We denote this by $Y \nrightarrow_\infty X\ [\,I\,]$.

(iii) Y does not Granger–cause X given I over the range of horizons $\mathcal{H}$ if $Y \nrightarrow_h X\ [\,I\,]$ for all $h \in \mathcal{H}$. We denote this by $Y \nrightarrow_{\mathcal{H}} X\ [\,I\,]$.

When it is clear from the context and there is no danger of confusion we drop the "given I" phrase in the above definitions. $\square$

When $Y \nrightarrow_h X\ [\,I\,]$, Y does not help predict X at horizon h. When $Y \nrightarrow_\infty X\ [\,I\,]$, the predictive effect dissipates in the long run, the limit taken in L^2 as this is the most natural mode of convergence in our setting. Of course, long run GNC does not rule out GC at finite horizons. Definition 3.2 (i) is due to DR although they require I to be conformable with X, which we do not. Definition 3.2 (iii) describes GNC over a range of horizons $\mathcal{H}$, typically an interval or half–line. Note that (iii) is derived from (i) and therefore inherits the properties of (i). For this reason, we focus most of our attention on (i) and (ii). In particular, (ii) generalizes Bruneau & Jondeau (1999) and Yamamoto & Kurozumi (2006) as they require $P(X(t+h)|I(t) + Y(\omega, t])$ and $P(X(t+h)|I(t))$ to have equal L^2 limits in h, whereas we do not require these limits to exist.

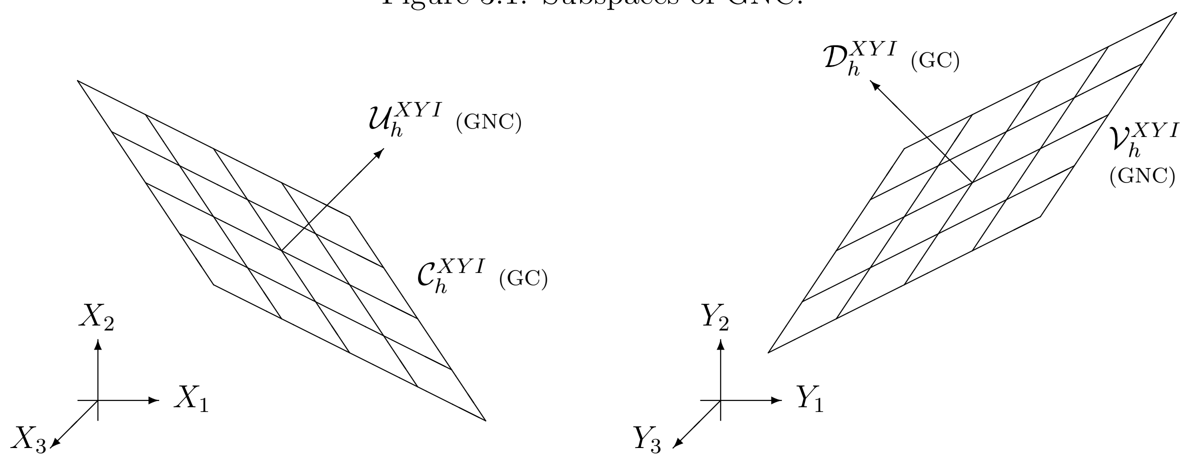
[6]This is similar to the idea of "screening off" in Hoover (2001) and Pearl (2009).

Example 3.1. Suppose X is a regular explosive AR(p) (see Assumption 2), Y is X, and $I(t) = X(\omega, t]$. Then clearly Y fails to Granger–cause X at any horizon, including the long run. However, it is not possible to formulate long run GNC according to the two aforementioned papers because the long run forecasts for X do not exist. □

We refer to the notions of GNC in Definition 3.2 as Cartesian GNC because they concern Cartesian components of the process $(X', Y')'$. Unfortunately, Cartesian GNC cannot capture the full range of dependence between X and Y.

Example 3.2. Let $n_X = 3$ and consider the left panel of Figure 3.1. Even if Y has a predictive effect on X at horizon h, it may be that it has an effect only along the subspace $\mathcal{C}_h^{XYI}$, while Y has no predictive effect on X along $\mathcal{U}_h^{XYI}$. This phenomenon may occur, for example, in a policy design framework where there are trade–offs between the X target variables *viz–a–viz* the Y policy instruments. In this case, $\mathcal{U}_h^{XYI}$ may emanate from a structural relationship between the X variables involving no Y's. There will therefore be policy prescriptions that allow the policymaker to hit any target in $\mathcal{C}_h^{XYI}$ but not targets in the direction of $\mathcal{U}_h^{XYI}$.

Figure 3.1: Subspaces of GNC.



Similarly, consider Y–space in the case $n_Y = 3$ in the right panel of Figure 3.1. It may be the case that variations of Y along $\mathcal{D}_h^{XYI}$ have a predictive effect on X but variations along $\mathcal{V}_h^{XYI}$ do not. In a policy design framework, this may occur if the policy instruments do not have independent effects on the target variables so that the policymaker effectively has less than n_Y instruments at his disposal. The policy design perspective is explored in greater detail in Section 7. □

To formalize this type of GNC then we define some new concepts.

Definition 3.3 (Subspace GNC)**.** Under Assumption 1, let $\mathcal{U} \subseteq \mathbb{R}^{n_X}$ and $\mathcal{V} \subseteq \mathbb{R}^{n_Y}$ be subspaces, and let $P_\mathcal{U}$ and $P_\mathcal{V}$ be orthogonal projection matrices onto $\mathcal{U}$ and $\mathcal{V}$ respectively, we have the following definitions:

(i) Y along $\mathcal{V}$ does not Granger–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 Granger–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 Granger–cause X along $\mathcal{U}$ given I over the range of horizons $\mathcal{H}$ if $P_\mathcal{V}Y \nrightarrow_\mathcal{H} P_\mathcal{U}X\ [I]$. We denote this by, $Y|_\mathcal{V} \nrightarrow_\mathcal{H} X|_\mathcal{U}\ [I]$.

When $\mathcal{U} = \mathbb{R}^{n_X}$ 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 will 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 when there is no danger of confusion. □

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

Lemma 3.1. Under Assumption 1, with $h \leq \infty$, $Y|_\mathcal{V} \nrightarrow_h X|_\mathcal{U}\ [I]$ if and only if $V'Y \nrightarrow_h U'X\ [I]$, where the columns of U are a basis for $\mathcal{U}$ and the columns of V are a basis for $\mathcal{V}$.

We are now ready to consider the properties of subspace GNC.

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

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

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

(iii) 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) 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 GNC across a range of horizons.

Lemma 3.2 gives the basic monotonicity ((i) and (ii)) and additivity ((iii) and (iv)) properties of subspace GNC. Intuitively, (i) implies that if Y fails to Granger–cause X then no linear function of Y can Granger–cause X. Likewise, (iv) implies that if Y fails to Granger–cause two different components of X, then it also fails to Granger–cause the two components jointly.

Lemma 3.2 generalizes Proposition 2.1 of DR in three directions. First, it considers all subspaces along which X and Y vary, whereas DR consider only the Cartesian components. Second, it considers long run GNC whereas DR consider only finite horizons. Third, DR require I to be conformable with $P_{\mathcal{U}}X$, which we do not.

Now for given processes X and Y, we would like to define what we mean by "the subspace along which X is not Granger–caused by Y" as well as the "the subspace along which Y fails to Granger–cause X" in line with the illustration in Figure 3.1. We do that using the notion of maximality: a subspace is maximal in the class of subspaces having a particular property if it is not properly contained in any other subspace having that property.

Lemma 3.3. Under Assumption 1, for $h \leq \infty$, the maximal subspace $\mathcal{U}$ along which $Y \nrightarrow_h X|_{\mathcal{U}}\ [\,I\,]$ exists and is unique. Similarly, the maximal subspace $\mathcal{V}$ along which $Y|_{\mathcal{V}} \nrightarrow_h X\ [\,I\,]$ also exists and is unique. An identical set of results hold for GNC across a range of horizons.

Lemma 3.3 proves existence and uniqueness of subspaces of GNC at any horizon for any two processes in L^2.[7] We can now formalize these subspaces in the following definition.

Definition 3.4 (Subspace of GC and GNC)**.** For $h \leq \infty$, define:

$\mathcal{U}_h^{XYI}$: the maximal subspace $\mathcal{U}$ such that $Y \nrightarrow_h X|_{\mathcal{U}}\ [\,I\,]$.

$\mathcal{C}_h^{XYI}$: the orthogonal complement of $\mathcal{U}_h^{XYI}$.

$\mathcal{V}_h^{XYI}$: the maximal subspace $\mathcal{V}$ such that $Y|_{\mathcal{V}} \nrightarrow_h X\ [\,I\,]$.

$\mathcal{D}_h^{XYI}$: the orthogonal complement of $\mathcal{V}_h^{XYI}$.

We also define, U_h^{XYI}, C_h^{XYI}, V_h^{XYI}, and D_h^{XYI} to be matrices whose columns are bases for $\mathcal{U}_h^{XYI}$, $\mathcal{C}_h^{XYI}$, $\mathcal{V}_h^{XYI}$, and $\mathcal{D}_h^{XYI}$ respectively. The subspaces $\mathcal{U}_{\mathcal{H}}^{XYI}$, $\mathcal{C}_{\mathcal{H}}^{XYI}$, $\mathcal{V}_{\mathcal{H}}^{XYI}$, and $\mathcal{D}_{\mathcal{H}}^{XYI}$ and the matrices $U_{\mathcal{H}}^{XYI}$, $C_{\mathcal{H}}^{XYI}$, $V_{\mathcal{H}}^{XYI}$, and $D_{\mathcal{H}}^{XYI}$ are defined similarly for GNC across horizons $\mathcal{H}$. □

Thus, at horizon h, Y Granger–causes X along $\mathcal{C}_h^{XYI}$ but not along $\mathcal{U}_h^{XYI}$. Likewise, variations of Y along $\mathcal{D}_h^{XYI}$ have a predictive effect on X at horizon h, but not variations

[7] Maximality can also be used to define the subspace associated with any feature of the process *à la* Engle & Kozicki (1993) so long as the feature is additive in the same sense as in Lemma 3.2 (iii) and (iv). Thanks to George Kapetanios for pointing me to this connection.

along $\mathcal{V}_h^{XYI}$. The columns of U_h^{XYI} are the linear combinations of the X's that are not predicted by Y at horizon h, while the columns of C_h^{XYI} are the linear combinations of the X's that are predicted by Y. Finally, the columns of V_h^{XYI} are the linear combinations of the Y's that have no predictive effect on X, while the columns of D_h^{XYI} are the linear combinations of the Y's that do have a predictive effect on X. Note that these and the other matrices listed in Definition 3.4 are unique modulo left multiplication by non–singular matrices. The reader may find it useful to consult Figure 3.1 as a summary of the notation in Definition 3.4, keeping in mind that calligraphic font refers to subspaces while italic capital letters refer to matrices.

