Testing Subspace Granger Causality

Majid M. Al-Sadoon*

August 15, 2017

Abstract

The methodology of multivariate Granger non–causality testing at various horizons is extended to allow for inference on its directionality. Empirical manifestations of these subspaces are presented and useful interpretations for them are provided. Simple vector autoregressive models are used to estimate these subspaces and to find their dimensions. The methodology is illustrated by an application to empirical monetary policy, where a conditional form of Okun's law is demonstrated as well as a statistical monetary policy reaction function to oil price changes.

Keywords: Granger causality, VAR model, rank testing, Okun's law, policy trade–offs.

*Universitat Pompeu Fabra & Barcelona GSE. 25-27 Calle Ramon Trias Fargas. 08005 Barcelona, Spain. Tel.: +34 93 542 2682. E-mail address: majid.alsadoon@gmail.com. The software package accompaniment to this paper can be found on the author's website: https://www.upf.edu/web/majid-al-sadoon/.

1 Introduction

The concepts of Granger causality (GC) and Granger non–causality (GNC) developed by Wiener (1956) and Granger (1969) are fundamental concepts in time series analysis (see e.g. the surveys of Geweke (1984) or Hamilton (1994)). Many extensions have been proposed to the basic concept throughout the years. To name some of these extensions, we have multivariate analysis (Tjøstheim, 1981), enlarged information sets (Hsiao, 1982), variable horizons (Dufour & Renault, 1998; Dufour et al., 2006), graphical modelling techniques (Eichler, 2007), measurement under linearity (Dufour & Taamouti, 2010), GC priority (Jarociński & Maćkowiak, 2017), measurement under non–linearity (Song & Taamouti, 2017), and second order GC (Dufour & Zhang, 2015).

Recently, Al-Sadoon (2014) has shown that some of the multivariate notions of GC proposed above may not give a full characterization of the structure of dynamic dependence of the system and proposed the extensions to subspace Granger causality (SGC) and subspace Granger non–causality (SGNC). The basic idea is that if a vector process Y helps predict the vector process X at horizon h, the predictive effect may be limited to a subspace in two different ways: (i) Y may predict comovements of X in some directions but not in all directions (i.e. GC is limited to a subspace of X–space) or (ii) comovements of Y in certain (but not all) directions may have a predictive effect on X (i.e. GC is limited to a subspace of Y–space). Al-Sadoon (2014) shows that SGNC in a VAR process is equivalent to rank restrictions on the VAR coefficients rather than the zero block restrictions typically studied in the literature. This accords with T. W. Anderson's seminal contribution that the proper extension of zero univariate restrictions to the multivariate setting is rank restrictions rather than zero block restrictions (Anderson, 1951).

Whereas Al-Sadoon (2014) provided the necessary and sufficient conditions for SGNC in population, the objective of this paper is the statistical testing of SGNC and estimation of subspaces of GNC. The paper employs the method of (p, h) autoregressions (also known as direct VAR forecasting models in the forecasting literature) to estimate the relevant coefficient matrices, just as in Dufour et al. (2006). As is well known, the residuals in such equations are moving averages and therefore hypothesis testing requires the use of HAC estimators. We follow Dufour et al. (2006) in using the Bartlett–Newey–West estimator (Newey & West, 1987). The rank tests are carried out using the QR test statistic of Al-Sadoon (2017b) for

its computational expediency in bootstrapping and the subspace estimation procedure follows Robin & Smith (2000) in utilizing the sequential procedure which tests rank $0, 1, \ldots$ until acceptance. The methodology is also extended to the $I(d)$ case by employing results of Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996) of augmenting the regression equation by redundant lags to achieve standard asymptotics.

As this paper is targeted towards practitioners, the paper takes the following steps in order to accommodate their needs. First, it devotes substantial space to the interpretation and uses of SGC, focusing in particular on cases where endogeneity may be present, as this is likely to be the case in most empirical applications. Second, the Matlab code for the test (`SGNC.m`) and for the data–driven evaluation of its small sample performance (`SSP_SGNC.m`) has been made available and as user–friendly as possible, allowing the practitioner to adjust a wide range of parameters of the test (e.g. the dataset, lag length, horizons, trends, seasonality, etc.).

The new methodology is illustrated by an application to US macroeconomic data. The dataset consists of monthly observations of the monetary policy variable constructed by Romer & Romer (2004), the producer price index for finished goods, the industrial production index, the civilian unemployment rate, and the West Texas Intermediate spot price for oil for the period January 1966 – December 1996 and is not seasonally adjusted (see Section 4 for the data codes and sources). We find that monetary policy predicts variations of industrial production and unemployment growth with a trade–off of around 3% higher unemployment for every 1% fall of industrial production over horizons 1–5. This trade–off doubles at horizons 6 but falls gradually after that. This we interpret as a conditional form of Okun's law. We also find a statistical reaction function of monetary policy to oil prices. In particular, observed decreases of the monetary policy indicator of around 0.15–0.20% in response to 1% increase in oil prices have no predictive effects on unemployment growth.

The paper is organized as follows. Section 2 motivates and reviews the idea of SGC. Section 3 discusses estimation and inference. Section 4 is an empirical illustration of the methodology. Section 5 concludes. Appendices A and B consist of further mathematical results. Finally, the software package accompaniment to this paper (`SGC.rar`) contains all of the data and code necessary to generate the tables, figures, and pretests of this paper and is available on the author's website: `https://www.upf.edu/web/majid-al-sadoon/`.

2 Multivariate Granger Causality in VAR Models

In this section we discuss multivariate GC and its extension to subspace GC. This is accomplished primarily through empirical examples rather than mathematical formalism. The reader desiring a more formal and general discussion is referred to Al-Sadoon (2014).

2.1 Theory of Subspace Granger Causality

In this paper we will be concerned with the n–dimensional VAR(p) process,

$$W(t+1) = \mu^{(1)}(t) + \sum_{j=1}^{p} \pi_j^{(1)} W(t+1-j) + a(t+1), \quad t = p, \ldots, T, \tag{1}$$

where $\mu^{(1)}(t)$ is a n–dimensional deterministic trend and $a(t)$ is a martingale difference sequence with respect to the information set generated by W, with $\mathbb{E}(a(t)a'(t)) = \Omega$ positive definite. The first p observations of W are assumed given.

We will be interested in the predictability of components of $W(t+h)$ with respect to current and past components of W and for that we will need the following representation, which we obtain by iterating equation (1) forwards,

$$W(t+h) = \mu^{(h)}(t) + \sum_{j=1}^{p} \pi_j^{(h)} W(t+1-j) + \sum_{j=0}^{h-1} \psi_j a(t+h-j), \quad t = p, \ldots, T-h, \tag{2}$$

where $\mu^{(h)}(t) = \sum_{j=0}^{h-1} \psi_j \mu^{(1)}(t+h-j)$ for $h \geq 1$. It will be convenient to assume that the deterministic trend satisfies $\mu^{(h)}(t) = \gamma(h) D^{(h)}(t)$, where $D^{(h)}(t)$ is an observable k–dimensional deterministic trend and $\gamma(h)$ is an $n \times k$ coefficient matrix. This is certainly the case for polynomials and seasonal dummies. Dufour & Renault (1995) derive the following formulae for the coefficient matrices $\{\pi_j^{(h)}\}$ and impulse responses $\{\psi_j\}$,

$$\pi_j^{(h+1)} = \pi_{j+h}^{(1)} + \sum_{l=1}^{h} \pi_{h-l+1}^{(1)} \pi_j^{(l)} = \pi_{j+1}^{(h)} + \pi_1^{(h)} \pi_j^{(1)}, \qquad j, h \geq 1 \tag{3}$$

$$\psi_h = \pi_1^{(h)}, \qquad h \geq 1. \tag{4}$$

The representation (2) forms the basis of the Dufour et al. (2006) (henceforth, DPR) analysis of GC as well as this paper's analysis. This model was proposed by Shibata (1980) and has since found a great number of applications in the time series literature (e.g. Bhansali (2002), Jorda (2005), and Al-Sadoon (2014)).

Now partition W as $W(t) = (X'(t), Y'(t), Z'(t))'$, $t = 1, \ldots, T$, where the dimensions of the components X, Y, and Z are n_X, n_Y, and n_Z respectively and partition the coefficient matrices conformably with W as

$$\pi_j^{(h)} = \begin{bmatrix} \pi_{XXj}^{(h)} & \pi_{XYj}^{(h)} & \pi_{XZj}^{(h)} \\ \pi_{YXj}^{(h)} & \pi_{YYj}^{(h)} & \pi_{YZj}^{(h)} \\ \pi_{ZXj}^{(h)} & \pi_{ZYj}^{(h)} & \pi_{ZZj}^{(h)} \end{bmatrix}, \qquad j, h \geq 1. \tag{5}$$

Dufour & Renault (1998) define h–step GNC as follows: Y fails to Granger cause X at horizon h if at every time t the prediction of $X(t+h)$ does not depend on current or past Y. We will denote this by $Y \nrightarrow_h X$. Equation (2) suggests $Y \nrightarrow_h X$ depends on the coefficient matrices $\pi_{XY1}^{(h)}, \ldots, \pi_{XYp}^{(h)}$. This is indeed the case as we make clear in the following result.

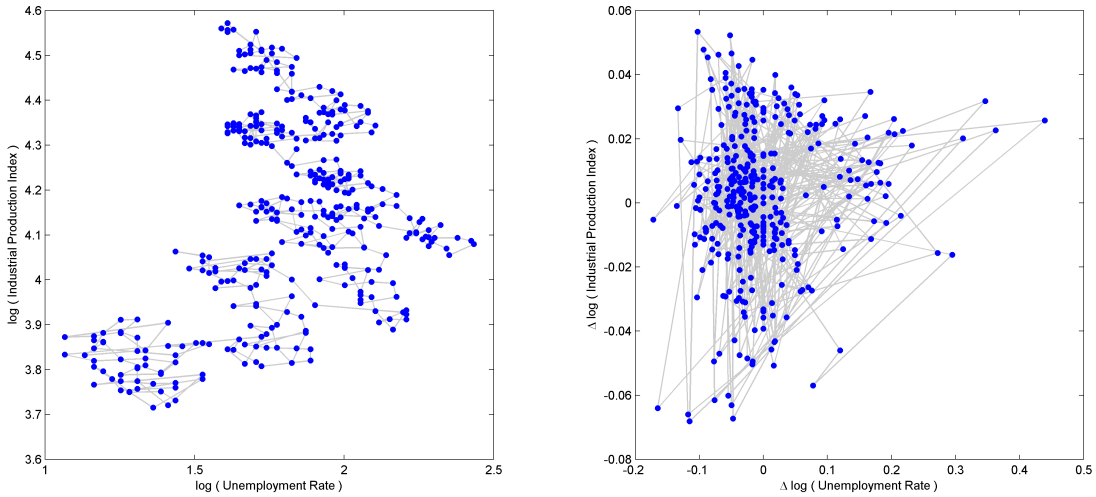
Result 2.1 (Dufour & Renault (1998))**.** $Y \nrightarrow_h X$ if and only if $\pi_{XYj}^{(h)} = 0$ for $j = 1, \ldots, p$.

Note that $Y \nrightarrow_h X$ does not preclude GC at horizon $h + j$ for some $j \geq 1$ because Y may Granger cause Z at horizon h, while Z Granger causes X at horizon j. This chain of GC from Y to X through Z explains why h-step GC is so important for understanding predictive effects as Dufour & Renault (1998) have made clear.

Now Al-Sadoon (2014) has argued that the form of GC proposed by Dufour & Renault (1998) does not capture the full structure of dynamic dependence in multivariate time series. In particular, if we fail to reject GNC, it may still be the case that GC is limited to a particular subspace. This is illustrated empirically in the following two examples.

Example 2.1 (Target Subspace GNC)**.** Consider US monthly data on the industrial production index and civilian unemployment from January 1966 to December 1996. We are interested in looking at the predictability of these variables in terms of the monetary policy variable derived in Romer & Romer (2004). Figure 1 plots the log of the industrial production index against the log of civilian unemployment. The negative comovements of unemployment and output visible in the figure gave rise to Okun's eponymous law (Okun, 1962).