We now have all the elements necessary for subspace and long run GNC. The gist of the rest of the paper is that a large variety of time series concepts can be understood in terms of the properties of the predictive effect process (3.1) by varying the cross sections of X, the cross sections of Y, the information set I, and the horizon h.[8]

4 Long Run Granger Causality

If the long run behavior of a series depends on its history, then the effects of some disturbances in its history never dissipate and are therefore permanent. If, on the other hand, the long run behavior of a series is unrelated to any past disturbance, the series is in a sense weakly dependent. Thus whether a series Granger–causes itself in the long run or not can be indicative of strong or weak dependence respectively. Here we develop this idea and consider its relationship to ρ–mixing and cotrendedness.

Suppose Assumption 1 holds with $I(t) = H = \text{sp}\{1\}$ for all $t > \varpi$, then $\Delta_h^{XXH}(t) = P(X(t+h)|H + X(\omega, t]) - \mathbb{E}(X(t+h))$. If for every $t > \varpi$ this quantity converges to zero as h tends to infinity, then X fails to Granger–cause itself in the long run and it is expected to revert to its mean in the long run. More generally, however, X may have a random effect and so we can think of its "mean" slightly more broadly.

Example 4.1. Let X be weakly stationary, of zero–mean, and let $\omega = -\infty$. Suppose the predictable part of its Wold decomposition is time invariant and has non–zero variance (Brockwell & Davis, 1991, Theorem 5.7.1). Then X has a random effect and will not revert to its mean; it reverts back to the random effect. If, on the other hand, H is set to the remote information

[8]Thanks to an anonymous referee for pointing this out.

set of X, then $\Delta_h^{XXH}(t)$ factors out the random effect and converges to zero with h. $\square$

This suggests the following definition.

Definition 4.1 (L^2–mean–reversion)**.** Under Assumption 1, let H be a closed subspace of L^2. If $\mathcal{U}_\infty^{XXH} = \mathbb{R}^{n_X}$, then we say that X is L^2–mean–reverting with respect to H. The subspace $\mathcal{U}_\infty^{XXH}$ is referred to as the subspace of L^2–mean–reversion of X with respect to H. When there is no danger of confusion, we drop the "with respect to H" phrase. $\square$

As we saw in the previous example, the choice of H in Definition 4.1 is not arbitrary but depends on the particular problem at hand. It is intended to remove any predictable components, such as deterministic trends and random effects.

L^2–mean–reversion is not only weaker than weak stationarity (as we saw in the previous example), it is also weaker than ρ–mixing, which is the most natural concept of mixing in an L^2 context (see Pourahmadi (2001) or Davidson (1994)). Recall that X in Assumption 1 is ρ–mixing if for all $t > \omega$ and $h \geq 1$, $\rho(h,t) = \sup\{\text{corr}(Y,Z) : Y \in L^2(\mathscr{X}_{(\omega,t]}), Z \in L^2(\mathscr{X}_{[t+h,\infty)})\}$ converges to zero as h goes to infinity.

Proposition 4.1 (L^2–mean–reversion and ρ–Mixing)**.** Under Assumption 1, if X is zero–mean, ρ–mixing, and L^2 bounded, then it is L^2–mean–reverting with respect to any closed subspace of its remote information set.

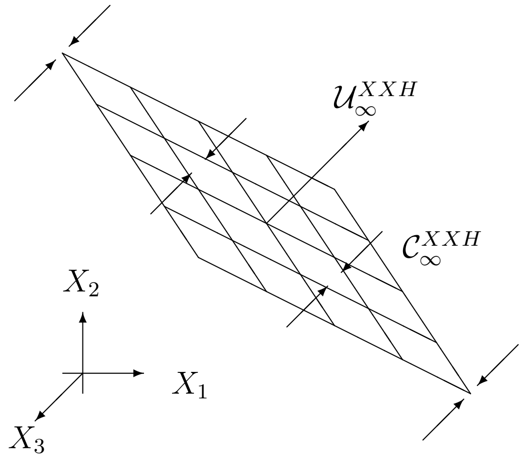
In our previous example, X was L^2–mean–reverting relative to the remote information set but not ρ–mixing because the predictable part of X remains a source of correlation between the past and the future at all forecast horizons.[9]

Further geometric insight can be gleaned from decomposing an L^2 process X into an L^2–mean–reverting process, $P_{\mathcal{U}}X$ and a process which is not L^2–mean reverting, $(I_{n_X} - P_{\mathcal{U}})X$, where $\mathcal{U} = \mathcal{U}_\infty^{XXH}$. We see in Figure 4.1 that deviations away from the plain $\mathcal{C}_\infty^{XXH}$ are predicted to dissipate in the long run, while variations along the plain are common non–mean–reverting processes that cancel out after appropriate linear combination. In this case U_∞^{XXH} may be interpreted as equilibrium relationships between the X variables. We will see in Section 6 that cointegration is a manifestation of L^2–mean–reversion.[10]

[9] A more interesting example of an L^2–mean–reverting series that is not α–mixing and therefore not ρ–mixing can be found in Andrews (1984).

[10] Figure 4.1 is a three–dimensional version of Figure 3.1 in Johansen (1995).

Figure 4.1: L^2–Mean–Reversion Along a Subspace.



We close this section with a generalization of a result by Granger (1988b) who shows that in a cointegrated bivariate VAR, at least one of the variables must Granger–cause the other. Granger comments that,

> "This is a somewhat surprising result, when taken at face value, as cointegration is concerned with the long run and equilibrium, whereas the causality in mean is concerned with short–run forecastability. However, what it essentially says is that for a pair of series to have an attainable equilibrium, there must be some causation between them to provide the necessary dynamics." (Granger, 1988b, p. 203)

A generalization to multivariate processes in L^2 is that for a multivariate L^2 process to be L^2–mean–reverting along a subspace, there must be subspace GC at every horizon.

Theorem 4.1 (L^2–Mean–Reversion and Long run Subspace GNC)**.** Under Assumption 1, suppose H is a closed subspace of L^2 and X is not L^2–mean–reverting with respect to H. Then for each $1 \leq h \leq \infty$, there are subspaces $\mathcal{C}, \mathcal{D} \subseteq \mathbb{R}^{n_X}$ such that $X|_{\mathcal{D}} \not\rightarrow_h X|_{\mathcal{C}}\ [\,H\,]$ fails.

Unlike Granger's result, ours does not require the representation theory of $I(1)$ processes and it considers all linear combinations of the process and not just its Cartesian components.

5 Linear Invertible Processes

We now specialize the theory to linear invertible L^2 processes. This allows us to extend results by DR on h–step GNC and develop new results for long run GNC.

We sharpen Assumption 1 by requiring that X and Y be components of a larger linear invertible process, W. The next set of assumptions is adopted from DR.

Assumption 2. $W = \{W(t) = (X'(t), Y'(t), Z'(t))' : t > \omega = -\infty\}$ is a stochastic process 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 representation,

$$W(t) = \mu(t) + \sum_{j=1}^{\infty} \pi_j W(t-j) + a(t), \qquad t > \varpi, \tag{5.1}$$

where the infinite series is assumed to converge in L^2 for all $t > \varpi$. When $\varpi > -\infty$, $\{W(t) : -\infty < t \leq \varpi\}$ is a given set of initial conditions.

We assume that $\mu(t) \in H$ for all $t > \varpi$, where H is a closed subspace of L^2. H is always included in the information sets we utilize so that μ is always predictable. It accounts for any deterministic trend or random effect in the data.

The innovations process $\{a(t) : t > \varpi\}$ is a sequence of uncorrelated random vectors in L^2, with $\mathbb{E}(a(t)) = 0$, $\mathbb{E}(a(t)a'(t)) = \Omega(t)$, and $a(t)$ is uncorrelated with $H + W(-\infty, t-1]$ for all $t > \varpi$. When $\Omega(t)$ is positive definite for all $t > \varpi$ we say that W is regular.

We will be interested in two types of GC tests:

(i) GC from Y to X. 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, $I(t) = H + X(-\infty, t] + P_{\mathcal{V}^\perp} Y(-\infty, t] + Z(-\infty, t]$ for $t > \varpi$, which consists of all available information excluding the contribution of variations in Y along $\mathcal{V}$.

(ii) GC from W to itself. Here we will assume that the subspaces $\mathcal{U}, \mathcal{V} \subseteq \mathbb{R}^n$ are given and work with the information set $J(t) = H + P_{\mathcal{V}^\perp} W(-\infty, t]$ for $t > \varpi$. Thus $J(t)$ includes all available information excluding the variation of W along $\mathcal{V}$.

It will be convenient to consider the demeaned process of W, which we denote by $\widehat{W} = \{\widehat{W}(t) = W(t) - P(W(t)|H) : t \in \mathbb{Z}\}$. Clearly, $\widehat{W}(t) = \sum_{j=1}^{\infty} \pi_j \widehat{W}(t-j) + a(t)$ for all $t > \varpi$. The demeaned process is partitioned conformably with W as $\widehat{W} = (\widehat{X}', \widehat{Y}', \widehat{Z}')'$. □

The class of processes in Assumption 2 includes stationary and non–stationary invertible VARMA (see e.g. Lütkepohl (2006)) and long–memory processes (see e.g. Palma (2007)). Lemma 6.4 of Pourahmadi (2001) provides a full characterization of the class of stationary processes with the representation (5.1).

The h–period–ahead forecasts of W are of the form,

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

We will refer to $\{\pi_j^{(h)}\}_{j=1}^{\infty}$ as the projection coefficient matrices at horizon h and we will partitioned them conformably with W as,

$$\pi_j^{(h)} = \begin{bmatrix} \pi_{XXj}^{(h)} & \pi_{XYj}^{(h)} & \pi_{XZj}^{(h)} \\ \pi_{YXj}^{(h)} & \pi_{YYj}^{(h)} & \pi_{YZj}^{(h)} \\ \pi_{ZXj}^{(h)} & \pi_{ZYj}^{(h)} & \pi_{ZZj}^{(h)} \end{bmatrix}, \qquad j, h \geq 1.$$

DR have shown that GNC from Y to X depends on the matrices $\{\pi_{XYj}^{(h)}\}_{j,h\geq 1}$. This remains true for subspace GNC.

Theorem 5.1 (Characterization of Subspace GNC at Horizon $h < \infty$)**.** Under Assumption 2 and for $1 \leq h < \infty$, $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\ [\,I\,]$ if $P_{\mathcal{U}} \pi_{XYj}^{(h)} P_{\mathcal{V}} = 0$ for all $j \geq 1$. The converse holds when W is regular.

Theorem 5.1 states that the generalization from Cartesian GNC to subspace GNC involves nothing more than linear restrictions on the projection coefficient matrices when $\mathcal{U}$ and $\mathcal{V}$ are known. On the other hand if we are interested in finding the subspaces of GNC then, according to the theorem, we have a reduced rank regression problem *à la* Anderson (1951) in which,

$$\begin{aligned} &\mathcal{U}_h^{XYI} \text{ is the left null space of the matrix } \begin{bmatrix} \pi_{XY1}^{(h)} & \pi_{XY2}^{(h)} & \cdots \end{bmatrix}. \\ &\mathcal{V}_h^{XYI} \text{ is the right null space of the matrix } \begin{bmatrix} \pi_{XY1}^{(h)} \\ \pi_{XY2}^{(h)} \\ \vdots \end{bmatrix}. \end{aligned} \tag{5.2}$$

The problem of estimating $\mathcal{V}_h^{XYI}$ can be seen as a variant of the problem considered by Sargent & Sims (1977). Their purpose is to extract from a large set of variables all of the information useful for prediction. Then $(D_h^{XYI})'Y$ is a set of "indexes" that summarize the information in Y useful for predicting X.

On the other hand,

$$\mathcal{U}^{XYI}_{[1,h]} \text{ is the left null space of the matrix } \begin{bmatrix} \pi^{(1)}_{XY1} & \cdots & \pi^{(h)}_{XY1} & \pi^{(1)}_{XY2} & \cdots & \pi^{(h)}_{XY2} & \cdots \end{bmatrix}.$$

$$\mathcal{V}^{XYI}_{[1,h]} \text{ is the right null space of the matrix } \begin{bmatrix} \pi^{(1)}_{XY1} \\ \vdots \\ \pi^{(h)}_{XY1} \\ \pi^{(1)}_{XY2} \\ \vdots \\ \pi^{(h)}_{XY2} \\ \vdots \end{bmatrix}. \tag{5.3}$$

We will see in the next section that in the case of VARMA processes, we can confine our rank tests to finite matrices rather than the infinite matrices above.

Remark 5.1. At the outset, Theorem 5.1 seems to generalize Theorem 3.1 of DR (the special case, $\mathcal{U} = \mathbb{R}^{n_X}$ and $\mathcal{V} = \mathbb{R}^{n_Y}$). However, a judicious choice of coordinates yields that it is in fact equivalent to DR's result. This follows from the fact that subspace GNC is nothing more than Cartesian GNC in a rotated coordinate system. With U and V is in Lemma 3.1 and $h \leq \infty$, $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}\ [I]$ if and only if $\widetilde{Y} \nrightarrow_h \widetilde{X}\ [I]$, where $\widetilde{Y} = V'Y$ and $\widetilde{X} = U'X$. Theorem 5.1 merely applies DR's Theorem 3.1 to the transformed series $\widetilde{W} = (\widetilde{X}', \widetilde{Y}', \widetilde{Z}')' = (X'U, Y'V, Z', X'U_\perp, Y'V_\perp)'$. The matrices $[U\ U_\perp]$ and $[V\ V_\perp]$ may be chosen to have orthonormal columns so that $W \mapsto \widetilde{W}$ is just a rotation. □

Long run GNC is more subtle to deal with than its finite horizon counterpart. Assumption 2 allows us to obtain necessary conditions for long run GNC but sufficiency requires stronger assumptions.

Theorem 5.2 (Characterization of Long Run Subspace GNC)**.** Under Assumption 2, $Y|_{\mathcal{V}} \nrightarrow_\infty X|_{\mathcal{U}}\ [I]$ if $\sup_{s \leq \tau} \mathbb{E}\|P_{\mathcal{V}}\widehat{Y}(s)\|^2 < \infty$ for some $\tau \in \mathbb{Z}$ and $\lim_{h\to\infty} \sum_{j=1}^{\infty} \|P_{\mathcal{U}}\pi^{(h)}_{XYj}P_{\mathcal{V}}\| = 0$. Conversely, if $Y|_{\mathcal{V}} \nrightarrow_\infty X|_{\mathcal{U}}\ [I]$ and W is regular, then $\lim_{h\to\infty} P_{\mathcal{U}}\pi^{(h)}_{XYj}P_{\mathcal{V}} = 0$ for all $j \geq 1$.

To ensure long run GNC we need the initial values along $\mathcal{V}$ of the demeaned Y series to be L^2–bounded as well as the uniform convergence of the projection coefficient matrices across all lags. The former condition is ensured under trend stationarity, for example, or if we assume that the initial values of $\widehat{W}$ are L^2–bounded. The latter condition can be ensured if $\lim_{h\to\infty} P_{\mathcal{U}}\pi^{(h)}_{XYj}P_{\mathcal{V}} = 0$ and this limit in h is interchangeable with summation over j.[11] We will see that these conditions are greatly simplified for VARMA processes.

Long run GC should be contrasted with "long run effects" from the dynamic multiplier literature (Lütkepohl, 2006, p. 392). Long run GC indicates long run predictability and, as we have seen in the last section, does not occur in weakly stationary processes. On the other hand, the "long run effect" of Y on X measures the response of X to a permanent change in Y and is given as $\sum_{j=1}^{\infty}\pi_{XYj}$, which may or may not be zero in a weakly stationary process.

Following the same line of argument as that used in DR's Theorem 3.2 and Remark 5.1, we can also find the necessary and sufficient restrictions for up to horizon h subspace GNC.

Theorem 5.3 (Characterization of Subspace GNC up to Horizon h)**.** Under Assumption 2 and for $h \geq 2$, $Y|_{\mathcal{V}} \nrightarrow_{[1,h]} X|_{\mathcal{U}}\ [I]$ if: (a) $P_{\mathcal{U}}\pi_{XYj}P_{\mathcal{V}} = 0$ for all $j \geq 1$ and (b) $P_{\mathcal{U}}\pi^{(k)}_{XX1}P_{\mathcal{U}^\perp}\pi_{XYj}P_{\mathcal{V}} + P_{\mathcal{U}}\pi^{(k)}_{XY1}P_{\mathcal{V}^\perp}\pi_{YYj}P_{\mathcal{V}} + P_{\mathcal{U}}\pi^{(k)}_{XZ1}\pi_{ZYj}P_{\mathcal{V}} = 0$ for all $j \geq 1$, $1 \leq k < h$. The converse holds when W is regular.

In fact, a much more general formulation of Theorems 5.1 – 5.3 is possible as we summarize in the following theorem.

Theorem 5.4 (General GNC in Linear Invertible Processes)**.** Under Assumption 2 and for $1 \leq h < \infty$,

(i) $W|_{\mathcal{V}} \nrightarrow_h W|_{\mathcal{U}}\ [J]$ if $P_{\mathcal{U}}\pi^{(h)}_j P_{\mathcal{V}} = 0$ for all $j \geq 1$. The converse holds when W is regular.

(ii) $W|_{\mathcal{V}} \nrightarrow_\infty W|_{\mathcal{U}}\ [J]$ if $\sup_{s\leq\tau}\mathbb{E}\|P_{\mathcal{V}}\widehat{W}(s)\|^2 < \infty$ for some $\tau \in \mathbb{Z}$ and $\lim_{h\to\infty}\sum_{j=1}^{\infty}\|P_{\mathcal{U}}\pi^{(h)}_j P_{\mathcal{V}}\| = 0$.[12] Conversely, if $W|_{\mathcal{V}} \nrightarrow_\infty W|_{\mathcal{U}}\ [J]$ and W is regular, then $\lim_{h\to\infty} P_{\mathcal{U}}\pi^{(h)}_j P_{\mathcal{V}} = 0$ for all $j \geq 1$.

(iii) $W|_{\mathcal{V}} \nrightarrow_{[1,h]} W|_{\mathcal{U}}\ [J]$ if: (a) $P_{\mathcal{U}}\pi_j P_{\mathcal{V}} = 0$ for all $j \geq 1$ and (b) $P_{\mathcal{U}}\pi^{(k)}_1 P_{\mathcal{U}^\perp}\pi_j P_{\mathcal{V}} = 0$ for all $j \geq 1$, $1 \leq k < h$. The converse holds when W is regular.

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

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

[11] For example, this is possible under the dominated convergence theorem if $\|P_{\mathcal{U}}\pi^{(h)}_{XYj}P_{\mathcal{V}}\| \leq r_j$ for all $j, h \geq 1$ for some sequence satisfying $\sum_{j=1}^{\infty} r_j < \infty$.

[12] Just as before, a sufficient condition for uniform convergence is $\lim_{h\to\infty} P_{\mathcal{U}}\pi^{(h)}_j P_{\mathcal{V}} = 0$ and $\|P_{\mathcal{U}}\pi^{(h)}_j P_{\mathcal{V}}\| \leq r_j$ for all $j, h \geq 1$, where r_j is summable.

Additionally, estimates of the subspaces of GNC can be obtained just as before by stacking the projection coefficient matrices appropriately and conducting the appropriate test.

The case $h = 1$ in Theorem 5.4 (i) has been studied by Box & Tiao (1977), Velu et al. (1986), and Otter (1990) in the context of stationary processes for the purpose of model reduction and improving forecasts. Here $(C_1^{WWJ})'W$ is predictable by current and past values of W but $(U_1^{WWJ})'W$ is not. On the other hand, $(D_1^{WWJ})'W$ helps predict W but $(V_1^{WWJ})'W$ does not. The other results are straightforward generalizations of previous results.

6 VARMA Processes

We next specialize the theory further to invertible VARMAs as these models are quite popular in empirical work and the specialization is rather elegant. We also discuss the relationship between VARMA stability and L^2–mean–reversion, considering cointegration as a special case.

We will require the following set of standard assumptions.

Assumption 3. With $\pi(z) = \sum_{j=1}^{\infty} \pi_j z^j$, let $I_n - \pi(z) = \theta^{-1}(z)\phi(z)$, where $\phi(z) = I_n - \sum_{j=1}^{p} \phi_j z^j$ and $\theta(z) = I_n + \sum_{j=1}^{q} \theta_j z^j$. □

Assumption 3 simplifies the structure of the projection coefficient matrices by assuming that $\pi(z)$ is rational. From linear system theory (Kailath, 1980; Hannan & Deistler, 1988), this implies that the projection coefficient matrices are recursively and finitely generated, allowing us to find truncation rules useful for empirical testing of GNC.

Theorem 6.1 (Truncation Rules for Subspace GNC in VARMA Processes I)**.** Under Assumptions 2 – 3 and for $1 \leq h < \infty$,

(i) $Y|_{\mathcal{V}} \not\rightarrow_h X|_{\mathcal{U}}\ [I]$ if $P_{\mathcal{U}} \pi_{XYj}^{(h)} P_{\mathcal{V}} = 0$ for all $1 \leq j \leq p + (n-1)q$.