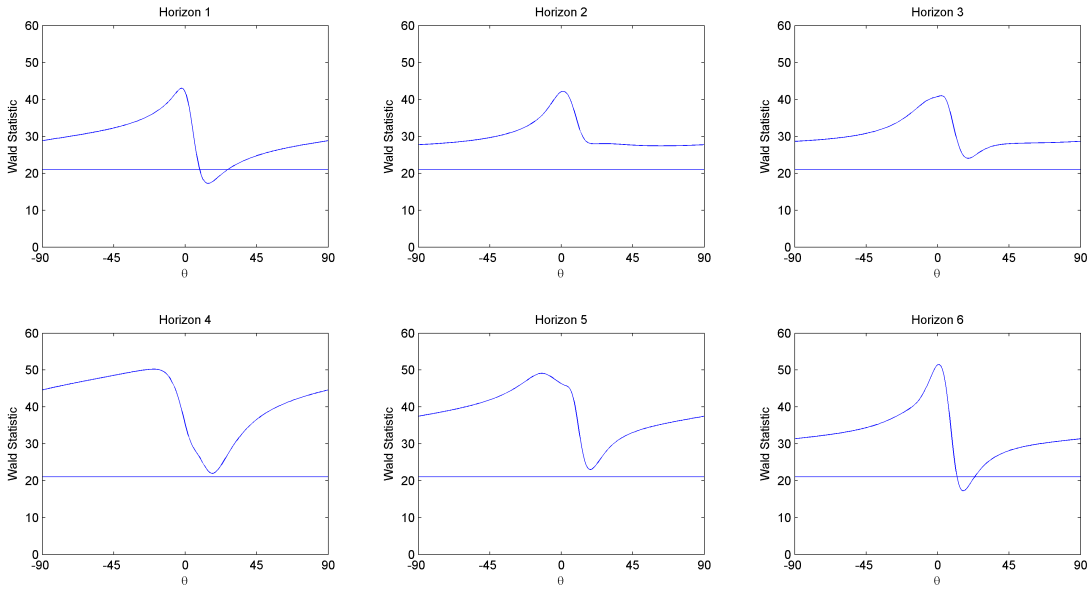
Figure 1: Index of Industrial Production vs. Unemployment



The structure evident in the left panel disappears entirely once we look at the differenced series. Thus, the unconditional form of Okun's law for the differenced series, *"growth rates of industrial production and unemployment exhibit negative comovements,"* is evidently false. We may, however, consider formulating a conditional form of Okun's law, *"monetary policy predicts negative comovements of the growth rates of industrial production and unemployment."* Here, the comovements that we consider are conditional on monetary policy. The conditional form of Okun's law is motivated by the same macroeconomic considerations as the unconditional form of Okun's law: monetary policy, as a driver of aggregate demand, will tend to push industrial production and unemployment in opposite directions.

How can we check whether the conditional form of Okun's law is consistent with the data? One solution would be to form linear combinations of the differenced industrial production index and the unemployment rate series and see which linear combinations are Granger caused by monetary policy. Let r stand for the Romer & Romer (2004) monetary policy measure. Let y and u stand for the differenced logs of the industrial production index and the unemployment rate respectively; these are our *targets* in this exercise. We will also take into account differenced inflation, π, and the differenced log of oil prices, o, as they are important

Figure 2: Naive SGNC Tests for the Predictive Effect of r on $\cos(\theta)y + \sin(\theta)u$



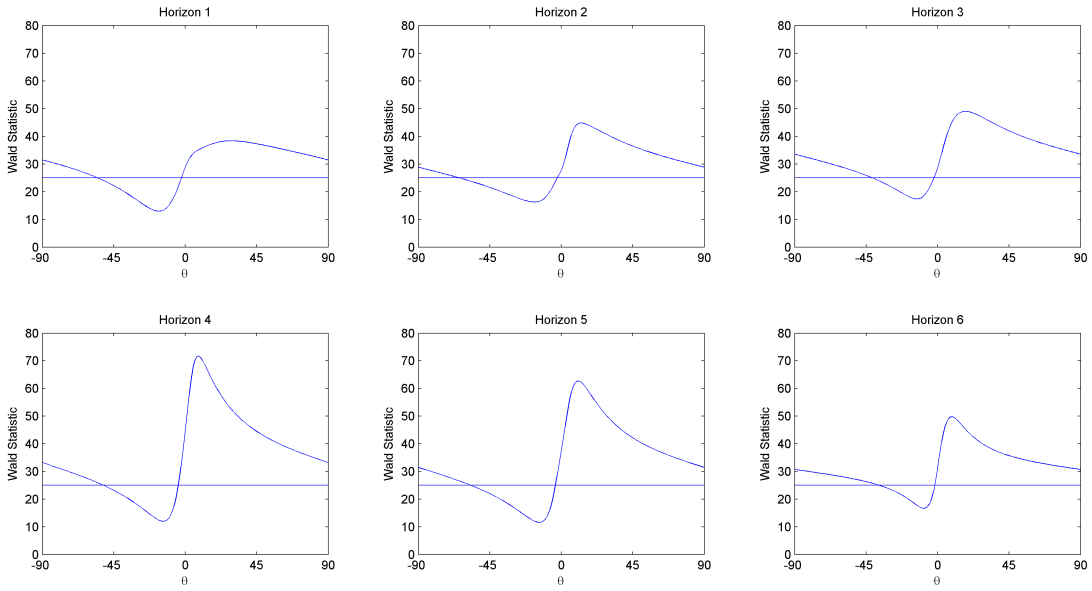
determinants of the dynamics in our sample. We then transform the data as

$$\begin{bmatrix} r \\ \pi \\ y \\ u \\ o \end{bmatrix} \mapsto \begin{bmatrix} r \\ \pi \\ I_1(\theta) \\ I_2(\theta) \\ o \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \cos(\theta) & \sin(\theta) & 0 \\ 0 & 0 & -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r \\ \pi \\ y \\ u \\ o \end{bmatrix} \tag{6}$$

and test the hypothesis $r \nrightarrow_h I_1(\theta)$ for each θ in the range $[-90°, 90°)$ for $h = 1, \ldots, 6$, using the DPR method with 12 lags and including a constant and seasonal dummies (Section 4 provides much more detail on our modelling choices). When GC is detected for $0 \leq \theta < 90°$, this implies that monetary policy predicts variations of (y, u) along positive directions. Instead, we expect that any GC should be confined to θ in the range $[-90°, 0)$. Figure 2 performs just such an exercise for horizons from 1 to 6. The horizontal line represents the critical value at 5% significance of the GC tests. It is evident that although r helps predict each of y and u, it has a stronger predictive effect for some linear combinations than others. For horizons 1 and 6, in particular, there are directions along which the variation of y and u cannot be attributed to r. Thus a conditional form of Okun's law persists in the differenced data. □

Example 2.2 (Predictor Subspace GNC)**.** Suppose we are interested in the joint predictive effect (over the same period and forecast horizons) of monetary policy and oil price growth on the one hand and unemployment growth on the other hand. Individually, both variables

Figure 3: Naive SGNC Tests for the Predictive Effect of $\cos(\theta)r + \sin(\theta)o$ on u



Granger cause unemployment growth (see Section 4). However, one may naturally ask: do all comovements of the monetary policy indicator and oil price growth predict variations in unemployment growth? Just as we answered the question in Example 2.1 by rotating the target space, here we will rotate the predictor space to form the following linear combinations

$$\begin{bmatrix} r \\ \pi \\ y \\ u \\ o \end{bmatrix} \mapsto \begin{bmatrix} I_1(\theta) \\ \pi \\ y \\ u \\ I_2(\theta) \end{bmatrix} = \begin{bmatrix} \cos(\theta) & 0 & 0 & 0 & \sin(\theta) \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -\sin(\theta) & 0 & 0 & 0 & \cos(\theta) \end{bmatrix} \begin{bmatrix} r \\ \pi \\ y \\ u \\ o \end{bmatrix} \tag{7}$$

and measure the predictive effect of $I_1(\theta)$ for u. When GC is found for $\theta \in [0, 90°)$, then u is predicted by positive comovements r and o, the *predictors*. We expect GNC for $\theta \in [-90°, 0)$ as negative comovements of r and o should have predictive effects on u that cancel each other out. This is precisely what we find in Figure 3 using the DPR test as in Example 2.1. Again, the horizontal line represents the critical value at 5% significance. Thus certain negative comovements of monetary policy and oil price growth fail to predict unemployment growth. In other words, what we obtain is a statistical policy reaction function that relates observed variations in oil price growth to observed variations in monetary policy that neutralize the effect on expected unemployment growth. □

Although the peaks in Figures 2 and 2 can be interpreted as directions that maximize

conditional predictability, the purpose of the exercise is to illustrate empirical instances of what Al-Sadoon (2014) has termed subspace GNC, or SGNC for short (SGC is defined similarly relative to GC); namely, there are certain directions along which there is no GC. Readers interested in estimating the directions of the peaks may pursue the partial canonical correlation approach provided in Appendix A. We say that Y along subspace $\mathcal{V} \subseteq \mathbb{R}^{n_Y}$ fails to Granger cause X along subspace $\mathcal{U} \subseteq \mathbb{R}^{n_X}$ at horizon h if the components of Y along $\mathcal{V}$ do not help predict X along $\mathcal{U}$ at horizon h. We denote it by $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}$. Al-Sadoon (2014) shows that $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}$ if and only if $V'Y \nrightarrow_h U'X$ for any matrices V and U whose columns form bases of $\mathcal{V}$ and $\mathcal{U}$ respectively. The requisite restrictions for this sort of GNC are as follows.

Result 2.2 (Al-Sadoon (2014))**.** $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}$ if and only if, $U'\pi^{(h)}_{XY\,j}V = 0$ for all $1 \leq j \leq p$, where V and U span $\mathcal{V}$ and $\mathcal{U}$ respectively.

Since SGNC is GNC of a linear combination of Y for a linear combination of X, all of the insights in Dufour & Renault (1998) continue to hold. In particular, if $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}$, it may still be the case that Y along $\mathcal{V}$ Granger causes X along $\mathcal{U}$ at a horizon $h + j$ with $j \geq 1$, either through Z, $V'_\perp Y$, or $U'_\perp X$, where $V_\perp$ and $U_\perp$ are orthogonal complements to V and U respectively. Thus, there may be chains of GC that run not only through Z but also through the subspaces orthogonal to $\mathcal{V}$ and $\mathcal{U}$.

Now if $\mathcal{U}$ and $\mathcal{V}$ are known then testing for SGNC is easily done by employing a Wald test as in DPR. However, as we saw in the examples above, we will typically not know a priori along which subspaces GNC occurs. We are then lead to the following notions of SGNC.

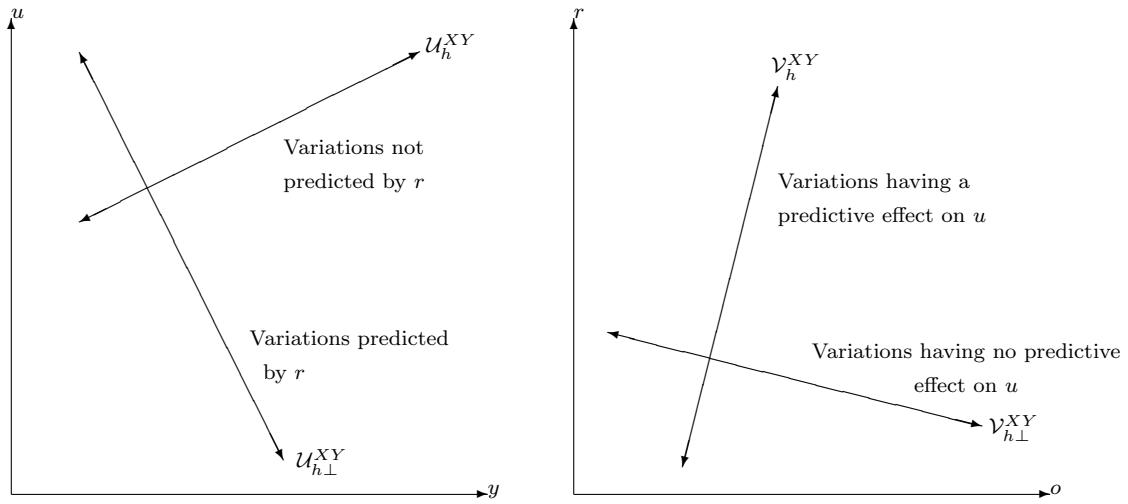
Define *the Subspace of Target GNC at horizon h* to be the maximal subspace $\mathcal{U}$ along which $Y \nrightarrow_h X|_{\mathcal{U}}$ and denote it by $\mathcal{U}^{XY}_h$. Its orthogonal complement $\mathcal{U}^{XY}_{h\perp}$ is defined as *the Subspace of Target GC at horizon h*. In Example 2.1, $\mathcal{U}^{XY}_h$ corresponds to the positively sloped lines along which comovements of the growth rates of industrial production and unemployment were not predictable by monetary policy, while $\mathcal{U}^{XY}_{h\perp}$ corresponds to the negatively sloped line along which monetary policy does have predictive effects (see the left panel in Figure 4).

Result 2.2, implies that $\mathcal{U}^{XY}_h$ is exactly the left null space of

$$C^{\text{target}} = \begin{bmatrix} \pi^{(h)}_{XY1} & \cdots & \pi^{(h)}_{XYp} \end{bmatrix}. \tag{8}$$

On the other hand, the column space of C^{target} is $\mathcal{U}^{XY}_{h\perp}$.

Figure 4: Subspaces of Granger Non–Causality in Examples 2.1 and 2.2.



Define similarly *the Subspace of Predictor GNC at horizon h* to be the maximal subspace $\mathcal{V}$ along which $Y|_{\mathcal{V}} \nrightarrow_h X$ and denote it by $\mathcal{V}_h^{XY}$. Its orthogonal complement $\mathcal{V}_{h\perp}^{XY}$ is defined as *the Subspace of Predictor GC at horizon h*. In Example 2.2, $\mathcal{V}_h^{XY}$ corresponds to the negatively sloped lines along which comovements of monetary policy and the growth rate of oil prices have no predictive effect on the growth rate of industrial production, while $\mathcal{V}_{h\perp}^{XY}$ corresponds to the positively sloped line along which comovements of monetary policy and the growth rate of oil prices do have predictive effects (see the right panel in Figure 4).