(ii) $Y|_{\mathcal{V}} \not\rightarrow_\infty X|_{\mathcal{U}}\ [I]$ if $\lim_{h\to\infty} P_{\mathcal{U}} \pi_{XYj}^{(h)} P_{\mathcal{V}} = 0$ for all $1 \leq j \leq p + (n-1)q$.

(iii) $Y|_{\mathcal{V}} \not\rightarrow_{[1,\infty)} X|_{\mathcal{U}}\ [I]$ if $P_{\mathcal{U}} \pi_{XYj}^{(h)} P_{\mathcal{V}} = 0$ for all $1 \leq j \leq p + (n-1)q$ and $1 \leq h \leq (p + (n-1)q)(n - \dim(\mathcal{U}) - \dim(\mathcal{V})) + 1$.

Theorem 6.1 (i) and (iii) generalize DR's Proposition 4.5, which concerns Cartesian GNC in VARs. Theorem 6.1 specializes the sufficiency parts of Theorems 5.1 – 5.3 to the VARMA case. When the subspaces are known, the restrictions above are easily tested as linear restrictions of the kind considered in (Lütkepohl, 2006, section 12.4). On the other hand if we are interested in finding the subspaces of GNC at horizon $h < \infty$, then we test the rank of the appropriate matrix in (5.2) or (5.3) truncated at the $p+(n-1)q$ block (Al-Sadoon, 2009a, 2011). Long run subspace GNC testing requires more care since it involves restrictions on matrices which may not be defined in the limit. However, for cointegrated VARMAs, the limits of the projection coefficient matrices do exist (Johansen, 1995) and so,

$$\begin{aligned}
&\mathcal{U}_\infty^{XYI} \text{ is the left null space of the matrix } \begin{bmatrix} \lim_{h\to\infty}\pi_{XY1}^{(h)} & \cdots & \lim_{h\to\infty}\pi_{XYp+(n-1)q}^{(h)} \end{bmatrix}. \\
&\mathcal{V}_\infty^{XYI} \text{ is the right null space of the matrix } \begin{bmatrix} \lim_{h\to\infty}\pi_{XY1}^{(h)} \\ \vdots \\ \lim_{h\to\infty}\pi_{XYp+(n-1)q}^{(h)} \end{bmatrix}.
\end{aligned} \tag{6.1}$$

The Cartesian GNC variants of these tests for VARs have been proposed by Bruneau & Jondeau (1999) and Yamamoto & Kurozumi (2006).

The truncation rules for GNC tests from W to itself are summarized in the next result.

Theorem 6.2 (Truncation Rules for Subspace GNC in VARMA Processes II)**.** Under Assumptions 2 – 3 and for $1 \leq h < \infty$,

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

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

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

We now turn to the relationship between stability in VARMA and L^2–mean–reversion. Suppose the columns of U are a basis for the subspace $\mathcal{U} \in \mathbb{R}^n$, then $U'\widehat{W}$ satisfies a vector difference equation driven by a that can be described as being stable or unstable according to whether its autoregressive part has all its roots outside the closed unit disk or not (Lütkepohl, 1984; Hannan & Deistler, 1988). Under stability, the effect of any initial conditions dissipates in the long run and so we should expect that stability of $U'\widehat{W}$ implies L^2–mean–reversion along $\mathcal{U}$ with respect to H. The converse is not true in general. For example, with zero initial conditions, $\widehat{W}$ may be L^2–mean–reverting yet have unstable modes that a is unable to excite due to the singularity of $\Omega(t)$.

Theorem 6.3 (Stability and Long run Subspace GNC in VARMA Processes)**.** Under Assumptions 2 – 3, with subspace $\mathcal{U} \subseteq \mathbb{R}^n$ and $U \in \mathbb{R}^{n \times u}$ defined as in Lemma 3.1, if $U'\widehat{W}$ is stable then $\mathcal{U} \subseteq \mathcal{U}_{\infty}^{WWH}$. The converse holds when W is regular.

As is well known, if W is a cointegrated VARMA, the cointegrated component of W satisfies a stable vector difference equation driven by a (Johansen, 1995; Lütkepohl, 2006). Therefore, under regularity, the cointegration space is exactly the subspace of L^2–mean–reversion.

7 Controllability[13]

We now consider the relationship between controllability and subspace GNC. Here we show that non–controllability is a special case of subspace GNC at all forecast horizons. We also find that Kalman's celebrated controllability decomposition theorem is a partial converse to DR's separation condition.

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. Examples include linear systems (Kailath, 1980; Hannan & Deistler, 1988), Kalman filters (Anderson & Moore, 1979), linear quadratic control (Hansen & Sargent, 2005), and economic policy (Pitchford & Turnovsky, 1976; Preston & Pagan, 1982). Here, we consider the model most commonly encountered in the literature.

Assumption 4. Let $Y = \{Y(t) \in \mathbb{R}^{n_Y} : t \leq 0\} \subset L^2$ be policy variables chosen by the policymaker. Let $X = \{X(t) \in \mathbb{R}^{n_X} : t \geq 0\}$ consist of target variables generated as,

$$\begin{aligned} Z(t) &= AZ(t-1) + BY(t-1) + \varepsilon(t), && t > 0 \\ X(t) &= CZ(t) + \eta(t), && t \geq 0. \end{aligned}$$

We assume that ε and η are white noise processes 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 (for simplicity) $\mathbb{E}(Z(0)) = 0$. The purpose of the policymaker is to choose a sequence of Y's to pursue some

[13] It may interest the reader to know that the problem of subspace GC was initially approached from this perspective – i.e. from an attempt to test for controllability in linear optimal control problems.

objective (e.g. minimizing a loss function). We have,

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

We will work with the information set $H = \{0\}$ and denote by $\mathscr{T}$ the class of L^2 processes orthogonal to $Z(0)$, ε, and η. Here we take $\omega = -1$ and $\varpi = 0$. □

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, the range of expected values of X that are reachable by some choice of Y is the image of the sequence of matrices $\{CA^jB\}_{j=0}^{\infty}$. 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. The left null space of this matrix is the unreachable part of the control system. Thus, the system is controllable (in the sense that any target is reachable in expectation) if and only if the output controllability matrix is of full rank.[14]

In contrast, the theory of GC allows us to approach the problem from a different point of view. The deterministic Y considered by the engineering literature is just one among many possible exogenous processes that allow us to investigate the causal effect of Y on X. $Y \in \mathscr{T}$ is akin to a randomly assigned treatment from which the causal effect on X can be directly measured by the predictive effect (see e.g. White & Lu (2010)). For $Y \in \mathscr{T}$, $\Delta_h^{XYH}(t) = \sum_{j=0}^{t+h-1} CA^jBP(Y(t+h-j-1)|Y(\omega,t])$ for $t > 0$ and so the causal effect of Y on X is clearly along the image of the output controllability matrix.

Theorem 7.1. Under Assumption 4, the subspace $\mathcal{U} \subseteq \mathbb{R}^{n_X}$ is unreachable if and only if $\mathcal{U} \subseteq \mathcal{U}_{[1,\infty)}^{XYH}$ for all $Y \in \mathscr{T}$. Thus the unreachable part of the control system is $\bigcap_{Y\in\mathscr{T}} \mathcal{U}_{[1,\infty)}^{XYH}$.

It follows that the system is controllable if and only if $\mathcal{C}_{[1,\infty)}^{XYH} = \mathbb{R}^{n_X}$ for some $Y \in \mathscr{T}$. Therefore controllability is a special case of GC. Since controllability is necessary for linear quadratic optimal control (Hansen & Sargent, 2005), it follows that this particular form of GC is necessary for the existence of optimal controls. This result should be contrasted with Granger (1988a) who showed the necessity of GC only for a subclass of control problems,

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

where the policymaker gives zero weight to all variables except for one – we place no such *a priori* restrictions.

The connection between GC and controllability is still more intimate. Suppose that Z is measurable directly and with no error (i.e. $C = I_{n_X}$ and $\eta = 0$). Then Kalman's controllability decomposition states that X is decomposable into two parts, one influenced directly by Y and another which is influenced by Y neither directly nor indirectly (through the first component). In particular, it is possible to write the system as,

$$\begin{bmatrix} \widetilde{X}_1(t) \\ \widetilde{X}_2(t) \end{bmatrix} = \begin{bmatrix} * & * \\ 0 & * \end{bmatrix} \begin{bmatrix} \widetilde{X}_1(t-1) \\ \widetilde{X}_2(t-1) \end{bmatrix} + \begin{bmatrix} * \\ 0 \end{bmatrix} Y(t-1) + \begin{bmatrix} * \\ * \end{bmatrix} \varepsilon(t), ^{15}$$

where $\widetilde{X}_1$ is the component of X along the reachable subspace of the system and $\widetilde{X}_2$ is the component along the unreachable subspace. Thus starting with $Y \nrightarrow_{[1,\infty)} \widetilde{X}_2 \; [I_{\widetilde{X}_2}]$, Kalman concludes that $(Y', \widetilde{X}_1')' \nrightarrow_1 \widetilde{X}_2 \; [I_{\widetilde{X}_2}]$, where $I_{\widetilde{X}_2}(t) = \widetilde{X}_2(\omega, t]$. On the other hand, a special case of DR's separation condition (their Proposition 2.4) is that if $(Y', \widetilde{X}_1')' \nrightarrow_1 \widetilde{X}_2 \; [I_{\widetilde{X}_2}]$ then $Y \nrightarrow_{[1,\infty)} \widetilde{X}_2 \; [I_{\widetilde{X}_2}]$. Thus, Kalman's controllability decomposition is, quite simply, a partial converse of the separation condition of DR.

8 Linear Rational Expectations Equilibria

The final section of this paper considers the subspace GNC restrictions entailed by rational expectations. We also show that the forward component of a linear rational expectations equilibrium (LREE) model always lies inside a subspace of GNC.

We employ a slightly specialized variant of the Sims (2002) formulation.

Assumption 5. Suppose $Y = \{Y(t) \in \mathbb{R}^{n_Y} : t \geq 0\}$ is a Gaussian process with L^2–bounded forecasts, i.e. $\sup_{h \geq 1} \|P(Y(t+h)|Y(\omega, t])\|^2 < \infty$ for all $t \geq 0$. Let $X_0 \in \mathbb{R}^{n_X}$ be a given Gaussian initial condition. And let $A, B, C \in \mathbb{R}^{n_X \times n_X}$, $D \in \mathbb{R}^{n_X \times n_Y}$. A solution to the LREE model,

$$AX(t) = BX(t-1) + C\eta(t) + DY(t), \qquad t > 0, \tag{8.1}$$

$$X(0) = X_0 \tag{8.2}$$

[15]The symbol '$*$' represents possibly non–zero blocks.

is a nonexplosive linear process $X = \{X(t) \in \mathbb{R}^{n_X} : t \geq 0\}$ satisfying (8.1) – (8.2), where $\eta(t)$ consists of expectational errors satisfying $P(\eta(t+1)|I(t)) = 0$ for all $t > 0$, and $I(t) = \overline{\text{sp}}\{X_0\} + Y(\omega, t]$ is conformable with X. We assume the system is well–specified so that it is impossible to eliminate the X variables.[16] Here we take, $\omega = -1$, $\varpi = 0$, and $H = \{0\}$. $\square$

Assumption 5 specializes Sims' assumptions in two ways: (i) it require all exogenous variables to be Gaussian and utilizes best linear predictors instead of conditional expectations. There is no loss of generality here as best linear predictors are conditional expectations under Gaussianity and these assumptions allows us to use the L^2 theory we have developed above. Our results continue to hold for non–Gaussian exogenous variables if we substitute the projection operators by conditional expectation operators. (ii) it excludes explosive solutions – although unit roots are allowed. Extensions to more general growth conditions are straightforward (see Sims (2002)). In fact, both specializations are very common in practice.

Some GC relations are immediately evident from (8.1). Using the fact that $P(\eta(t+1)|I(t)) = 0$, we can write $AP(X(t+1)|I(t)) = BX(t) + DP(Y(t+1)|I(t))$ for $t > 0$. It follows that conditional on the one–period–ahead forecasts of Y and BX, X along $\text{im}(B'_\perp)$ fails to Granger–cause X along $\text{im}(A')$. Now this result may not be particularly enlightening as Y is often unobservable. However, if there are equations with no unobservable variables and S is a selection matrix for these equations, then $X|_{\text{im}((SB)'_\perp)} \nrightarrow_1 X|_{\text{im}(A'S')} \; [I_{(SB)X}]$, where $I_{(SB)X}(t) = SBX(\omega, t]$. If observable variables are excluded from the equations selected by S (i.e. $(SB)'_\perp X$ includes observable variables), then we can test some of the GNC implication of LREE by testing the GNC of these excluded observable variables on SAX.

Example 8.1. Consider the textbook LREE model,

$$X_1(t) + X_2(t) = (1+r)X_1(t-1) + Y(t)$$
$$X_2(t) = X_2(t-1) + \eta(t).$$

Here X_1 is the capital stock, X_2 is consumption, η is an expectational error, Y is income (all expressed as deviations from steady state), and $r > 0$ is the interest rate (Hansen & Sargent, 2005). The second equation implies that $P(X_2(t+1)|I(t)) = X_2(t)$, which contains no unobservable variables. Therefore, given past consumption, any other variable will fail to

[16]Technically, A and B have no zero eigenvalues corresponding to the same eigenvector.

Granger–cause consumption. This result is due to Hall (1978), although he did not state it in terms of GC. □

We can glean more GNC implications by solving the model. Sims finds a transformation $\widetilde{X} = Q'X$, where Q is an orthogonal matrix such that the system decouples into two components, one having nonexplosive roots and therefore iterated backwards and another having explosive roots and therefore iterated forwards. The first component is called the backward component, while the second is called the forward component. Subject to existence conditions, Sims (2002) obtains the class of all solutions as,

$$\begin{bmatrix} I_k & * \\ 0 & I_{n_X-k} \end{bmatrix} \begin{bmatrix} \widetilde{X}_1(t) \\ \widetilde{X}_2(t) \end{bmatrix} = \begin{bmatrix} * & * \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{X}_1(t-1) \\ \widetilde{X}_2(t-1) \end{bmatrix} + \begin{bmatrix} * \\ 0 \end{bmatrix} Y(t) + \begin{bmatrix} * \\ 0 \end{bmatrix} \eta(t)$$
$$+ \begin{bmatrix} 0 \\ * \end{bmatrix} \sum_{s=t+1}^{\infty} F^{s-t-1} G P(Y(s)|I(t)), \quad t > 0, \tag{8.3}$$

where $\widetilde{X}_1 = Q_1'X$ is the backward component, $\widetilde{X}_2 = Q_2'X$ is the forward component, and F is a stable matrix.[17] The solution is unique if and only if η disappears in (8.3); that is, when uniqueness fails, an arbitrary innovation process (a sunspot) may influence the system.

Regardless of the uniqueness of X, $\widetilde{X}_2$ is determined uniquely by the forecasts of Y. In practice, Y consists of weakly dependent structural shocks (Woodford, 2003), making $\widetilde{X}_2$ itself weakly dependent. This then imparts further GNC restrictions along subspaces as delineated in the following result.

Theorem 8.1. Under Assumption 5, suppose a solution (8.3) exists. If Y is white noise, then $\mathrm{im}(Q_2) \subseteq \mathcal{V}_{[1,\infty)}^{XXH} \cap \mathcal{U}_{[1,\infty)}^{XXH}$. More generally, if Y is L^2–mean–reverting, then $\mathrm{im}(Q_2) \subseteq \mathcal{U}_{\infty}^{XXH}$.

It follows that for a LREE driven by a weakly dependent Y, the forward component is always inside a subspace of GNC. Thus some subspaces of GNC may be generated by the LREE cross–equation restrictions in a nontrivial way. In particular, if X is not L^2–mean–reverting but Y is, the subspace of L^2–mean–reversion of X contains the forward component of the model. Hansen & Sargent (2005) and Juselius (2008) have developed similar results on the cointegration relationships generated by LREE in highly restricted LREE models. Theorem

[17] The infinite sum in the last expression exists in L^2 because F is stable and the forecasts of Y are bounded in L^2. The importance of the boundedness of the forecasts is discussed further in Onatski (2006).

8.1, on the other hand, makes no assumptions on the structure of the system and is concerned with further GNC implications of LREE than just cointegration.

Example 8.2. The unique solution to the model in Example 8.1 is,

$$X_1(t) = X_1(t-1) + \frac{1}{1+r}Y(t) - \frac{r}{1+r}\sum_{s=t+1}^{\infty}\left(\frac{1}{1+r}\right)^{s-t} P(Y(s)|I(t))$$
$$X_2(t) = rX_1(t-1) + \frac{r}{1+r}Y(t) + \frac{r}{1+r}\sum_{s=t+1}^{\infty}\left(\frac{1}{1+r}\right)^{s-t} P(Y(s)|I(t)).$$

As a linear process driven by Y and its forecasts, X is unstable as it's autoregressive part has a unit root. However, there is a linear combination of the X's which may be weakly dependent if Y is. If Y is white noise, then $rX_1(t) - X_2(t) = 0$ for all $t > 0$ and so X does not vary along $[\ r \quad 1\ ]'$. Thus, $rX_1 - X_2$ cannot help forecast X and X cannot improve upon a forecast of zero for $rX_1 - X_2$. Formally, $\operatorname{im}\left([\ r \quad 1\ ]'\right) \in \mathcal{V}_{[1,\infty)}^{XXH} \cap \mathcal{U}_{[1,\infty)}^{XXH}$. On the other hand, if Y is L^2–mean–reverting, then $rX_1(t) - X_2(t) = -r\sum_{s=t+1}^{\infty}\left(\frac{1}{1+r}\right)^{s-t-1} P(Y(s)|I(t))$ is always forecasted to revert to its mean in the long run and so $\operatorname{im}\left([\ r \quad 1\ ]'\right)$ is inside $\mathcal{U}_{\infty}^{XXH}$. In particular, if Y is a weakly stationary VARMA process, then $\operatorname{im}\left([\ r \quad 1\ ]'\right)$ is inside the cointegration space of X. □

9 Conclusion

This paper has presented geometric and long run aspects of GC within L^2 processes and some of it's subclasses. These aspects elucidate the connection between GNC on the one hand and L^2–mean–reversion, cointegration, controllability, and linear rational expectations on the other hand. They also allow for extensions of various GNC tests proposed by DR for linear invertible processes including VARMA. The geometric and long run aspect of GNC also open the door for many venues of future research as we now discuss.

First, subspace GNC can be seen as reduced rank regression applied to GC analysis. This begs the question of whether it can be developed from a canonical correlations point of view as well.[18] One would suspect that $\mathcal{U}_h^{XYI}$ and $\mathcal{V}_h^{XYI}$ are subspaces associated with zero canonical correlations and this is indeed the case for simple processes such as VARs (Al-Sadoon, 2010).[19]

[18] Reinsel & Velu (1998) discuss the relationship between canonical correlations and reduced rank regression.

[19] Otter (1991) used canonical correlation analysis to study Cartesian GC. His analysis could be adapted to study subspace GNC in stationary multivariate series.

It remains to be seen whether it can be developed under the most rudimentary Assumption 1. Similarly, it would be of interest to study the subspaces associated with canonical correlations of one, particularly its relation to long run subspace GC (Hannan & Poskitt, 1988; Poskitt, 2000).[20]

Second, the exclusion of subspaces of GNC 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 subspace GC can be further applied for model reduction. In the same vain, it would be interesting to see how Bayesian analysis can be conducted using meaningful subspace GNC priors.

Third, subspace GNC is concerned with the predictability and predictive effects of linear cross–sectional weighted averages of time series data (Lemma 3.1), we can therefore begin to think about whether it is possible to extend the theory to allow for the analysis of the predictability and predictive effects of non–linear cross sectional functions of the data. Such a theory would be applicable to testing the GNC implications of non–linear Euler equations, just as we saw in the linear case in Example 8.1. It would also be applicable to non–linear cointegration analysis where a non–linear function of the data is stationary and so predicted to revert back to its mean in the long run. In this regard, the non–linear approaches to GC of Engle et al. (1983) and Florens & Mouchart (1982) would be particularly useful.

Finally, we have introduced a new concept of long run GNC, which encompasses the concepts of Bruneau & Jondeau (1999) and Yamamoto & Kurozumi (2006). There is, however, a frequency–domain concept of long run causality (Granger & Lin (1995), Hosoya (1991), and Hosoya (2001)). It would be fruitful to clarify the extent of overlap between the two concepts and compare the testing procedures proposed by each perspective.

10 Appendix

Proof of Lemma 3.1. Recall that $P_{\mathcal{U}} = U(U'U)^{-1}U'$ and $P_{\mathcal{V}} = V(V'V)^{-1}V'$ (see e.g. Theorem 2.5.1 of Brockwell & Davis (1991) and the subsequent remarks). Clearly, $P_{\mathcal{V}}Y(\omega,t] = V'Y(\omega,t]$. Now for $h < \infty$, $\Delta_h^{P_{\mathcal{U}}XP_{\mathcal{V}}YI}(t) = P_{\mathcal{U}}\Delta_h^{XP_{\mathcal{V}}YI}(t) = U(U'U)^{-1}U'\Delta_h^{XV'YI}(t)$, which is zero if and only if $U'\Delta_h^{XV'YI}(t) = \Delta_h^{U'XV'YI}(t)$ is zero. For the long run case simply substitute

[20]Thanks are due to an anonymous referee for raising this point.