Result 2.2, implies that $\mathcal{V}_h^{XY}$ is exactly the right null space of

$$C^{\text{predictor}} = \begin{bmatrix} \pi_{XY1}^{(h)} \\ \vdots \\ \pi_{XYp}^{(h)} \end{bmatrix}. \tag{9}$$

On the other hand, the row space of $C^{\text{predictor}}$ is $\mathcal{V}_{h\perp}^{XY}$.

Note that $\mathcal{U}_{h\perp}^{XY}$ and $\mathcal{V}_{h\perp}^{XY}$ do not necessarily point in the directions indicated by the peaks in Figures 2 and 3. They simply indicate directions orthogonal to $\mathcal{U}_h^{XY}$ and $\mathcal{V}_h^{XY}$ in which GC is present. The objective here is to decompose the target space or the predictor space into non-predictive and predictive directions. Of course, as we have seen in Examples 2.1 and 2.2, the angles between the most predictive and least predictive directions are not necessarily 90°; however, the ranking of directions according to predictive ability is not the objective here. That is the objective of the partial canonical correlations perspective that we detail in Appendix A.

Because they are null spaces, subspaces of GNC are obtainable only under a rank restriction. Thus we are lead naturally to reduced–rank regression. Our procedure for estimating these subspaces exactly mirrors cointegration analysis. To estimate $\mathcal{U}_h^{XY}$ (resp. $\mathcal{V}_h^{XY}$), we will first estimate the rank of C^{target} (resp. $C^{\text{predictor}}$) then, based on this estimated rank, obtain an estimate of the left (resp. right) null space. Identifying restrictions may then be imposed to interpret the estimated relationships, just like the identifying restrictions imposed in cointegration analysis (see e.g. Juselius (2006) or Garratt et al. (2006)). In particular, one may either treat these relationships as directionless trade–offs (e.g. y and u in Example 2.1) or may normalize the relationships so that a certain subset of variables is seen as responding to the other variables (e.g. r responding to o in Example 2.2).

2.2 Interpretation and Utility of Subspace Granger Causality

Before we move on to estimation and inference, it is important to consider the correct interpretation of GC and its extension, SGC, especially as problems of interpretation have dogged GC since its inception (see e.g. Hoover (2001) and Hamilton (1994)).

Recent work by White & Lu (2010), White et al. (2011), and White & Pettenuzzo (2014) has emphasized that, under a form of conditional exogeneity of the predictors, GC implies causality. However, these conditions are not likely to hold in many empirical applications. For example, while many researchers may be comfortable considering the Romer & Romer (2004) series to be conditionally exogenous, its strong similarity to the federal funds rate series raises endogeneity flags for others. How are we to interpret GC and its extension SGC in these cases? The answer is best understood by recalling the difference between a causal effect and a predictive effect. A causal effect tells us what to expect due to a *manipulation* of the predictor, while a predictive effect tell us what to expect due to an *observed* change in the predictor. The difference is illustrated in the following basic example.

Example 2.3. Let $x = \beta' y + \delta' z + u$ be a linear structural equation that determines the causal dependence of x on y, z, and u. Let $E(u|y,z) = \gamma' y + \theta' z$ with $\gamma \neq 0$ so that y is endogenous. To complete the analogy to (1), the regression residual is given by $a = u - E(u|y,z) = x - (\beta + \gamma)y - (\gamma + \theta)z$. Then the causal effect of a change in y from y_0 to y_1 is $\beta'(y_1 - y_0)$, whereas the predictive effect of that change is $(\beta + \gamma)'(y_1 - y_0)$. In the language of Pearl's causal framework, the causal effect is $E(x|do(y = y_1, z = z_0)) - E(x|do(y = y_0, z = z_0))$,

whereas the predictive effect is $E(x|y = y_1, z = z_0) - E(x|y = y_0, z = z_0)$ (Pearl, 2000). The two coincide when $\gamma = 0$ or, more generally, when conditional exogeneity holds (i.e. $E(u|y, z) = E(u|z)$). □

GC is a manifestations of a predictive effect and SGC goes a step further in determining the directions in target space and/or predictor space along which the predictive effect is present. Neither GC nor SGC has any causal meaning without conditional exogeneity. So why bother with GC in empirical practice?

It is certainly the case that causal effects are often the centre of attention in empirical research. However, this is often due to the fact that the objective of the exercise is to prescribe policy. On the other hand, for an outsider who *observes* and does not *manipulate* the instruments of policy, the predictive effect is the more relevant quantity because it tells them what to expect after the *observed* change in policy. In Example 2.3, knowledge of β does not help us obtain the predictive effect, we need $\beta + \gamma$ in order to obtain the predictive effect. For a more concrete example, a market participant may be interested in knowing whether recent variations in monetary policy should lead them to adjust their predictions for GDP. The causal effect is of no use to such a person because (on its own) it does not allow the observer to revise their prediction. Similarly, the manifestations of SGC we have seen in Examples 2.1 and 2.2 are to be interpreted as effects of *observations* (predictive effects) rather than effects of *manipulations* (causal effects). This distinction is important to bear in mind whenever conditional exogeneity is suspect in a regression-based study. It is, incidentally, the very same distinction that exists between generalized impulse responses (Koop et al., 1996) and structural impulse responses (Sims, 1980).

Note finally that SGC provides a natural way of interpreting VAR estimates of the predictive effect of Y on X. The idea is best illustrated in analogy to the 2–dimensional VAR(1) model where the GC of X relative to Y is assessed based on a single coefficient. Based on that coefficient's magnitude we can assess the strength of GC and based on its sign we can assess the direction of GC. For higher dimensional VAR(p) models, we can still assess the strength of GC but its direction cannot be read directly from the signs of the individual elements of $\{\pi^{(h)}_{XYj} : j = 1, \ldots, p\}$ because usually the signs are not uniformly either positive or negative. SGC allows the researcher to retrieve the directional information that was visible in the simpler model.

3 Estimation and Inference

Except for the empirical manifestations of SGNC and the interpretation of SGNC, we have so far only reviewed the basic theory of SGNC put forth by Al-Sadoon (2014). The rest of this paper is dedicated to statistical estimation and inference.

We have already conducted simplistic tests for SGNC in Examples 2.1 and 2.2. The exercise can be seen as a special case of the test for common features proposed by Engle & Kozicki (1993); here the common feature is predictability by Y. However, it is well known that this procedure controls the level but not the size of the test. Moreover, it is well known that tests based on asymptotic critical values lead to over–rejection in small–samples (Dufour & Khalaf, 2003). Thus, the procedure employed in Examples 2.1 and 2.2 can be improved substantially. This section proposes tests of SGNC for both stationary and non–stationary data.

3.1 Estimation and Inference for Stationary VARs

One approach to estimating C^{target} or $C^{\text{predictor}}$ is to regress $W(t)$ on $W(t-1),\ldots,W(t-p)$ in (1) to obtain estimates of $\{\pi^{(1)}_{XY_j}\}$, then iterate using (3) to obtain estimates of $\{\pi^{(h)}_{XY_j}\}$. However, as is well known, these estimates may have a singular asymptotic covariance matrix (see the example in section 3.6.4 of Lütkepohl (2006)). Lutkepohl & Burda (1997) and Dufour & Valery (2016) have suggested regularizing the covariance matrix. On the other hand, Duplinskiy (2014) proposed bootstrapping the non–standardized non–pivotal statistic. Finally, Lütkepohl (2006) p. 108 has suggested imposing the zero restrictions on the coefficients directly.

In this paper we will opt for simplicity and for a procedure that yields an asymptotically pivotal statistic. This is to avoid the difficulty of implementation and/or power loss known to occur in the procedures above. The discussion below is informal. Readers interested in the technical details are referred to Appendix B.

First, we follow DPR in estimating the coefficients in (2) by regressing $W(t+h)$ on $D^{(h)}(t)$, $W(t)$,..., $W(t-p+1)$, obtaining an estimator $\widehat{B}_h$ for

$$B_h = \begin{bmatrix} \gamma(h) & \pi^{(h)}_1 & \cdots & \pi^{(h)}_p \end{bmatrix}, \tag{10}$$

which is $\sqrt{T}$–consistent and asymptotically normal under fairly general regularity conditions. Ω can be estimated as the variance of the OLS residuals in (1), call this estimator $\widehat{\Omega}$. The

impulse responses are then consistently estimated by iterating (3) and (4).

Now the asymptotic variance of $\widehat{B}_h$, call it Ξ_h, requires some care in estimating it because the residuals of the model (2) are autocorrelated for $h > 1$. We follow DPR in utilizing the Bartlett–Newey–West estimator, which requires a bandwidth $m(T)$ to be specified and we will consider two choices for this bandwidth. (i) In the small–b case, $m(T) \to \infty$ as $T \to \infty$ but a slower rate than T. This makes the estimator consistent for the asymptotic variance of $\widehat{B}_h$. This is the common approach found in the literature (Hall, 2005; Cushing & McGravey, 1999; den Haan & Levin, 1997). (ii) In the fixed–b case, we fix $m(T) = bT$ for $b \in (0, 1]$. This makes the estimator inconsistent for the asymptotic variance of $\widehat{B}_h$, although Wald test statistics using this estimator remain asymptotically pivotal in our context. This approach is more recent (Kiefer et al., 2000; Kiefer & Vogelsang, 2002a,b, 2005) and has found great success in controlling for over–rejection in small samples, a serious problem in GC testing. We will compare the performance of small–b and fixed–b statistics and also employ the bootstrap in the next section.

Now depending on what test we are interested in conducting, we may obtain estimates of C^{target} or $C^{\text{predictor}}$ from $\widehat{B}_h$ using the correspondences in (8) or (9) respectively. To simplify the notation in the subsequent analysis, we will simply write C and $\widehat{C}$ and denote their dimensions by m (numbers of rows) and l (number of columns) respectively. We may also extract an estimator for the asymptotic variance of $\widehat{C}$ from the estimator of the asymptotic variance of $\widehat{B}_h$, call it $\widehat{\Theta}$.

We have argued in the previous section that target and predictor SGNC restrictions amount to rank restriction hypotheses of the form

$$H_0(r) : \text{rank}(C) = r, \tag{11}$$

where $r < \min\{m, l\}$. Just as in cointegration analysis, we will test this hypothesis against the alternative

$$H_1(r) : \text{rank}(C) > r. \tag{12}$$

Various options are available for this test. The original analysis of Anderson (1951) can be applied to our regression model but because Ξ_h is not of the Kronecker product form for $h > 1$, Anderson's test statistic may not be asymptotically pivotal (Robin & Smith, 2000). Robin and Smith show that, under $H_0(r)$, it converges in distribution to a weighted sum

of independent $\chi^2(1)$ random variables with weights that depend on nuisance parameters. They show that when the weights are estimated consistently, the test has the correct size asymptotically. However, the presence of nuisance parameters in the asymptotic distribution makes this option less attractive than available alternatives. Cragg & Donald (1996), Cragg & Donald (1997), and Kleibergen & Paap (2006) propose alternative consistent tests based on statistics that are asymptotically pivotal. However, these statistics utilize computationally costly algorithms that may slow down performance as we bootstrap (see the on-line appendix to Al-Sadoon (2017b)), so we opt for the least computationally costly test statistic that is also asymptotically pivotal in our setting, the QR statistic proposed in Al-Sadoon (2017b).

We will sketch the intuitive idea of the QR test, leaving the formal details to Al-Sadoon (2017b). Let $\widehat{C} = \widehat{U}\widehat{S}\widehat{V}'$ be the QR decomposition with pivoting (Golub & Van Loan, 1996, Algorithm 5.4.1). This decomposition is obtained by permuting the columns of $\widehat{C}$ ($\widehat{V}$ is a permutation matrix) so that it can be factorized by the Gram–Schmidt algorithm into a product of an orthogonal matrix (typically denoted by Q, here $\widehat{U}$) and a block upper triangular matrix (typically denoted by R, here $\widehat{S}$). Now partition $\widehat{S}$ as $\left[\begin{smallmatrix} \widehat{S}_{11} & \widehat{S}_{12} \\ 0 & \widehat{S}_{22} \end{smallmatrix}\right]$ so that $\widehat{S}_{11} \in \mathbb{R}^{r\times r}$. The basic idea behind the test is that $\widehat{S}_{22}$ is small when $\widehat{C}$ approaches a rank–r matrix and bounded away from zero when $\widehat{C}$ approaches a matrix of rank higher than r. Setting, $\widehat{N}_r = \widehat{U}\left[\begin{smallmatrix} 0 \\ I_{m-r} \end{smallmatrix}\right]$ and $\widehat{M}_r = \widehat{V}\left[\begin{smallmatrix} -\widehat{S}_{11}^{-1}\widehat{S}_{12} \\ I_{l-r} \end{smallmatrix}\right]$, we have that $\widehat{N}_r'\widehat{C}\widehat{M}_r = \widehat{S}_{22}$ and we may base our inference on the statistic

$$F = T\text{vec}'(\widehat{S}_{22})\{(\widehat{M}_r \otimes \widehat{N}_r)'\widehat{\Theta}(\widehat{M}_r \otimes \widehat{N}_r)\}^{-1}\text{vec}(\widehat{S}_{22}). \tag{13}$$

The plug-in principle of Al-Sadoon (2017b) implies that, under $H_0(r)$, F has the same asymptotic behaviour as the infeasible statistic

$$T\text{vec}'(N_r'\widehat{C}M_r)\{(M_r \otimes N_r)'\widehat{\Theta}(M_r \otimes N_r)\}^{-1}\text{vec}(N_r'\widehat{C}M_r), \tag{14}$$

where N_r and M_r span the left and right null spaces of the population matrix C. The advantage of the plug-in principle is that the asymptotics of (14) are very simple since it is only a Wald statistic. It immediately follows that when the small–b covariance estimator is utilized, $F \xrightarrow{d} \chi^2((m-r)(l-r))$ under $H_0(r)$. When the fixed–b covariance estimator is utilized and $b = 1$, $F \xrightarrow{d} W'(1)\left(2\int_0^1 (W(s) - sW(1))(W(s) - sW(1))'ds\right)^{-1} W(1)$ under $H_0(r)$, where W is a standard Brownian motion of dimension $(m-r)(l-r)$. Limiting distributions for $b \in (0,1)$ can be found in Kiefer & Vogelsang (2005); here we will limit our discussion to

$b = 1$ as this yields the least size distortions (Kiefer & Vogelsang, 2005). Al-Sadoon (2017b) proves that both statistics have an asymptotic power of 1 under $H_1(r)$ and their local power is identical to the Cragg & Donald (1996), Cragg & Donald (1997), and Kleibergen & Paap (2006) counterparts, so there is no loss in efficiency when using the QR statistic.