"is zero" in the last sentence with "goes to zero in L^2." $\square$

Proof of Lemma 3.2. We prove the case of GNC at horizon h. The case of GNC across a range of horizons 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,

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

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 GNC 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. □

Proof of Lemma 3.3. We prove only the case of GNC at horizon h. The case of GNC across a range of horizons follows a similar argument. To prove existence consider the collection of all subspaces $\mathcal{U}$ such that $Y \nrightarrow_h X|_{\mathcal{U}}\ [I]$ 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.[21] Uniqueness is proven by noting that if $\mathcal{U}_1$ and $\mathcal{U}_2$ are maximal then by Lemma 3.2 (iv) again $Y \nrightarrow_h X|_{\mathcal{U}_1+\mathcal{U}_2}\ [I]$. 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 for $\mathcal{V}$ follows a similar argument, utilizing Lemma 3.2 (iii). □

[21] Artin (1991) provides a number of examples of uses of Zorn's lemma in algebra.

Proof of Proposition 4.1. Let H be a closed subspace of the remote information set of X.

$$\begin{aligned}
\mathbb{E}\left\|\Delta_h^{XXH}(t)\right\|^2 &= \sum_{i=1}^{n_X}\langle P(X_i(t+h)|X(\omega,t]-H), P(X_i(t+h)|X(\omega,t]-H)\rangle \\
&= \sum_{i=1}^{n_X}\langle X_i(t+h), P(X_i(t+h)|X(\omega,t]-H)\rangle \\
&\leq \sum_{i=1}^{n_X}\rho(h,t)\left(\mathbb{E}|X_i(t+h)|^2\right)^{\frac{1}{2}}\left(\mathbb{E}|P(X_i(t+h)|X(\omega,t]-H)|^2\right)^{\frac{1}{2}} \\
&\leq \sum_{i=1}^{n_X}\rho(h,t)\mathbb{E}|X_i(t+h)|^2 \\
&\leq \rho(h,t)\sup_{s>\omega}\mathbb{E}\|X(s)\|^2.
\end{aligned}$$

The first equality follows from the fact that for any closed subspaces $H_1 \subseteq H_2$, $P(X|H_2) - P(X|H_1) = P(X|H_2 - H_1)$ for all $X \in L^2$ (Pourahmadi, 2001, Theorem 9.18). The second equality follows from the fact that the projection operator is self–adjoint (Pourahmadi, 2001, Theorem 9.17 (a)). The first inequality follows from the definition of ρ–mixing. The second inequality follows from the properties of the projection operator. The third inequality follows from the L^2 boundedness of X. □

Proof of Theorem 4.1. According to DR's Proposition 2.3, if $\mathcal{U}_1^{XXH} = \mathbb{R}^{n_X}$, then $\mathcal{U}_{[1,\infty)}^{XXH} = \mathbb{R}^{n_X}$ and so $\mathcal{U}_\infty^{XXH} = \mathbb{R}^{n_X}$. It follows that $\mathcal{C}_h^{XXH} \neq \{0\}$ for every $1 \leq h \leq \infty$. Now choose $\mathcal{C} = \mathcal{C}_h^{XXH}$ and $\mathcal{D} = \mathbb{R}^{n_X}$. □

Proof of Theorem 5.1. Follows from Theorem 3.1 of DR and Remark 5.1. □

Proof of Theorem 5.2. We will prove the Cartesian version of the theorem (i.e. the case $\mathcal{U} = \mathbb{R}^{n_X}$ and $\mathcal{V} = \mathbb{R}^{n_Y}$). The subspace case then follows from Remark 5.1.

Noting that for all $t > \varpi$,

$$\Delta_h^{XYI}(t) = \sum_{j=1}^{\infty}\pi_{XYj}^{(h)}\{Y(t+1-j) - P(Y(t+1-j)|I(t))\},$$

we have,

$$\begin{aligned}
\mathbb{E}\|\Delta_h^{XYI}(t)\|^2 &= \mathbb{E}\left\|\sum_{j=1}^{\infty}\pi_{XYj}^{(h)}\Delta_{1-j}^{YYI}(t)\right\|^2 \\
&\leq \mathbb{E}\left(\sum_{j=1}^{\infty}\|\pi_{XYj}^{(h)}\Delta_{1-j}^{YYI}(t)\|\right)^2 \\
&\leq \mathbb{E}\left(\sum_{j=1}^{\infty}\|\pi_{XYj}^{(h)}\|\|\Delta_{1-j}^{YYI}(t)\|\right)^2,
\end{aligned}$$

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

$$= \mathbb{E} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \|\pi_{XYj}^{(h)}\| \|\pi_{XYk}^{(h)}\| \|\Delta_{1-j}^{YYI}(t)\| \|\Delta_{1-k}^{YYI}(t)\|$$

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \|\pi_{XYj}^{(h)}\| \|\pi_{XYk}^{(h)}\| \mathbb{E}\{\|\Delta_{1-j}^{YYI}(t)\| \|\Delta_{1-k}^{YYI}(t)\|\},$$

by the Fubini–Tonelli theorem,

$$\leq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \|\pi_{XYj}^{(h)}\| \|\pi_{XYk}^{(h)}\| \left(\mathbb{E}\|\Delta_{1-j}^{YYI}(t)\|^2\right)^{\frac{1}{2}} \left(\mathbb{E}\|\Delta_{1-k}^{YYI}(t)\|^2\right)^{\frac{1}{2}},$$

by the Cauchy–Schwartz inequality,

$$\leq \left(\sum_{j=1}^{\infty} \|\pi_{XYj}^{(h)}\|\right)^2 \sup_{k\leq 1} \mathbb{E}\|\Delta_{1-k}^{YYI}(t)\|^2$$

$$\leq \left(\sum_{j=1}^{\infty} \|\pi_{XYj}^{(h)}\|\right)^2 \sup_{s\leq t} \mathbb{E}\|Y(s) - P(Y(s)|H)\|^2,$$

because projecting onto H produces a larger mean squre error than projecting onto $I(t)$,

$$= \left(\sum_{j=1}^{\infty} \|\pi_{XYj}^{(h)}\|\right)^2 \sup_{s\leq t} \mathbb{E}\|\widehat{Y}(s)\|^2$$

$$\leq \left(\sum_{j=1}^{\infty} \|\pi_{XYj}^{(h)}\|\right)^2 \max\{\max_{\tau<s\leq t} \mathbb{E}\|P_{\mathcal{U}}\widehat{Y}(s)\|^2, \sup_{s\leq\tau} E\|P_{\mathcal{U}}\widehat{Y}(s)\|^2\},$$

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

The converse is proven similarly to Theorem 3.1 of DR. Let,

$$\Delta_h^{XYI}(t) = \sum_{j=1}^{\infty} \left[\pi_{XXj}^{(h)} - \phi_{XXj}^{(h)} \quad \pi_{XYj}^{(h)} \quad \pi_{XZj}^{(h)} - \phi_{XZj}^{(h)}\right] W(t-j).$$

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, $\sum_{j=1}^{\infty}[\pi_{XXj}^{(h)} - \phi_{XXj}^{(h)} \quad \pi_{XYj}^{(h)} \quad \pi_{XZj}^{(h)} - \phi_{XZj}^{(h)}]\mathbb{E}(W(t-j)a'(t)) \to 0$. Since $\mathbb{E}(W(t-j)a'(t)) = \Omega(t) > 0$ for $j = 0$ and is zero otherwise, this implies that $[\pi_{XX1}^{(h)} - \phi_{XX1}^{(h)} \quad \pi_{XY1}^{(h)} \quad \pi_{XZ1}^{(h)} - \phi_{XZ1}^{(h)}] \to 0$ 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 $[\pi_{XX2}^{(h)} - \phi_{XX2}^{(h)} \quad \pi_{XY2}^{(h)} \quad \pi_{XZ2}^{(h)} - \phi_{XZ2}^{(h)}] \to 0$. Continuing on with this process proves that, $\lim_{h\to\infty} \pi_{XYj}^{(h)} = 0$ for all $j \geq 1$. $\square$

Proof of Theorem 5.3. Follows from Theorem 3.2 of DR and Remark 5.1. □

Proof of Theorem 5.4. Similar to the proofs of Theorems 5.1 – 5.3 and is omitted. □

The next proof utilize the following notation. Set $\pi^{(h)}(z) = \sum_{j=1}^{\infty} \pi_j^{(h)} z^j$, with $\pi^{(1)}(z) = \pi(z)$ and partition it conformably with the projection coefficient matrices so that,

$$\pi^{(h)}(z) = \begin{bmatrix} \pi_{XX}^{(h)}(z) & \pi_{XY}^{(h)}(z) & \pi_{XZ}^{(h)}(z) \\ \pi_{YX}^{(h)}(z) & \pi_{YY}^{(h)}(z) & \pi_{YZ}^{(h)}(z) \\ \pi_{ZX}^{(h)}(z) & \pi_{ZY}^{(h)}(z) & \pi_{ZZ}^{(h)}(z) \end{bmatrix}.$$

We will also need the impulse response operator $\psi(w) = \sum_{h=1}^{\infty} \psi_h z^h$ defined by the equation $I_n + \psi(w) = (I_n - \pi(w))^{-1}$. Dufour & Renault (1995) show that $\psi_h = \pi_1^{(h)}$ for all $h \geq 1$. Assumption 3 ensures that all of these power series have a non–zero radius of convergence.

We will also need the following result, which generalizes Lemma A.4 of DR.

Lemma 10.1. Suppose $b(z)$ is a power series with a non–zero radius of convergence with $b(0) \neq 0$ and $c^h(z)$ is a polynomial of degree p, indexed by h. If $a^h(z) = \sum_{i=0}^{\infty} a_i^h z^i = b(z)c^h(z)$, then $\{\lim_{h\to\infty} a_i^h\}_{i=0}^{\infty} = \{0\}$ if and only if $\{\lim_{h\to\infty} a_i^h\}_{i=0}^{p} = \{0\}$. Clearly, when $c^h(z)$ is independent of h, $\{a_i^h\}_{i=0}^{\infty} = \{0\}$ if and only if $\{a_i^h\}_{i=0}^{p} = \{0\}$.

Proof of Lemma 10.1. Suppose $\{\lim_{h\to\infty} a_i^h\}_{i=0}^{p} = \{0\}$. From the multiplication rule for power series,

$$a_i^h = \sum_{j=0}^{\min\{i,p\}} b_{i-j} c_j^h. \tag{10.1}$$

Now for $h = 0$, $\lim_{h\to\infty} a_0^h = b_0 \lim_{h\to\infty} c_0^h$ and since $b_0 \neq 0$, this implies that $\lim_{h\to\infty} c_0^h = 0$. Suppose now that $\lim_{h\to\infty} c_i^h = 0$ for $i = 0, 1, \ldots, k < p$. Then by (10.1), $\lim_{h\to\infty} a_{k+1}^h = \sum_{j=0}^{k+1} b_{k+1-j} \lim_{h\to\infty} c_j^h = b_0 \lim_{h\to\infty} c_{k+1}^h$. Since $b_0 \neq 0$, this again implies that $\lim_{h\to\infty} c_{k+1}^h = 0$. Therefore, $\lim_{h\to\infty} c_i^h = 0$ for $i = 0, \ldots, p$. By (10.1) this implies that $\lim_{h\to\infty} a_i^h = 0$ for all $i \geq 0$. □

Proof of Theorem 6.1. We prove the theorem from the Cartesian perspective. The subspace version then follows from Remark 5.1.

(i) The proof is in two steps.

Step 1: $\pi_{XYj} = 0$ for all $j \geq 1$ 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 10.1 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: $\pi^{(h)}_{XYj} = 0$ for all $j \geq 1$ if $\pi^{(h)}_{XYj} = 0$ for all $1 \leq j \leq p + (n-1)q$.

We prove our claim by showing that $\pi^{(h)}_{XY}(z)$ 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 $\pi^{(i)}_{XY}(z)$ 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 10.1.

We will need the following equation, which can be found in DR,

$$\pi^{(h+1)}_j = \pi^{(h)}_{j+1} + \pi^{(h)}_1 \pi_j, \qquad j, h \geq 1. \tag{10.2}$$

Therefore,

$$\pi^{(h+1)}(z) = z^{-1}\pi^{(h)}(z) + \pi^{(h)}_1(\pi(z) - I_n), \qquad h \geq 1. \tag{10.3}$$

It follows that for $h \geq 2$,

$$\pi^{(h)}_{XY}(z) = z^{-1}\pi^{(h-1)}_{XY}(z) + \pi^{(h-1)}_{XX1}\pi_{XY}(z) + \pi^{(h-1)}_{XY1}(\pi_{YY}(z) - I_{n_Y}) + \pi^{(h-1)}_{XZ1}\pi_{ZY}(z). \tag{10.4}$$

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 $\pi^{(h-1)}_{XY}(z)$ 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 $z = 0$, $z^{-1}\pi^{(h-1)}_{XY}(z)$ is representable by a ratio with a numerator of degree $p + (n-1)q - 1$ and the denominator $\det(\theta(z))$.

(ii) The proof is in two steps.

Step 1. $Y \nrightarrow_\infty X\ [I]$ if $\lim_{h\to\infty} \pi^{(h)}_{XYj} = 0$ for all $j \geq 1$.

The main idea of the proof of this step is that $\Delta^{XYI}_h(t)$ is expressible as the product of a non–random matrix of finite dimensions with a vector in L^2, where the matrix shrinks to zero as $h \to \infty$ and the vector is independent of h. The assumption of rational $\pi(z)$ is crucial here because it ensures the finiteness of the dimensions and the independence of the vector from h.

First, write the projection coefficient matrices in state space form as $\pi_j^{(h)} = C'A^hF_j$ where,

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

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

and F_j is the j–th block of n columns of F (Lütkepohl, 2006, section 11.5). For $t > \varpi + q$,

$$
P(W(t+h)|H + W(-\infty, t]) = \sum_{k=0}^{h-1} \pi_1^{(k)} \mu(t+h-k) + C'A^h\widetilde{W}(t),
$$

where $\widetilde{W}(t) = [\,W'(t)\; W'(t-1)\; \cdots\; W'(t-p+1)\;\; a'(t)\; a'(t-1)\; \cdots\; a'(t-q+1)\,]'$. It is this aspect of VARMA, the fact that the information of the infinite past can be condensed into a finite vector $\widetilde{W}(t)$, that allows us to prove the result.

Now for $t > \varpi + q$,

$$\Delta_h^{XYI}(t) = E_X' C' A^h \{\widetilde{W}(t) - P(\widetilde{W}(t)|I(t))\},$$

where E_X is an $n \times n_X$ matrix that selects the columns associated with X from any matrix of n columns. Because, $\widetilde{W}(t) = \sum_{j=1}^{\infty} F_j W(t+1-j)$, $\widetilde{W}(t) - P(\widetilde{W}(t)|I(t)) = \sum_{j=1}^{\infty} F_j E_Y \Delta_{1-j}^{YYI}(t)$, where E_Y is defined similarly to E_X. This implies that $\widetilde{W}(t) - P(\widetilde{W}(t)|I(t))$ lies in the subspace $\mathcal{W} = \sum_{j=1}^{\infty} \mathrm{im}(F_j E_Y)$. To prove that $\Delta_h^{XYI}(t)$ converges to zero, it suffices to prove that $E_X' C' A^h P_{\mathcal{W}}$ converges to zero. Now because $\mathcal{W}$ is a subspace of Euclidean space, there exists a finite number m such that $\mathcal{W} = \mathrm{im}\left(\left[\, F_1 E_Y \ \cdots \ F_m E_Y \,\right]\right)$. It then follows that $E_X' C' A^h P_{\mathcal{W}}$ converges to zero if and only if $\lim_{h\to\infty} E_X' C' A^h F_j E_Y = 0$ for $j = 1, \dots, m$. Since $\pi_{XYj}^{(h)} = E_X' C' A^h F_j E_Y$, the result follows.

On the other hand, for $\varpi < t \le \varpi + q$ and $h > l = \max\{p, q\}$,

$$\begin{aligned}
\Delta_h^{XYI}(t) &= P(X(t+h)|H + W(-\infty, t]) - P(X(t+h)|I(t)) \\
&= P(P(X(t+h)|H + W(-\infty, t+l])|H + W(-\infty, t]) \\
&\quad - P(P(X(t+h)|H + W(-\infty, t+l])|I(t)) \\
&= E_X' C' A^{h-l} \{P(\widetilde{W}(t+l)|H + W(-\infty, t]) - P(\widetilde{W}(t+l)|I(t))\},
\end{aligned}$$

and the vector in curly brackets can be written as,

$$\Delta_l^{\widetilde{W}YI}(t) = \left[\begin{array}{c} \Delta_l^{WYI}(t) \\ \Delta_{l-1}^{WYI}(t) \\ \vdots \\ \Delta_{l-p+1}^{WYI}(t) \\ \hline 0 \\ \vdots \\ 0 \end{array}\right] = \left[\begin{array}{ccc} \pi_1^{(l)} & \pi_2^{(l)} & \cdots \\ \pi_1^{(l-1)} & \pi_2^{(l-1)} & \cdots \\ \vdots & \vdots & \\ \pi_1^{(l-p+1)} & \pi_2^{(l-p+1)} & \cdots \\ \hline 0 & 0 & \cdots \\ \vdots & \vdots & \\ 0 & 0 & \cdots \end{array}\right] \left[\begin{array}{c} E_Y \Delta_0^{YYI}(t) \\ E_Y \Delta_{-1}^{YYI}(t) \\ \vdots \\ \vdots \\ \\ \end{array}\right].$$

Clearly the matrix in the above expression is $A^l F$. It follows, again, that $\Delta_l^{\widetilde{W}YI}(t)$ lies in the image of $\left[\, A^l F_1 E_Y \ A^l F_2 E_Y \ \cdots \ A^l F_m E_Y \,\right]$ for some finite m. It then follows, just as before, that $\lim_{h\to\infty} \Delta_h^{XYI}(t) = 0$ if $\lim_{h\to 0} \pi_{XYj}^{(h)} = \lim_{h\to 0} E_X' C' A^{h-l} A^l F_j E_Y = 0$ for $j = 1, \dots, m$.

Step 2. $\lim_{h\to\infty} \pi_{XYj}^{(h)} = 0$ for all $j \ge 1$ if $\lim_{h\to\infty} \pi_{XYj}^{(h)} = 0$ for all $1 \le j \le p + (n-1)q$.

From equation (10.2), 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 the first $p + (n-1)q - 1$ equations hold. Recall that $\pi_{XY}(z)$ and $\pi_{ZY}(z)$ have zeros at $z = 0$. Therefore, from (i), $z^{-1}\pi_{XY}(z)$ and $z^{-1}\pi_{ZY}(z)$ are also representable as ratios of $[(p + (n-1)q) - 1]$–order matrix polynomials and an nq–order polynomial. The result then follows from Lemma 10.1 applied to $\pi_{XX1}^{(h)}\pi_{XY}(z) + \pi_{XZ1}^{(h)}\pi_{ZY}(z)$.

(iii) The proof is in two steps.

Step 1: $\pi_{XYj}^{(h)} = 0$ for all $j, h \geq 1$ if $\pi_{XYj}^{(h)} = 0$ for all $j \geq 1$ and $1 \leq h \leq n_Z(p+(n-1)q)+1$.

From Lemma 3.2 of DR and Lemma A.3, $Y \nrightarrow_{[1,\infty)} X\ [I]$ if $\pi_{XY}(z) = 0$ and $\pi_{XZ}(z)(I_{n_Z} - \pi_{ZZ}(z))^{-1}\pi_{ZYj} = 0$ for all $j \geq 1$.[22] 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 10.1 that $\pi_{XZ}(z)(I_{n_Z} - \pi_{ZZ}(z))^{-1}\pi_{ZYj}$ 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 $1 \leq h \leq n_Z(p+(n-1)q)+1$ and all $j \geq 1$.

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

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

[22] A quicker proof of this than DR's proof is obtained by noting from equation (10.4) that $\pi^{(h)}(z) = 0$ for all $h \geq 1$ if and only if $\pi_{XY}(z) = 0$ and $\psi_{XZ}(w)\pi_{ZY}(z) = 0$. Assuming this and 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 converse follows from the same two blocks. On combining them we obtain $\psi_{XY}(w)(I_{n_Y} - \pi_{YY}(w) - \pi_{YZ}(w)(I_{n_Z} - \pi_{ZZ}(w))^{-1}\pi_{ZY}(w)) = 0$, which implies that $\psi_{XY}(w) = 0$. The XZ block then implies that $\psi_{XZ}(w)\pi_{ZY}(z) = 0$.

Proof of Theorem 6.2. Similar to the proof of Theorem 6.1 and is omitted. □

Proof of Theorem 6.3. Let $\breve{W} = U'\widehat{W}$ satisfy a stable vector difference equation with respect to a. Then employing the same type of construction as that used in the proof of Theorem 6.1 (ii), we can show that for all $t > \varpi$, $P(\breve{W}(t+h)|H + W(-\infty,t]) = C'A^{h-l}\widetilde{W}(t)$, where $\widetilde{W}(t)$ is a vector of L^2 variables and A is a stable matrix. But $P(\breve{W}(t+h)|H + W(-\infty,t]) = P(U'W(t+h)|H + W(-\infty,t]) - P(U'W(t+h)|H) = U'\Delta_h^{WWH}(t)$.