Of course, it is well known that hypothesis tests based on classical asymptotic theory give poor results in small samples (see e.g. Dufour & Khalaf (2002) and Camba-Mendez et al. (2003)). This is also the case in our setting as we show later on. Therefore, we will use a bootstrap or Monte Carlo testing method, which gives better size control in finite samples. The general form of the testing algorithm follows Dufour et al. (2006) and its asymptotic validity can be established by standard methods (see e.g. Dufour (2006)).

Algorithm 3.1. For a given rank r and horizon h,

1. Compute $\widehat{B}_1$ and $\widehat{\Omega}$.
2. Substituting $\widehat{B}_1$ into (3) and (4) to compute estimates of the first $h-1$ impulse responses, $\widehat{\psi}_j$ for $j = 0, \ldots, h-1$.
3. If $h > 1$, compute $\widehat{B}_h$ and $\widehat{\Xi}_h$.
4. Compute $\widehat{C}$ and $\widehat{\Theta}$.
5. Compute the rank statistic (13) and denote it by F_0.
6. Compute a rank restricted estimator $\overline{B}_h$ (see the discussion below).
7. For $i = 1, \ldots, N$,
   (a) Construct a sample of T observations using $\overline{B}_h$, $\{\widehat{\psi}_j\}_{j=0}^{h-1}$, and $\widehat{\Omega}$ in equation (2) (see the discussion below).
   (b) Compute the statistic (13) for the bootstrapped sample and denote it by F_i.
8. Compute the bootstraped p–value, $\widehat{p}_N = \frac{1}{N+1}\sum_{i=0}^{N} 1(F_i \geq F_0)$, where $1(\cdot)$ is the indicator function.

Two points in the algorithm need further elaboration. First, the bootstrap sample can be generated from either simulated or resampled residuals. In the first case, one obtains the bootstrap shocks by drawing from a multivariate distribution of mean zero and variance $\widehat{\Omega}$ then generates the samples from equation (2) using $\overline{B}_h$ and $\{\widehat{\psi}_j\}$ in place of the population parameters (Dufour & Khalaf (2003) refer to this type of test as a Local Monte Carlo test).

We may also generate the bootstrap shocks non–parametrically by drawing with replacement from the residuals of the regression in step (1). The researcher may also choose to simulate more than T data point to allow for "burn–in" and ensure the data's stationarity. All of these options are available to the researcher in the accompanying Matlab code to this paper.

The second point is that the construction of $\overline{B}_h$ can be carried out in a number of ways. One possibility is to replace $\widehat{C}$ with $\widehat{U}\begin{bmatrix}\widehat{S}_{11} & \widehat{S}_{12}\\ 0 & 0\end{bmatrix}\widehat{V}'$ in $\widehat{B}_h$. In our work, however, we have chosen to use the restricted OLS estimator imposing the restriction $\widehat{N}_r'C=0$ for target SGNC testing and the restriction $C\widehat{M}_r=0$ when testing for predictor SGNC (see equation (40)). The advantage of using the restricted OLS estimator is that it helps avoid generating models with explosive roots. Indeed, these were not encountered in any of our simulations.

Algorithm 3.1 specializes to the one proposed in DPR in the case where $H_0(0)$ is being tested. The author has verified that the algorithm replicates the empirical results of DPR.

Finally, the rank of C can be estimated in a variety of ways. One approach tests sequentially the hypotheses $H_0(0), H_0(1), \ldots, H_0(\min\{m,l\}-1)$ at a particular level of significance α until acceptance. This produces an estimate of the rank that asymptotically never underestimates the true rank but has an asymptotic probability of α of over-estimating. Thus, it is not consistent. It can be made consistent by testing at significance levels that decrease to zero at an appropriate rate with T (Robin & Smith, 2000). However, this and other consistent model selectors (e.g. index minimization in Al-Sadoon (2015)) have the tendency to choose models that are too restricted in small samples. Practitioners usually find it more appealing to exercise judgement in this context, especially when empirically interpretable relationships exist in the data, such as the relationships we found in Examples 2.1 and 2.2. Thus, we opt for the simpler approach of sequential testing proposed by Robin & Smith (2000), which seems to be the preferred approach in empirical cointegration analysis as well (Juselius, 2006; Garratt et al., 2006). Once the rank is estimated as $\widehat{r}$, we can estimate $\mathcal{U}_h^{XY}$ as $\text{span}(\widehat{N}_{\widehat{r}})$ in the case of target SGNC and we can estimate $\mathcal{V}_h^{XY}$ as $\text{span}(\widehat{M}_{\widehat{r}})$ in the case predictor SGNC. See Al-Sadoon (2017b) for more on the estimation of null spaces.

3.2 Estimation and Inference for Non–Stationary VARs

Suppose now that W is allowed to be $I(1)$. In that case, Ξ_h will no longer be invertible and therefore we will not be able to ensure the non–singularity of Θ. Toda & Phillips (1993) give

a detailed analysis of the problem and Lütkepohl (2006) provides a text–book analysis. As a result, the SGNC test may be invalid.

One solution that authors such as Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996) have proposed is to use a lag augmented VAR. These papers have shown that in estimating the model,

$$W(t+1) = \mu^{(1)}(t) + \sum_{j=1}^{p+1} \pi_j^{(1)} W(t+1-j) + a(t+1), \quad t = p+1, \ldots, T, \tag{15}$$

instead of (1) then the estimates of the coefficient matrices $\pi_j^{(1)}$, $1 \leq j \leq p$ are $\sqrt{T}$–consistent and have non–singular asymptotic covariance matrix. Thus Wald tests can proceed as usual. The same reasoning can be adapted to (2), where it is not difficult to show that in the regression,

$$W(t+h) = \mu^{(h)}(t) + \sum_{j=1}^{p+1} \pi_j^{(h)} W(t+1-j) + \sum_{j=0}^{h-1} \psi_j a(t+h-j), \quad t = p+1, \ldots, T-h \tag{16}$$

the estimates of the coefficient matrices $\pi_j^{(h)}$, $1 \leq j \leq p$ are also $\sqrt{T}$–consistent and have non–singular asymptotic covariance matrix. Once these are available, we can proceed as in the previous subsection to draw inference on the rank of C.

Toda & Yamamoto (1995) show that the approach can deal with the general $I(d)$ case just the same, i.e. by augmenting the model with d lags.

4 Empirical Illustration

The empirical problems to which we apply SGNC were introduced in Examples 2.1 and 2.2. Here we extend the analysis and employ the new methodology of SGC.

4.1 The Data and Model Specification

First, we elaborate on our modelling choices. The data includes three series from the Romer & Romer (2004) study, the monetary policy variable that they construct, the log of the producer price index for finished goods (Bureau of Labor Statistics series WPUSOP3000), and the log of the industrial production index (Board of Governors series B50001). To this we have added the log of the civilian unemployment rate (Federal Reserve Economic Data series UNRATENSA) as well as the log of the West Texas Intermediate spot price (Federal Reserve Economic Data

series ID OILPRICE). The data is monthly for the period January 1966 – December 1996 and is not seasonally adjusted.

Next, we checked for stationarity of the individual series. An augmented Dickey–Fuller test rejected the unit root hypothesis for the monetary policy variable. The other variables were visibly at odds with the stationary assumption and were differenced until the augmented Dickey–Fuller test rejected their non–stationarity. Therefore, we constructed the vector process (r, π, y, u, o), which consists of the raw monetary policy variable, the twice differenced log of the producer price index, the differenced log of the industrial production index, the differenced log of the unemployment rate, and the differenced log of oil prices.

Finally, we specified the model as in (1) with a constant and seasonal dummies. The number of lags was selected by minimizing AIC over lags 1–18 and including the seasonal dummies as exogenous variables; this resulted in 12 lags selected. As is well known, AIC tends to be more generous than consistent estimators of lag, which have a tendency to specify too few lags in small samples (McQuarrie & Tsai, 1998). Indeed, the Bayesian–Schwartz and Hannan–Quinn criteria select 1 lag and 3 lags respectively. Intuitively, including too many lags leads to over–rejection in small samples due to there being too many degrees of freedom. On the other hand, including too few of them would invalidate the asymptotic distribution results. We opt, as DPR do, to err on the side of too many lags and show below that over–rejection is not too big an issue for the objective of our study. All of the pretesting mentioned here can be replicated using the STATA file (`PRETESTING.do`), which is part of the software package accompaniment to this paper.

4.2 Size and Power

Before we employ the procedure proposed in this paper, it is important to check that appropriate inference can be drawn based on the sample of interest. Standard practice illustrates size and power in a Monte Carlo experiment, which attempts to emulate the characteristics of empirical data in terms of size, persistence, and other characteristics. However, such Monte Carlo results may be misleading because empirical data can deviate substantially from the Monte Carlo design. Thus, this paper follows DPR and uses the data to decide how well the inference procedure performs. We take it for granted that the large sample approximation holds for large enough samples, and check whether it holds for the sample under study. The

following algorithm estimates the actual size and power of the testing procedures proposed in this paper.

Algorithm 4.1 (Bootstrap Size and Power of a Target (resp. Predictor) SGNC Test)**.** For a given rank r, horizon h, and size α,

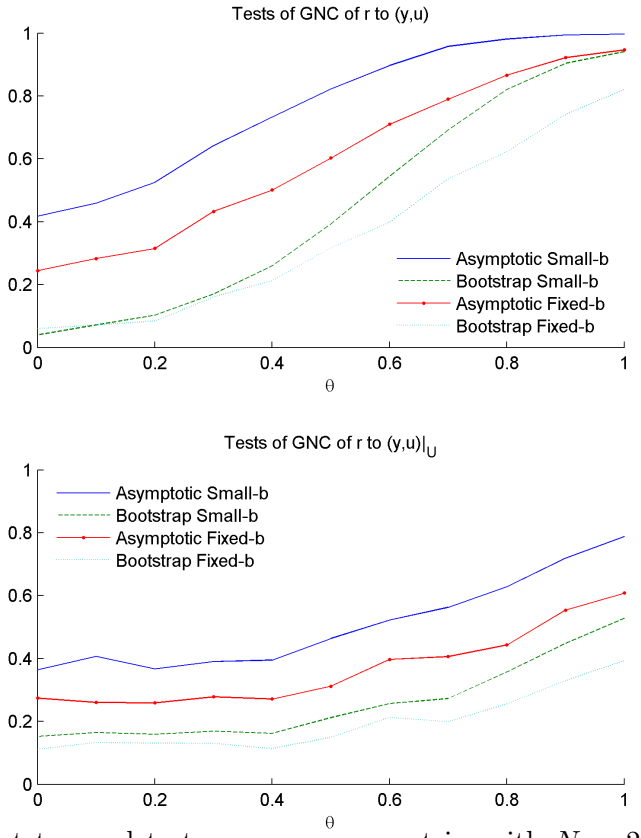
1. Compute $\widehat{B}_1$, $\widehat{\Omega}$, the first $h-1$ impulse responses, $\widehat{\psi}_j$ for $j = 0, \ldots, h-1$, and $\widehat{B}_h$.
2. Construct $\widehat{C}$ and $\widehat{N}_r$ (resp. $\widehat{M}_r$) as described in the previous section.
3. Obtain $\overline{B}_h$ subject to the restriction $Y \nrightarrow_h \widehat{M}_r' X$ (resp. $\widehat{N}_r' Y \nrightarrow_h X$).
4. For $\theta = 0 = \theta_1, \ldots, \theta_c = 1$
   (a) For $i = 1, \ldots, R$,
      i. Construct a sample of T observations using $(1-\theta)\overline{B}_h + \theta\widehat{B}_h$, $\{\widehat{\psi}_j\}_{j=0}^{h-1}$, and $\widehat{\Omega}$ in equation (2).
      ii. Test $H_0(r)$ for $Y \nrightarrow_h X|_{\mathcal{U}}$ (resp. $Y|_{\mathcal{V}} \nrightarrow_h X$) at significance α.
   (b) Compute the rejection rate of each test for the given θ.