Now suppose that $\breve{W} = U'\widehat{W}$ is unstable and regular. We will show that W cannot be L^2–mean–reverting along $\mathcal{U}$. We assume, without loss of generality, that (ϕ,θ) is left coprime, otherwise simply factor out a greatest common left divisor. Let $Q = [\,U\ U_\perp\,]$ and let $S(z)$ be a unimodular matrix that transforms $\phi(z)Q^{-1}$ to Hermite form (Hannan & Deistler, 1988, Lemma 2.2.2). Then we have,

$$S(z)\phi(z)Q^{-1} = \begin{bmatrix} \breve{\phi}(z) & 0 \\ * & * \end{bmatrix}, \qquad S(z)\theta(z) = \begin{bmatrix} \breve{\theta}(z) \\ * \end{bmatrix},$$

where $\breve{\phi}(z)$ is square and, along with $\breve{\theta}(z)$, has u rows. If $(\breve{\phi},\breve{\theta})$ is not left coprime, then there exists $0 \neq \xi \in \mathbb{R}^u$ and $\lambda \in \mathbb{C}$ such that $\xi'\breve{\phi}(\lambda) = 0$ and $\xi'\breve{\theta}(\lambda) = 0$ (Hannan & Deistler, 1988, Lemma 2.2.1). This then implies that $(\xi',0')S(\lambda)\phi(\lambda) = 0$ and $(\xi',0')S(\lambda)\theta(\lambda) = 0$, which contradicts our assumption that (ϕ,θ) is left coprime. Therefore, $(\breve{\phi},\breve{\theta})$ is left coprime. Likewise, if $\det(\breve{\phi}(0)) = 0$, then there exists $0 \neq \xi \in \mathbb{R}^u$ such that $\xi'S(0)\phi(0) = 0$ but this is impossible since $\phi(0) = I_n$ and $\det(S(0)) \neq 0$. Thus $\det(\breve{\phi}(0)) \neq 0$ and $\breve{W}$ is determined by $\breve{\phi}(L)\breve{W}(t) = \breve{\theta}(L)a(t)$ for $t > \varpi + q + \deg(S)$, where L is the lag operator.

Because $\breve{W}$ is unstable, there exists $0 \neq \xi_1 \in \mathbb{R}^u$ such that $\xi_1'\breve{\phi}(\lambda) = 0$ for some $\lambda \in \mathbb{C}$ with $0 \neq |\lambda| \leq 1$. Define the sequence $\xi_i' = \lambda^{-1}\xi_{i-1}' - \xi_1'\breve{\phi}_{i-1}$ for $i = 2,\ldots,\breve{p} = \deg(\breve{\phi})$. Combining these equations with the definition of ξ_1, we obtain $\xi_{\breve{p}}' = \lambda\xi_1'\breve{\phi}_{\breve{p}}$. Finally, defining $\breve{w}(t) = \sum_{i=1}^{\breve{p}} \xi_i'\breve{W}(t+1-i)$ we have $\breve{w}(t) = \lambda^{-1}\breve{w}(t-1) + \xi_1'\breve{\phi}(L)W(t) = \lambda^{-1}\breve{w}(t-1) + \xi_1'\breve{\theta}(L)a(t)$ for $t > \varpi + q + \deg(S)$.[23] For such a t and $h \geq \breve{q} = \deg(\breve{\theta})$,

$$P(\breve{w}(t+h)|H + W(-\infty,t]) = \lambda^{-h}\left(\breve{w}(t) + \sum_{i=1}^{\breve{q}} \lambda^i\xi_1'P(\breve{\theta}(L)a(t+i)|H + W(-\infty,t])\right).$$

If this quantity goes to zero in L^2 as $h \to \infty$, then its dot product with $a(t)$, $\lambda^{-h}\xi_1'\breve{\theta}(\lambda)\Omega(t)$ also goes to zero. But this later quantity does not converge to zero unless $\xi_1'\breve{\theta}(\lambda)\Omega(t) = 0$,

[23] This choice of linear combination isolates an unstable mode of $\breve{W}$ associated with λ. $(\xi_1',\ldots,\xi_p')'$ is an eigenvector of the companion matrix of $\breve{\phi}$ associated with λ^{-1}.

which is impossible since $\Omega(t)$ is non–singular and $(\breve{\phi}, \breve{\theta})$ is left coprime. It follows that the forecasts of $\breve{w}$ do not converge to zero in L^2 and therefore there is a component of $\breve{W}$ that does not have a long run forecast of zero. That is, $P(\breve{W}(t+h)|H+W(-\infty,t]) = U'\Delta_h^{WWH}(t)$ does not converge to zero. $\square$

Proof of Theorem 7.1. If $\mathcal{U}$ is unreachable, then $P_{\mathcal{U}}CA^jB = 0$ for all $j \geq 0$. Since for all $Y \in \mathscr{T}$,

$$\Delta_h^{XYH}(t) = \sum_{j=0}^{t+h-1} CA^jBP(Y(t+h-j-1)|Y(\omega,t]), \quad t > 0,$$

$\mathcal{U} \subseteq \mathcal{U}_{[1,\infty)}^{XYH}$. On the other hand if $\mathcal{U} \subseteq \mathcal{U}_{[1,\infty)}^{XYH}$, choose Y to be a white noise process with covariance matrix equal to the identity. Then for all $h \geq 1$, $0 = \Delta_h^{P_{\mathcal{U}}XYH}(t) = \sum_{j=0}^{t} P_{\mathcal{U}}CA^{j+h-1}BY(t-j)$. This implies that, $\mathbb{E}\left(\Delta_h^{P_{\mathcal{U}}XYH}(t)Y'(t)\right) = P_{\mathcal{U}}CA^{h-1}B = 0$ for all $h \geq 1$ and so $\mathcal{U}$ is unreachable. $\square$

Proof of Theorem 8.1. If Y is white noise, then (8.3) implies that $\widetilde{X}_2(t) = 0$ for $t > 0$. It follows that $\widetilde{X}_2$ cannot help forecast X and X cannot improve upon a forecast of zero for $\widetilde{X}_2$. That is, $\operatorname{im}(Q_2) \subseteq \mathcal{V}_{[1,\infty)}^{XXH}$ and $\operatorname{im}(Q_2) \subseteq \mathcal{U}_{[1,\infty)}^{XXH}$ respectively.

More generally, $\Delta_h^{Q_2'XXH}(t) = P(\widetilde{X}_2(t+h)|X(\omega,t]) = P(P(\widetilde{X}_2(t+h)|I(t))|X(\omega,t])$ and so it suffices to show that $P(\widetilde{X}_2(t+h)|I(t))$ converge to zero in L^2. By similar arguments to those used in the proof of Theorem 5.2, we can show that $\mathbb{E}\|P(\widetilde{X}_2(t+h)|I(t))\|^2 \leq \sum_j \sum_k \|F^{j+k-2}\|\|G\|^2(\mathbb{E}\|P(Y(t+j+h)|I(t))\|^2)^{\frac{1}{2}}(\mathbb{E}\|P(Y(t+k+h)|I(t))\|^2)^{\frac{1}{2}}$. Since Y is L^2–mean–reverting, each summand converges to zero as $h \to \infty$. Moreover, each summand is bounded above by $\|F^{j+k-2}\|\|G\|^2 \sup_{s\geq t} \mathbb{E}\|P(Y(s)|I(t))\|^2$, which is summable (Horn & Johnson, 1985, see problem 2 of section 5.7). Therefore $\mathbb{E}\|P(\widetilde{X}_2(t+h)|I(t))\|^2$ converges to zero by the dominated convergence theorem. $\square$

References

Al-Sadoon, M. M. (2009a). Causality along subspaces: Inference. Mimeo, University of Camrbridge.

Al-Sadoon, M. M. (2009b). Causality along subspaces: Theory. Cambridge working papers in economics, Faculty of Economics, University of Cambridge.

Al-Sadoon, M. M. (2010). *Causality Along Subspaces*. PhD thesis, University of Cambridge.

Al-Sadoon, M. M. (2011). The stochastic wald test. Mimeo, University of Camrbridge.

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.

Andrews, D. W. K. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability*, *21*(4), pp. 930–934.

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

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.

Davidson, J. (1994). *Stochastic Limit Theory*. Advanced Texts in Econometrics. Oxford, UK: Oxford University 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.

Dufour, J.-M. & Taamouti, A. (2010). Short and long run causality measures: Theory and inference. *Journal of Econometrics*, *154*(1), 42–58.

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.

Engle, R. F. & Kozicki, S. (1993). Testing for common features. *Journal of Business & Economic Statistics*, *11*(4), pp. 369–380.

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.

Granger, C. W. J. & Lin, J.-L. (1995). Causality in the long run. *Econometric Theory*, *11*(3), 530–536.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. *Journal of Political Economy*, *86*(6), pp. 971–987.

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

Hannan, E. J. & Deistler, M. (1988). *The Statistical Theory of Linear Systems*. Wiley Series in Probability and Mathematical Statistics. New York, NY, USA.: John Wiley & Sons, Inc.

Hannan, E. J. & Poskitt, D. S. (1988). Unit canonical correlations between future and past. *The Annals of Statistics*, *16*(2), pp. 784–790.

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

Hansen, L. P. & Sargent, T. J. (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control*, *2*(1), 7–46.

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.

Hsiao, C. (1982). Autoregressive modeling and causal ordering of economic variables. *Journal of Economic Dynamics and Control*, *4*(1), 243–259.

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

Juselius, M. (2008). Cointegration implications of linear rational expectation models. Research Discussion Papers 6/2008, Bank of Finland.

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

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.

Onatski, A. (2006). Winding number criterion for existence and uniqueness of equilibrium in linear rational expectations models. *Journal of Economic Dynamics and Control*, *30*(2), 323–345.

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.

Palma, W. (2007). *Long–Memory Time Series: Theory and Methods.* Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley and Sons, Inc.

Pearl, J. (2009). *Causality* (2 ed.). Cambridge, UK: Cambridge University Press.

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

Poskitt, D. S. (2000). Strongly consistent determination of cointegrating rank via canonical correlations. *Journal of Business & Economic Statistics*, *18*(1), pp. 77–90.

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.

Sims, C. A. (2002). Solving linear rational expectations models. *Computational Economics*, *20*(1-2), 1–20.

Tjøstheim, D. (1981). Granger-causality in multiple time series. *Journal of Econometrics*, *17*(2), 157 – 176.

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

White, H., Al-Sadoon, M., & Chalak, K. (2012). Rational expectations and causality: A settable systems view. Technical report.

White, H., Chalak, K., & Lu, X. (2010). Linking granger causality and the pearl causal model with settable systems. Boston College Working Papers in Economics 744, Boston College Department of Economics.

White, H. & Lu, X. (2010). Granger causality and dynamic structural systems. *Journal of Financial Econometrics*, *8*(2), 193–243.

White, H. & Pettenuzzo, D. (2011). Granger causality, exogeneity, cointegration, and economic policy analysis. Technical report.

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

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ, USA.: Princetone University Press.

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