The rejection rate at $\theta = 0$ is the empirical size of the test and gives us an indication of how well the test performs under the null. The rejection rate for $\theta = 1$ is the rejection rate at the estimated set of parameters and gives us an idea of the small-sample power of the test.

The Monte Carlo test in Algorithm 4.1 differs from the design utilized by DPR in that they impose GNC across a range of horizons in step 3, whereas we impose GNC only at a single horizon. The advantage of the DPR design is that it allows us to see the ability of the tests to be detect GC across a range of horizons under the alternative. However, the design of Algorithm 4.1 is more representative of both the null and the alternative hypotheses usually considered in practice and is easier to implement. Practitioners are recommended to run the simulations above for each particular test of interest. This is easily done using the accompanying Matlab code `SSP_SGNC.m`.

To conserve space, we illustrate by considering a small set of null hypotheses to test: $r \nrightarrow_1 (y,u)$, $(r,o) \nrightarrow_1 u$, $r \nrightarrow_1 (y,u)|_{\mathcal{U}}$, and $(r,o)|_{\mathcal{V}} \nrightarrow_1 u$. Results for analogous hypotheses at higher horizons paint a similar picture of the performance of the tests under consideration. We use $\alpha = 0.05$ and $R = 1000$ in Algorithm 4.1 to study the small–sample behaviour of the asymptotic small–b test, asymptotic fixed–b test, bootstrapped small–b test, and bootstrapped

Figure 5: Rejection Rates for SGNC Tests of r to (y, u) at Horizon 1.

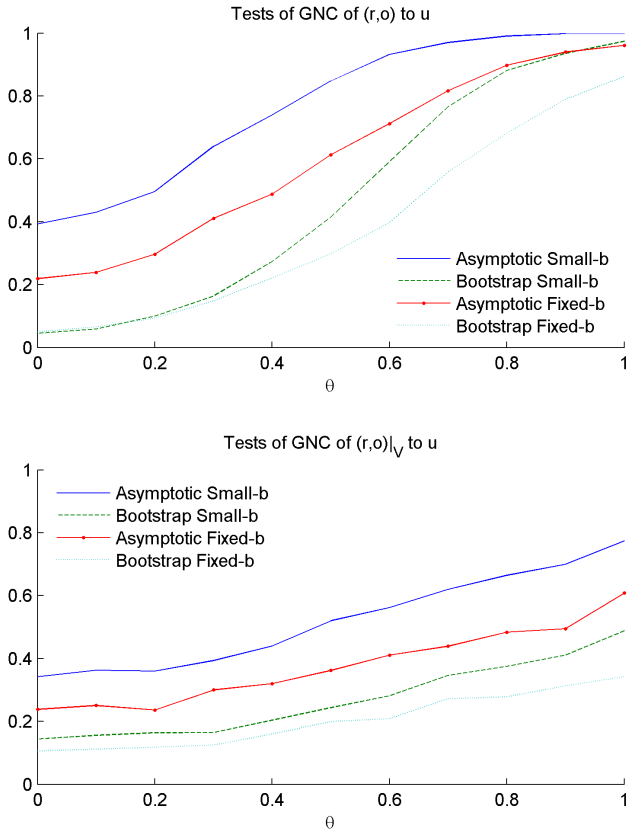


fixed–b test. The bootstrapped tests are non–parametric, with $N = 2000$ and a burn–in of 100 periods. The results are plotted in Figures 5 and 6.

The asymptotic tests have a serious over–rejection problem, with fixed–b tests significantly improving on small–b tests but without successfully matching the nominal size. The bootstrap versions of the tests control size much better across the four hypotheses tested. The bootstrapped tests of $r \nrightarrow_1 (y, u)$ and $(r, o) \nrightarrow_1 u$ have good size and power properties, with the bootstrapped small–b tests having higher power than the bootstrapped fixed–b tests. However, the bootstrapped tests of $r \nrightarrow_1 (y, u)|_{\mathcal{U}}$, and $(r, o)|_{\mathcal{V}} \nrightarrow_1 u$ are moderately oversized, with the bootstrapped fixed–b tests closer to the nominal size.

To summarize, asymptotic tests are to be avoided in favour of the alternative bootstrapping procedures. We can be confident about using the bootstrapped tests for testing $r \nrightarrow_h (y, u)$ and $(r, o) \nrightarrow_h u$ but must be cautious when testing either $r \nrightarrow_h (y, u)|_{\mathcal{U}}$ or $(r, o)|_{\mathcal{V}} \nrightarrow_h u$ because of the problem of over–rejection. Luckily, in our sample, none of the tests for $r \nrightarrow_h (y, u)|_{\mathcal{U}}$ or $(r, o)|_{\mathcal{V}} \nrightarrow_h u$ were rejected, so we need not worry about the over–rejection problem in this context.

Figure 6: Rejection Rates for SGNC Tests of (r, o) to u at Horizon 1.



4.3 Results

Given the size and power results above, we employ Algorithm 3.1 to find bootstrapped p–values based on small–b and fixed–b statistics to study target SGNC between r and (y, u) and predictor SGNC between (r, o) and u. We will base our inference primarily on the bootstrapped small–b test, except when the size of the test is in question, in which case we will consider the bootstrapped fixed–b test. We will utilize a non–parametric bootstrap with $N = 2000$, a burn–in of 100 periods, and test at the conventional 5% significance.

We begin by considering the univariate predictive effects (Table 1). We see that monetary policy predicts the growth of industrial production over the entire range of horizons we consider, 1–12. It predicts the growth of unemployment over horizons 4 and 5. On the other hand, oil price growth predicts unemployment growth over horizons 1, 3, and 4.

Consider next the target SGNC results given in Table 2(a). The results confirm our graphical analysis in Example 2.1. Monetary policy predicts negative comovements of industrial production and unemployment growth across a range of horizons. The trade–off between industrial production and unemployment growths predicted by monetary policy is estimated

Table 1: Univariate Results

Bootstrapped p–Values for Small–b SGNC Tests												
h	1	2	3	4	5	6	7	8	9	10	11	12
$r \nrightarrow_h y$	**0.005**	**0.008**	**0.014**	**0.044**	**0.009**	**0.005**	**0.007**	**0.004**	**0.015**	**0.009**	**0.011**	**0.007**
$r \nrightarrow_h u$	0.062	0.093	0.096	**0.008**	**0.036**	0.106	0.052	0.076	0.117	0.056	0.631	0.873
$o \nrightarrow_h u$	**0.039**	0.069	**0.039**	**0.044**	0.077	0.111	0.107	0.101	0.195	0.220	0.649	0.395
Bootstrapped p–Values for Fixed–b SGNC Tests												
$r \nrightarrow_h y$	**0.011**	**0.034**	0.226	0.199	**0.007**	**0.003**	**0.007**	**0.010**	**0.011**	**0.011**	0.076	**0.046**
$r \nrightarrow_h u$	0.091	0.109	0.381	0.620	0.143	0.214	0.180	0.129	0.078	0.192	0.644	0.771
$o \nrightarrow_h u$	0.156	0.271	0.136	0.106	0.124	0.348	0.075	**0.037**	**0.042**	0.158	0.416	0.297

at about 3% higher unemployment for every 1% fall of industrial production over horizons 1–5. This trade–off becomes quite severe at horizons 6 but falls gradually after that. The significance of these trade–offs follows from the univariate tests in Table 1. In particular, the estimated trade–off between u and y at horizon h is zero if and only if $r \nrightarrow_h u$ and this is rejected at horizons 4 and 5.

Consider next the predictor SGNC results given in Table 2(b). The results again confirm the graphical analysis of Example 2.2. There is a statistical reaction function of monetary policy to oil prices. Observed decreases of r of around 0.15–0.20% in response to 1% increase in oil prices have no predictive effects on unemployment growth. From Table 1, we see that indeed the trade–offs at horizons 1, 3, and 4 are statistically significant.

Visual inspection of the series π, y, u, and o does not yield anything too alarming about the stationarity assumption. One may, however, have misgivings about considering r to be stationary. In that case, we employ the methods of subsection 3.2. The results of these tests are given in Tables 3 and 4. Clearly, the qualitative empirical conclusions under stationarity remain intact when we employ our $I(1)$–robust method.

5 Conclusion

In this paper, we have presented an extension of GC that allows the researcher to estimate the directionality or the subspaces of GC. These subspaces have been shown to admit interesting empirical interpretations as conditional predictability trade–offs. The method was illustrated both graphically and statistically. In the remainder, we mention some possible venues of future research.

Table 2: Multivariate Results

Table 2(a). Bootstrapped p-values for Target SGNC Tests of $r \nrightarrow_h (y, u)$ and Subspaces of SGNC for the Horizons 1-12.

Small–b SGNC Test Results

h	1	2	3	4	5	6	7	8	9	10	11	12
$H_0(0)$	**0.022**	**0.001**	**0.004**	**0.015**	**0.011**	**0.043**	**0.010**	**0.002**	**0.031**	0.053	0.091	**0.035**
$H_0(1)$	0.788	0.460	0.329	0.398	0.492	0.239	0.261	0.172	0.159	0.143	0.004	0.941
$\mathcal{U}_h^{XY}$	$\begin{bmatrix}0.9448\\0.3276\end{bmatrix}$	$\begin{bmatrix}0.9471\\0.3209\end{bmatrix}$	$\begin{bmatrix}0.9594\\0.2821\end{bmatrix}$	$\begin{bmatrix}0.9622\\0.2722\end{bmatrix}$	$\begin{bmatrix}0.9436\\0.3311\end{bmatrix}$	$\begin{bmatrix}0.9886\\0.1508\end{bmatrix}$	$\begin{bmatrix}0.9858\\0.1681\end{bmatrix}$	$\begin{bmatrix}0.9763\\0.2163\end{bmatrix}$	$\begin{bmatrix}0.9749\\0.2224\end{bmatrix}$	I_2	I_2	$\begin{bmatrix}0.8757\\0.4829\end{bmatrix}$
(y, u) Tradeoff	2.8841	2.9518	3.4011	3.5350	2.8502	6.5542	5.8634	4.5137	4.3832	NA	NA	1.8136

Fixed–b SGNC Test Results

h	1	2	3	4	5	6	7	8	9	10	11	12
$H_0(0)$	**0.046**	0.070	0.092	0.590	**0.011**	**0.027**	**0.010**	**0.011**	**0.039**	0.078	0.248	**0.027**
$H_0(1)$	0.506	0.335	0.107	0.339	0.222	0.146	0.184	0.313	0.115	0.219	0.002	0.827
$\mathcal{U}_h^{XY}$	$\begin{bmatrix}0.9448\\0.3276\end{bmatrix}$	I_2	I_2	I_2	$\begin{bmatrix}0.9436\\0.3311\end{bmatrix}$	$\begin{bmatrix}0.9886\\0.1508\end{bmatrix}$	$\begin{bmatrix}0.9858\\0.1681\end{bmatrix}$	$\begin{bmatrix}0.9763\\0.2163\end{bmatrix}$	$\begin{bmatrix}0.9749\\0.2224\end{bmatrix}$	I_2	I_2	$\begin{bmatrix}0.8757\\0.4829\end{bmatrix}$
(y, u) Tradeoff	2.8841	NA	NA	NA	2.8502	6.5542	5.8634	4.5137	4.3832	NA	NA	1.8136

Table 2(b). Bootstrapped p-values for Predictor SGNC Tests of $(r, o) \nrightarrow_h u$ and Subspaces of SGNC for the Horizons 1-12.

Small–b SGNC Test Results

h	1	2	3	4	5	6	7	8	9	10	11	12
$H_0(0)$	**0.003**	**0.032**	**0.021**	**0.003**	**0.016**	0.089	0.088	0.076	0.164	0.112	0.501	0.607
$H_0(1)$	0.654	0.322	0.551	0.641	0.616	0.823	0.897	0.864	0.677	0.937	0.823	0.974
$\mathcal{V}_h^{XY}$	$\begin{bmatrix}-0.9823\\0.1874\end{bmatrix}$	$\begin{bmatrix}0.9968\\0.0801\end{bmatrix}$	$\begin{bmatrix}-0.9755\\0.2199\end{bmatrix}$	$\begin{bmatrix}-0.9832\\0.1824\end{bmatrix}$	$\begin{bmatrix}-0.9881\\0.1536\end{bmatrix}$	I_2	I_2	I_2	I_2	I_2	I_2	I_2
(r, o) Trade–off	0.1908	-0.0804	0.2254	0.1855	0.1554	NA	NA	NA	NA	NA	NA	NA

Fixed–b SGNC Test Results

h	1	2	3	4	5	6	7	8	9	10	11	12
$H_0(0)$	**0.030**	0.194	0.061	0.068	**0.031**	0.139	**0.049**	**0.035**	**0.046**	0.152	0.483	0.606
$H_0(1)$	0.501	0.481	0.527	0.440	0.364	0.747	0.855	0.736	0.645	0.911	0.833	0.955
$\mathcal{V}_h^{XY}$	$\begin{bmatrix}-0.9823\\0.1874\end{bmatrix}$	I_2	I_2	I_2	$\begin{bmatrix}-0.9881\\0.1536\end{bmatrix}$	I_2	$\begin{bmatrix}-0.9783\\0.2072\end{bmatrix}$	$\begin{bmatrix}-0.9832\\0.1825\end{bmatrix}$	$\begin{bmatrix}-0.9873\\0.1590\end{bmatrix}$	I_2	I_2	I_2
(r, o) Trade–off	0.1908	NA	NA	NA	0.1554	NA	0.2118	0.1856	0.1611	NA	NA	NA

Table 3: $I(1)$–Robust Univariate Results

Bootstrapped p–Values for Small–b SGNC Tests												
h	1	2	3	4	5	6	7	8	9	10	11	12
$r \nrightarrow_h y$	**0.015**	**0.006**	**0.004**	0.082	**0.008**	**0.004**	**0.006**	**0.008**	**0.030**	**0.007**	**0.019**	**0.016**
$r \nrightarrow_h u$	0.084	0.155	**0.021**	**0.027**	**0.044**	0.051	0.052	0.264	**0.023**	**0.050**	0.715	0.922
$o \nrightarrow_h u$	**0.029**	**0.040**	**0.010**	**0.045**	0.217	0.137	0.081	0.106	0.337	0.249	0.632	0.394
Bootstrapped p–Values for Fixed–b SGNC Tests												
$r \nrightarrow_h y$	**0.004**	0.075	0.169	0.060	**0.005**	**0.002**	**0.013**	**0.009**	**0.019**	**0.040**	0.130	0.138
$r \nrightarrow_h u$	0.221	0.194	0.383	0.568	0.193	0.219	0.213	0.141	**0.049**	0.182	0.706	0.854
$o \nrightarrow_h u$	0.246	0.124	0.060	0.127	0.326	0.261	**0.043**	**0.024**	0.103	0.123	0.417	0.314

There are many possible applications of SGNC besides the problem of finding economically interpretable relationships in the data. We mention three:

1. Imposing GNC and SGNC restrictions can have forecasting benefits in terms of reducing estimation error. Jarociński & Maćkowiak (2017) conducts such an exercise using GNC restrictions. Velu et al. (1986) and Camba-Mendez et al. (2003), on the other hand, impose restrictions that can be interpreted as SGNC restriction (Al-Sadoon, 2014). These GNC or SGNC restrictions become even more attractive when they admit economic interpretations (e.g. Example 2.1 and 2.2) as structural restrictions have been shown to improve the performance of empirical models (Garratt et al., 2006).

2. Dynamic structural models such as dynamic stochastic general equilibrium models can imply testable SGC restrictions (Al-Sadoon, 2014, 2017a). Thus, SGNC tests can be used as specification tests for these structural models.

3. The moments describing SGC describe dynamic interdependence between time series. Because of their dynamic nature, they may be more natural moments to match in calibration exercises than the unconditional moments prevalent in the calibration literature.

Although the procedure outlined in this paper can easily be extended to test causality up to horizon h, rather than just at a particular horizon h, there is still need for a simple long run causality test. Bruneau & Jondeau (1999) proposed such a test for cointegrated VARs. Unfortunately, Yamamoto & Kurozumi (2006) have found that the multivariate extension of the statistic can suffer from the same singularity issue we have considered in subsection 3.2. They propose a two–step procedure that estimates the rank of Θ then uses a generalized inverse. Clearly, a simpler solution is desirable.

Table 4: $I(1)$–Robust Multivariate Results.

Table 4(a). Bootstrapped p-values for Predictor SGNC Tests of $(r, o) \nrightarrow_h u$ and Subspaces of SGNC for the Horizons 1-12.

Small–b SGNC Test Results

h	1	2	3	4	5	6	7	8	9	10	11	12
$H_0(0)$	**0.024**	**0.005**	**0.003**	**0.026**	**0.017**	**0.024**	**0.036**	**0.016**	**0.028**	**0.032**	0.065	0.060
$H_0(1)$	0.300	0.275	0.220	0.754	0.728	0.247	0.311	0.338	0.133	0.177	0.002	0.960
$\mathcal{U}_h^{XY}$	$\begin{bmatrix}0.9783\\0.2073\end{bmatrix}$	$\begin{bmatrix}0.9584\\0.2854\end{bmatrix}$	$\begin{bmatrix}0.9638\\0.2667\end{bmatrix}$	$\begin{bmatrix}0.9511\\0.3089\end{bmatrix}$	$\begin{bmatrix}0.9495\\0.3136\end{bmatrix}$	$\begin{bmatrix}0.9883\\0.1524\end{bmatrix}$	$\begin{bmatrix}0.9827\\0.1854\end{bmatrix}$	$\begin{bmatrix}0.9723\\0.2339\end{bmatrix}$	$\begin{bmatrix}0.9773\\0.2120\end{bmatrix}$	$\begin{bmatrix}0.9792\\0.2030\end{bmatrix}$	I_2	I_2
(y, u) Tradeoff	4.7194	3.3585	3.6141	3.0792	3.0274	6.4843	5.3000	4.1561	4.6095	4.8233	NA	NA

Fixed–b SGNC Test Results

$H_0(0)$	**0.022**	0.079	0.108	0.372	**0.020**	**0.041**	**0.023**	**0.032**	**0.036**	0.054	0.115	**0.030**
$H_0(1)$	0.287	0.099	0.199	0.592	0.422	0.210	0.096	0.340	0.133	0.136	0.001	0.880
$\mathcal{U}_h^{XY}$	$\begin{bmatrix}0.9783\\0.2073\end{bmatrix}$	I_2	I_2	I_2	$\begin{bmatrix}0.9495\\0.3136\end{bmatrix}$	$\begin{bmatrix}0.9883\\0.1524\end{bmatrix}$	$\begin{bmatrix}0.9827\\0.1854\end{bmatrix}$	$\begin{bmatrix}0.9723\\0.2339\end{bmatrix}$	$\begin{bmatrix}0.9773\\0.2120\end{bmatrix}$	I_2	I_2	$\begin{bmatrix}0.8699\\0.4932\end{bmatrix}$
(y, u) Tradeoff	4.7194	NA	NA	NA	3.0274	6.4843	5.3000	4.1561	4.6095	NA	NA	1.7639

Table 4(b). Bootstrapped p-values for Predictor SGNC Tests of $(r, o)|_{\mathcal{V}} \nrightarrow_h u$ and Subspaces of SGNC for the Horizons 1-12.

Small–b SGNC Test Results

h	1	2	3	4	5	6	7	8	9	10	11	12
$H_0(0)$	**0.005**	**0.041**	**0.005**	**0.007**	**0.044**	0.065	0.070	0.125	0.150	0.097	0.435	0.561
$H_0(1)$	0.447	0.302	0.434	0.608	0.693	0.795	0.873	0.829	0.661	0.538	0.828	0.967
$\mathcal{V}_h^{XY}$	$\begin{bmatrix}-0.9883\\0.1528\end{bmatrix}$	$\begin{bmatrix}0.9976\\0.0696\end{bmatrix}$	$\begin{bmatrix}-0.9831\\0.1828\end{bmatrix}$	$\begin{bmatrix}-0.9877\\0.1565\end{bmatrix}$	$\begin{bmatrix}-0.9846\\0.1747\end{bmatrix}$	I_2	I_2	I_2	I_2	I_2	I_2	I_2
(r, o) Trade–off	0.1546	-0.0697	0.1859	0.1585	0.1774	NA	NA	NA	NA	NA	NA	NA

Fixed–b SGNC Test Results

$H_0(0)$	0.081	0.204	**0.047**	0.060	0.129	0.065	0.068	0.051	0.088	0.129	0.360	0.512
$H_0(1)$	0.349	0.655	0.551	0.408	0.478	0.764	0.859	0.569	0.795	0.362	0.807	0.975
$\mathcal{V}_h^{XY}$	I_2	I_2	$\begin{bmatrix}-0.9831\\0.1828\end{bmatrix}$	I_2	I_2	I_2	I_2	I_2	I_2	I_2	I_2	I_2
(r, o) Trade–off	NA	NA	0.1859	NA	NA	NA	NA	NA	NA	NA	NA	NA

A Non-Parametric Subspace Granger Causality

It is well known in the multivariate statistics literature that rank testing in a regression context is related to testing the significance of the smallest canonical correlations (Reinsel & Velu, 1998; Anderson, 2003). We now show that SGNC can be studied using the method of partial canonical correlations proposed by Reinsel (2003).

Suppose we have random vectors $\mathcal{X} \in \mathbb{R}^n$, $\mathcal{Z} \in \mathbb{R}^k$, $\mathcal{Y}_i \in \mathbb{R}^m$ for $i = 1, \ldots, p$, $\mathcal{Y} = (\mathcal{Y}_1', \ldots, \mathcal{Y}_p')'$, and let the variance matrix of $(\mathcal{X}', \mathcal{Y}', \mathcal{Z}')'$ be

$$\Sigma = \begin{bmatrix} \Sigma_{\mathcal{X}\mathcal{X}} & \Sigma_{\mathcal{X}\mathcal{Y}} & \Sigma_{\mathcal{X}\mathcal{Z}} \\ \Sigma_{\mathcal{Y}\mathcal{X}} & \Sigma_{\mathcal{Y}\mathcal{Y}} & \Sigma_{\mathcal{Y}\mathcal{Z}} \\ \Sigma_{\mathcal{Z}\mathcal{X}} & \Sigma_{\mathcal{Z}\mathcal{Y}} & \Sigma_{\mathcal{Z}\mathcal{Z}} \end{bmatrix} = \begin{bmatrix} \Sigma_{\mathcal{X}\mathcal{X}} & \Sigma_{\mathcal{X}\mathcal{Y}_1} & \cdots & \Sigma_{\mathcal{X}\mathcal{Y}_p} & \Sigma_{\mathcal{X}\mathcal{Z}} \\ \Sigma_{\mathcal{Y}_1\mathcal{X}} & \Sigma_{\mathcal{Y}_1\mathcal{Y}_1} & \cdots & \Sigma_{\mathcal{Y}_1\mathcal{Y}_p} & \Sigma_{\mathcal{Y}_1\mathcal{Z}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Sigma_{\mathcal{Y}_p\mathcal{X}} & \Sigma_{\mathcal{Y}_p\mathcal{Y}_1} & \cdots & \Sigma_{\mathcal{Y}_p\mathcal{Y}_p} & \Sigma_{\mathcal{Y}_p\mathcal{Z}} \\ \Sigma_{\mathcal{Z}\mathcal{X}} & \Sigma_{\mathcal{Z}\mathcal{Y}_1} & \cdots & \Sigma_{\mathcal{Z}\mathcal{Y}_p} & \Sigma_{\mathcal{Z}\mathcal{Z}} \end{bmatrix}.$$

Neither Σ nor any of its components are assumed to have any particular rank. In this setting, $\mathcal{X}$ takes the role of $X(t+h)$, $\mathcal{Y}$ takes the role of $(Y'(t), Y'(t-1), \ldots, Y'(t+1-p))'$, and $\mathcal{Z}$ takes the role of $(X'(t), X'(t-1), \ldots, X'(t+1-p), Z'(t), Z'(t-1), \ldots, Z'(t+1-p))'$.

The Frisch–Waugh theorem implies that the best linear predictor of $\mathcal{X}$ in terms of $\mathcal{Y}$ and $\mathcal{Z}$ is $\Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}}\mathcal{Y} + \Sigma_{\mathcal{X}\mathcal{Z}\cdot\mathcal{Y}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}\cdot\mathcal{Y}}\mathcal{Z}$, where $\Sigma^{\dagger}_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}}$ is the Moore–Penrose inverse of $\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}}$ and the partial covariance matrices are given by

$$\Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = \Sigma_{\mathcal{X}\mathcal{Y}} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Y}} \qquad \Sigma_{\mathcal{X}\mathcal{Z}\cdot\mathcal{Y}} = \Sigma_{\mathcal{X}\mathcal{Z}} - \Sigma_{\mathcal{X}\mathcal{Y}}\Sigma^{\dagger}_{\mathcal{Y}\mathcal{Y}}\Sigma_{\mathcal{Y}\mathcal{Z}}$$
$$\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}} = \Sigma_{\mathcal{Y}\mathcal{Y}} - \Sigma_{\mathcal{Y}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Y}} \qquad \Sigma_{\mathcal{Z}\mathcal{Z}\cdot\mathcal{Y}} = \Sigma_{\mathcal{Z}\mathcal{Z}} - \Sigma_{\mathcal{Z}\mathcal{Y}}\Sigma^{\dagger}_{\mathcal{Y}\mathcal{Y}}\Sigma_{\mathcal{Y}\mathcal{Z}}.$$

It follows that $\mathcal{Y}$ fails to predict $\mathcal{X}$ conditionally on $\mathcal{Z}$ along the left null space of $\Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}}$. This subspace corresponds directly to the subspace of target GNC. Next we will see how it may be obtained from the partial canonical correlations point of view.

Suppose we are interested in directions along which $\mathcal{X}$ and $\mathcal{Y}$ have the strongest correlation after conditioning on $\mathcal{Z}$. Thus, we are interested in the directions of strongest correlation between $\mathcal{X} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\mathcal{Z}$ and $\mathcal{Y} - \Sigma_{\mathcal{Y}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\mathcal{Z}$ and we must solve for

$$\begin{aligned}
\rho^1_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} &= \sup\{|\mathrm{corr}(\mathcal{U}, \mathcal{V})| : \mathcal{U} = x'(\mathcal{X} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\mathcal{Z}), \mathcal{V} = y'(\mathcal{Y} - \Sigma_{\mathcal{Y}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\mathcal{Z})\} \\
&= \sup\{|\mathrm{cov}(\mathcal{U}, \mathcal{V})| : \mathcal{U} = x'(\mathcal{X} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\mathcal{Z}), \mathcal{V} = y'(\mathcal{Y} - \Sigma_{\mathcal{Y}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\mathcal{Z}), \mathrm{var}(\mathcal{U}) = \mathrm{var}(\mathcal{V}) = 1\} \\
&= \max\{x'\Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}}y : x \in \mathbb{R}^n, y \in \mathbb{R}^{mp}, x'\Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}}x = y'\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}}y = 1\}.
\end{aligned}$$

This expression is identical to its counterpart in canonical correlation analysis except that the covariance matrices are replaced by partial covariance matrices. Solutions x_1 and y_1 to the above maximization problem are then used to find the canonical variates, $\mathcal{U}_1 = x_1'(\mathcal{X} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Z}}^{\dagger}\mathcal{Z})$ and $\mathcal{V}_1 = y_1'(\mathcal{Y} - \Sigma_{\mathcal{Y}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Z}}^{\dagger}\mathcal{Z})$ so that finally, $\rho^1_{XY\cdot Z} = \mathrm{cov}(\mathcal{U}_1, \mathcal{V}_1)$.

The next set of canonical variates is found by looking for the directions of maximum correlation between $\mathcal{X} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Z}}^{\dagger}\mathcal{Z}$ and $\mathcal{Y} - \Sigma_{\mathcal{Y}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Z}}^{\dagger}\mathcal{Z}$ among all possible directions uncorrelated with $\mathcal{U}_1$ and $\mathcal{V}_1$. Thus, we solve for

$$\rho^{i+1}_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = \sup\{\mathrm{corr}(\mathcal{U},\mathcal{V}) : \mathcal{U} = x'(\mathcal{X} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Z}}^{\dagger}\mathcal{Z}), \mathrm{cov}(\mathcal{U},\mathcal{U}_j) = 0, j = 1,\dots i,$$
$$\mathcal{V} = y'(\mathcal{Y} - \Sigma_{\mathcal{Y}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Z}}^{\dagger}\mathcal{Z}), \mathrm{cov}(\mathcal{V},\mathcal{V}_j) = 0, j = 1,\dots i\},$$

for $i \geq 1$ and this similarly reduces to

$$\rho^{i+1}_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = \max\{x'\Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}}y : x \in \mathbb{R}^n, y \in \mathbb{R}^{mp}, x'\Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}}x = y'\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}}y = 1,$$
$$x'\Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}}x_j = y'\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}}y_j = 0, j = 1,\dots,i\}.$$

This procedure terminates after $\min\{n, mp\}$ steps and obtains as many canonical correlations and pairs of canonical variates. Following Anderson (2003), the solution to the algorithm can be represented as the solutions to the eigenvalue problems,

$$\begin{bmatrix} -\lambda_i\Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}} & \Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} \\ \Sigma_{\mathcal{Y}\mathcal{X}\cdot\mathcal{Z}} & -\lambda_i\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad x_i'\Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}}x_j = \delta_{ij}, \qquad y_i'\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}}y_j = \delta_{ij},$$

$$\mathcal{U}_i = x_i'(\mathcal{X} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Z}}^{\dagger}\mathcal{Z}), \quad \mathcal{V}_i = y_i'(\mathcal{Y} - \Sigma_{\mathcal{Y}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Z}}^{\dagger}\mathcal{Z}), \quad \rho^i_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = \lambda_i = \mathrm{cov}(\mathcal{U}_i, \mathcal{V}_i), \qquad (17)$$

for $i, j = 1, \dots, \min\{n, mp\}$. The existence of the canonical variates in this case follows from standard linear algebra techniques.

Clearly the canonical variates associated with canonical correlations of zero define directions of uncorrelated conditional variation between $\mathcal{X}$ and $\mathcal{Y}$. That is, $\{x_i : \rho^i_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = 0, i = 1, \dots, \min\{n, mp\}\}$ are the directions along which variations in $\mathcal{X}$ are not attributable to variations in $\mathcal{Y}$ after controlling for $\mathcal{Z}$. This can easily be seen from equation (17) where if $\rho^i_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = \lambda_i = 0$ then $x_i'\Sigma_{\mathcal{X}\mathcal{Y}\cdot Z} = 0$. Thus, along the subspace, $\mathrm{span}\{x_i : \rho^i_{XY\cdot Z} = 0, i = 1, \dots, \min\{n, mp\}\}$, $\mathcal{Y}$ cannot help predict $\mathcal{X}$ over and above the predictive ability of $\mathcal{Z}$. The space spanned by these vectors is $\mathcal{U}_h^{XY}$ in the context of target SGNC.

Suppose that instead we are interested in the components of $\mathcal{Y}_i$ that best predict $\mathcal{X}$ after conditioning on $\mathcal{Z}$. In order to study this correlation, we need a device that allows us to

look at the correlation of $\mathcal{X}$ with the individual components of $\mathcal{Y}$. Thus we will consider canonical correlations between $\widetilde{\mathcal{X}} = (\phi \otimes I_n)\mathcal{X}$ and $\widetilde{\mathcal{Y}} = (\phi \otimes I_m)'\mathcal{Y}$, where the random vector $\phi = (\phi_1, \ldots, \phi_p)'$ is independent of, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, and satisfies $\mathbb{E}(\phi\phi') = I_p$. We will also need $\widetilde{\mathcal{Z}} = (\phi \otimes I_k)\mathcal{Z}$. This construction makes sense because the covariance between $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}}$ is,

$$\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}} = \begin{bmatrix} \Sigma_{\mathcal{X}\mathcal{Y}_1} \\ \vdots \\ \Sigma_{\mathcal{X}\mathcal{Y}_p} \end{bmatrix},$$

which is the matrix that describes the joint covariation of the components of $\mathcal{Y}$ with $\mathcal{X}$. The matrix that describes this covariation after factoring out the effect of $\mathcal{Z}$ is $\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}} = \Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}} - \Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Z}}}\Sigma^{\dagger}_{\widetilde{\mathcal{Z}}\widetilde{\mathcal{Z}}}\Sigma_{\widetilde{\mathcal{Z}}\widetilde{\mathcal{Y}}}$ and it is easy to check that it simplifies to,

$$\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}} = \begin{bmatrix} \Sigma_{\mathcal{X}\mathcal{Y}_1\cdot\mathcal{Z}} \\ \vdots \\ \Sigma_{\mathcal{X}\mathcal{Y}_p\cdot\mathcal{Z}} \end{bmatrix} = \begin{bmatrix} \Sigma_{\mathcal{X}\mathcal{Y}_1} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Y}_1} \\ \vdots \\ \Sigma_{\mathcal{X}\mathcal{Y}_p} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\Sigma_{\mathcal{Z}\mathcal{Y}_p} \end{bmatrix}$$

Similarly, we find that $\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{X}}\cdot\widetilde{\mathcal{Z}}} = (I_p \otimes \Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}})$, while $\Sigma_{\widetilde{\mathcal{Y}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}} = \sum_{i=1}^{p} \Sigma_{\mathcal{Y}_i\mathcal{Y}_i\cdot\mathcal{Z}}$.

Now define the first canonical correlation analogously to the above as,

$$\theta^1_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = \rho^1_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}} = \max\{\widetilde{x}'\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}}\widetilde{y} : \widetilde{x} \in \mathbb{R}^{np}, \widetilde{y} \in \mathbb{R}^m, \widetilde{x}'\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{X}}\cdot\widetilde{\mathcal{Z}}}\widetilde{x} = \widetilde{y}'\Sigma_{\widetilde{\mathcal{Y}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}}\widetilde{y} = 1\}.$$

Having found the first canonical correlation, $\theta^1_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}}$, the optimal vectors $\widetilde{x}_1$ and $\widetilde{y}_1$, and the associated canonical variates $\widetilde{U}_1 = \widetilde{x}_1'(\widetilde{\mathcal{X}} - \Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Z}}}\Sigma^{\dagger}_{\widetilde{\mathcal{Z}}\widetilde{\mathcal{Z}}}\widetilde{\mathcal{Z}})$ and $\widetilde{V}_1 = \widetilde{y}_1'(\widetilde{\mathcal{Y}} - \Sigma_{\widetilde{\mathcal{Y}}\widetilde{\mathcal{Z}}}\Sigma^{\dagger}_{\widetilde{\mathcal{Z}}\widetilde{\mathcal{Z}}}\widetilde{\mathcal{Z}})$, we proceed recursively for $i \geq 1$ as,

$$\theta^{i+1}_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = \rho^{i+1}_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}} = \max\{\widetilde{x}'\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}}\widetilde{y} : \widetilde{x} \in \mathbb{R}^{np}, \widetilde{y} \in \mathbb{R}^m, \widetilde{x}'\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{X}}\cdot\widetilde{\mathcal{Z}}}\widetilde{x} = \widetilde{y}'\Sigma_{\widetilde{\mathcal{Y}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}}\widetilde{y} = 1,$$
$$\widetilde{x}'\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{X}}\cdot\widetilde{\mathcal{Z}}}\widetilde{x}_j = \widetilde{y}'\Sigma_{\widetilde{\mathcal{Y}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}}\widetilde{y}_j = 0, j = 1, \ldots, i\}.$$

The problem then reduces to solving the linear set of equations,

$$\begin{bmatrix} -\lambda_i\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{X}}\cdot\widetilde{\mathcal{Z}}} & \Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}} \\ \Sigma_{\widetilde{\mathcal{Y}}\widetilde{\mathcal{X}}\cdot\widetilde{\mathcal{Z}}} & -\lambda_i\Sigma_{\widetilde{\mathcal{Y}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}} \end{bmatrix} \begin{bmatrix} \widetilde{x}_i \\ \widetilde{y}_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \widetilde{x}_i'\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{X}}\cdot\widetilde{\mathcal{Z}}}\widetilde{x}_j = \delta_{ij}, \qquad \widetilde{y}_i'\Sigma_{\widetilde{\mathcal{Y}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}}\widetilde{y}_j = \delta_{ij},$$
$$\widetilde{\mathcal{U}}_i = \widetilde{x}_i'(\widetilde{\mathcal{X}} - \Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Z}}}\Sigma^{\dagger}_{\widetilde{\mathcal{Z}}\widetilde{\mathcal{Z}}}\widetilde{\mathcal{Z}}), \quad \widetilde{\mathcal{V}}_i = \widetilde{y}_i'(\widetilde{\mathcal{Y}} - \Sigma_{\widetilde{\mathcal{Y}}\widetilde{\mathcal{Z}}}\Sigma^{\dagger}_{\widetilde{\mathcal{Z}}\widetilde{\mathcal{Z}}}\widetilde{\mathcal{Z}}), \quad \theta^i_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}} = \lambda_i = \mathrm{cov}(\widetilde{\mathcal{U}}_i, \widetilde{\mathcal{V}}_i), \qquad (18)$$

for $i, j = 1, \ldots, \min\{np, m\}$.

Again, the canonical variates associate with canonical correlations of zero define directions of uncorrelated conditional variation between $\mathcal{X}$ and the components of $\mathcal{Y}$. That is, $\{\widetilde{y}_i$:

$\theta^i_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = 0, i = 1, \ldots, \min\{np, m\}\}$ are the directions along which variations in $\mathcal{X}$ are not attributable to the variations of the components of $\mathcal{Y}$ after controlling for $\mathcal{Z}$. This can easily be seen from equation (18) where if $\lambda_i = 0$ then $\Sigma_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}\cdot\widetilde{\mathcal{Z}}}\widetilde{y}_i = 0$, which is equivalent to $\Sigma_{\mathcal{X}\mathcal{Y}_j\cdot\mathcal{Z}}\widetilde{y}_i = 0$ for $j = 1, \ldots, p$. Thus, along the subspace, $\text{span}\{\widetilde{y}_i : \theta^i_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = 0, i = 1, \ldots, \min\{np, m\}\}$ variations of the p components cannot help predict $\mathcal{X}$ over and above the predictive ability of $\mathcal{Z}$. The space spanned by these vectors is $\mathcal{V}_h^{XY}$ in the context of predictor SGNC.

B Technical Appendix

This subsection provides some of the technical material omitted from Section 3. Define the following set of matrices

$$\mathsf{Y}_h = B_h\mathsf{X}_h + \mathsf{U}_h, \quad \mathsf{Y}_h = \begin{bmatrix} W(p+h) & \cdots & W(T) \end{bmatrix}, \quad B_h = \begin{bmatrix} \gamma(h) & \pi_1^{(h)} & \cdots & \pi_p^{(h)} \end{bmatrix}, \quad (19)$$

$$\mathsf{X}_h = \begin{bmatrix} \mathsf{X}_h(p) & \cdots & \mathsf{X}_h(T-h) \end{bmatrix}, \quad \mathsf{U}_h = \begin{bmatrix} \mathsf{U}_h(p) & \cdots & \mathsf{U}_h(T-h) \end{bmatrix}, \quad (20)$$

$$\mathsf{X}_h(t) = \begin{bmatrix} D^{(h)}(t) \\ W(t) \\ W(t-1) \\ \vdots \\ W(t-p+1) \end{bmatrix}, \quad \mathsf{U}_h(t) = \sum_{j=0}^{h-1} \psi_j a(t+h-j). \quad (21)$$

Then the OLS estimator

$$\widehat{B}_h = \mathsf{Y}_h\mathsf{X}_h'(\mathsf{X}_h\mathsf{X}_h')^{-1} \quad (22)$$

is $\sqrt{T}$–consistent under fairly general regularity conditions. Ω can be estimated consistently by

$$\widehat{\Omega} = \frac{1}{T}(\mathsf{Y}_1 - \widehat{B}_1\mathsf{X}_1)(\mathsf{Y}_1 - \widehat{B}_1\mathsf{X}_1)'. \quad (23)$$

The impulse responses are also consistently estimated by iterating (3) and (4).

Now, two points need to be kept in mind: (i) if the regression contains unbounded deterministic trends, we will need to rescale in the asymptotic analysis and (ii) the errors in the regression have an MA($h-1$) structure and so the asymptotic covariance of $\widehat{B}_h$ is not of the Kroncker product form for $h > 1$. To address (i) we will assume the existence of a diagonal rescaling matrix Q_T such that the dataset $Q_T^{-1}\mathsf{X}_h$ satisfies the usual regularity conditions. This is certainly true for polynomial trends, where each term of the form t^ν needs to

be rescaled by $T^{\nu+1}$ (Hamilton, 1994, Chapter 16). To address (ii), we write

$$\sqrt{T}\mathrm{vec}((\widehat{B}_h - B_h)Q_T) = \left(\left(\frac{Q_T^{-1}\mathsf{X}_h\mathsf{X}_h'Q_T^{-1}}{T}\right)^{-1} \otimes I_n\right) \mathrm{vec}\left(\frac{\mathsf{U}_h\mathsf{X}_h'Q_T^{-1}}{\sqrt{T}}\right) \tag{24}$$

$$= \left(\left(\frac{Q_T^{-1}\mathsf{X}_h\mathsf{X}_h'Q_T^{-1}}{T}\right)^{-1} \otimes I_n\right) \frac{1}{\sqrt{T}} \sum_{t=p}^{T-h} Q_T^{-1}\mathsf{X}_h(t) \otimes \mathsf{U}_h(t). \tag{25}$$

Since $\mathsf{U}_h(t)$ is an MA$(h-1)$ process, the summands $Q_T^{-1}\mathsf{X}_h(t) \otimes \mathsf{U}_h(t)$ are serially correlated at lags 1 through $h-1$ and, because $a(t)$ is martingale difference process, there is no serial correlation beyond that lag. Using standard results (e.g. Section 6.3 of White (2001) and Chapter 16 of Hamilton (1994)),

$$\frac{1}{\sqrt{T}} \sum_{t=p}^{T-h} Q_T^{-1}\mathsf{X}_h(t) \otimes \mathsf{U}_h(t) \xrightarrow{d} N(0, \Psi_h), \tag{26}$$

where

$$\Psi_h = \lim_{T\to\infty} \sum_{j=-h+1}^{h-1} \mathrm{cov}(Q_T^{-1}\mathsf{X}_h(t) \otimes \mathsf{U}_h(t), Q_T^{-1}\mathsf{X}_h(t-j) \otimes \mathsf{U}_h(t-j)), \tag{27}$$

and

$$\widehat{\Gamma}_h = \frac{Q_T^{-1}\mathsf{X}_h\mathsf{X}_h'Q_T^{-1}}{T} \xrightarrow{p} \Gamma_h. \tag{28}$$

Both Ψ_h and Γ_h are positive definite under the usual regularity assumptions. The asymptotic distribution of our estimator is then

$$\sqrt{T}\mathrm{vec}((\widehat{B}_h - B_h)Q_T) \xrightarrow{d} N\left(0, \Xi_h\right), \tag{29}$$

where the asymptotic covariance matrix of $\widehat{B}_h$ is given by

$$\Xi_h = (\Gamma_h^{-1} \otimes I_n)\Psi_h(\Gamma_h^{-1} \otimes I_n). \tag{30}$$

Now, as is well known, sample analogues can be substituted in for Γ_h but not for Ψ_h because the sample analogue is not guaranteed to be positive definite (Hamilton, 1994, p. 281). Following Dufour et al. (2006), we opt again for simplicity and convenience and utilize a Bartlett–Newey–West estimator of the form

$$\widehat{\Psi}_h = \sum_{j=-m(T)+1}^{m(T)-1} \left(1 - \frac{|j|}{m(T)}\right) \widehat{\mathrm{cov}}(Q_T^{-1}\mathsf{X}_h(t) \otimes \widehat{\mathsf{U}}_h(t), Q_T^{-1}\mathsf{X}_h(t-j) \otimes \widehat{\mathsf{U}}_h(t-j)), \tag{31}$$

where

$$\widehat{\mathsf{U}}_h(t) = W(t+h) - \widehat{B}_h\mathsf{X}_h(t) \tag{32}$$

and $m(T)$, commonly known as the bandwidth, satisfies $m(T) \to \infty$ and $m(T)/T^{\frac{1}{4}} \to 0$ (see Hall (2005), Cushing & McGravey (1999), and den Haan & Levin (1997)). With this estimator of Ψ_h, our estimator for the asymptotic covariance matrix of $\widehat{B}_h$ is

$$\widehat{\Xi}_h = (\widehat{\Gamma}_h^{-1} \otimes I_n)\widehat{\Psi}_h(\widehat{\Gamma}_h^{-1} \otimes I_n). \tag{33}$$

The estimator above requires the bandwidth to grow infinitely large but at a slower rate than T. A recent literature has allowed the bandwidth to behave as $m(T) = bT$ for $b \in (0, 1]$ (Kiefer et al., 2000; Kiefer & Vogelsang, 2002b,a, 2005). This fixed–bandwidth approach makes $\widehat{\Psi}_h$ inconsistent although test statistics using this estimator remain asymptotically pivotal in our context. This theory, commonly known as fixed–b theory to distinguish it from the small–b theory above, has found great success in controlling for over–rejection in small samples, a serious problem in GC testing.

SGNC amounts to restrictions on matrices which are linear transformations of B_h. In particular if $L \in \mathbb{R}^{n \times n_X}$ selects the X elements of W and $R \in \mathbb{R}^{n \times n_Y}$ selects the Y elements then

$$C^{\text{target}} = \begin{bmatrix} \pi^{(h)}_{XY1} & \cdots & \pi^{(h)}_{XYp} \end{bmatrix} = L' B_h \begin{bmatrix} 0_{k \times n_Y p} \\ I_p \otimes R \end{bmatrix}, \tag{34}$$

while

$$C^{\text{predictor}} = \begin{bmatrix} \pi^{(h)}_{XY1} \\ \vdots \\ \pi^{(h)}_{XYp} \end{bmatrix} = \sum_{i=1}^{p} (e_i \otimes L') B_h \begin{bmatrix} 0_{k \times n_Y} \\ (e_i \otimes I_n)R \end{bmatrix}, \tag{35}$$

where $e_i \in \mathbb{R}^p$ is the i-th standard basis vector. A generic expression for both C^{target} and $C^{\text{predictor}}$ is

$$C = \sum_{i=1}^{q} L_i B_h R_i, \tag{36}$$

where $\sum_{i=1}^{q} R_i' \otimes L_i$ is of full rank. In the case of C^{target}, $q = 1$, $L_1 = L'$ and $R_1 = \begin{bmatrix} 0_{k \times n_Y p} \\ I_p \otimes R \end{bmatrix}$ in (34). In the case of $C^{\text{predictor}}$, $q = p$, $L_i = (e_i \otimes L')$ and $R_i = \begin{bmatrix} 0_{k \times n_Y} \\ (e_i \otimes I_n)R \end{bmatrix}$ in (35). In each case, $\sum_{i=1}^{q} R_i' \otimes L_i$ is of full rank because the mappings $B_h \mapsto C^{\text{target}}$ and $B_h \mapsto C^{\text{predictor}}$ are surjective. Our estimator is then

$$\widehat{C} = \sum_{i=1}^{p} L_i \widehat{B}_h R_i \tag{37}$$

and its asymptotic covariance matrix is given by

$$\Theta = \left(\sum_{i=1}^{p} R_i' \otimes L_i\right)(\Gamma_h^{-1} \otimes I_n)\Psi_h(\Gamma_h^{-1} \otimes I_n)\left(\sum_{i=1}^{p} R_i \otimes L_i'\right). \tag{38}$$

It can be estimated by plugging in any of the estimators we proposed in the previous section

$$\widehat{\Theta} = \left(\sum_{i=1}^{p} R_i' \otimes L_i\right)(\widehat{\Gamma}_h^{-1} \otimes I_n)\widehat{\Psi}_h(\widehat{\Gamma}_h^{-1} \otimes I_n)\left(\sum_{i=1}^{p} R_i \otimes L_i'\right). \tag{39}$$

Finally, we mention that the restricted OLS estimator of $\overline{B}_h$ is given by

$$\operatorname{vec}(\overline{B}_h) = \left(I_{n(np+k)} - \{(\mathsf{X}_h\mathsf{X}_h')^{-1} \otimes I_n\}D_h^{r\prime}\{D_h^r((\mathsf{X}_h\mathsf{X}_h')^{-1} \otimes I_n)D_h^{r\prime}\}^{-1}D_h^r\right)\operatorname{vec}(\widehat{B}_h), \tag{40}$$

where $D_h^r = \sum_{i=1}^{q}(R_i' \otimes \widehat{N}_r' L_i) = (I_{n_Y p} \otimes \widehat{N}_r)' \sum_{i=1}^{q}(R_i' \otimes L_i)$ when testing for target SGNC and $D_h^r = (\widehat{M}_r \otimes I_{n_X p})' \sum_{i=1}^{q}(R_i' \otimes L_i)$ when testing for predictor SGNC. Note that this restricted estimator does not depend on the particular identification of $\widehat{N}_r$ and $\widehat{M}_r$.

Acknowledgements

I am grateful to two anonymous referees for their insightful comments and helpful suggestions. I would also like to thank Lynda Khalaf, Sean Holly, Hashem Pesaran, George Kapetanios, Robert Engle, Oscar Jorda, Jesus Gonzalo, Geert Mesters, Barbara Rossi, Tatevik Sekhposyan, Marek Jarocinski, and seminar participants at the European Central Bank. All remaining errors are my own. Some of the results of this paper formed part of my Phd thesis at the University of Cambridge. Research for this paper was supported by Spanish Ministry of Economy and Competitiveness projects ECO2012-33247 and ECO2015-68136-P (MINECO/FEDER, UE) and Fundación BBVA Scientific Research Grant PR16-DAT-0043.

References

Al-Sadoon, M. M. (2014). Geometric and long run aspects of granger causality. *Journal of Econometrics, 178, Part 3*(0), 558 – 568.

Al-Sadoon, M. M. (2015). A general theory of rank testing. Working Papers 750, Barcelona Graduate School of Economics.

Al-Sadoon, M. M. (2017a). The linear systems approach to linear rational expectations models. *Forthcoming, Econometric Theory.*

Al-Sadoon, M. M. (2017b). A unifying theory of tests of rank. *Journal of Econometrics*, *199*(1), 49 – 62.

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

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis, 3rd Edition*. Hoboken, NJ: John Wiley and Sons Inc.

Bhansali, R. J. (2002). Multi-step forecasting. In M. P. Clements & D. F. Hendry (Eds.), *A Companion to Economic Forecasting*, volume 1 (pp. 207–221). Blackwell Publishing Ltd.

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.

Camba-Mendez, G., Kapetanios, G., Smith, R. J., & Weale, M. R. (2003). Tests of rank in reduced rank regression models. *Journal of Business & Economic Statistics*, *21*(1), 145–155.

Cragg, J. G. & Donald, S. G. (1996). On the asymptotic properties of ldu-based tests of the rank of a matrix. *Journal of the American Statistical Association*, *91*(435), 1301–1309.

Cragg, J. G. & Donald, S. G. (1997). Inferring the rank of a matrix. *Journal of Econometrics*, *76*(1-2), 223–250.

Cushing, M. J. & McGravey, M. G. (1999). Covariance matrix estimation. In L. Maátyás (Ed.), *Generalized Method of Moments Estimation*, volume 1 (pp. 63–95). Cambridge University Press.

den Haan, W. J. & Levin, A. T. (1997). A practitioner's guide to robust covariance matrix estimation. In G. S. Maddala & C. R. Rao (Eds.), *Handbook of Statistics*, volume 15 (pp. 299–342). Elsevier Science B. V.

Dolado, J. & Lütkepohl, H. (1996). Making wald tests work for cointegrated var systems. *Econometric Reviews*, *15*(4), 369–386.

Dufour, J.-M. (2006). Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics*, *133*(2), 443 – 477.

Dufour, J.-M. & Khalaf, L. (2002). Simulation based finite and large sample tests in multivariate regressions. *Journal of Econometrics*, *111*(2), 303 – 322.

Dufour, J.-M. & Khalaf, L. (2003). Monte carlo test methods in econometrics. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics* chapter 23, (pp. 494–519). Blackwell

Publishing Ltd.

Dufour, J.-M., Pelletier, D., & Renault, E. (2006). Short run and long run causality in time series: inference. *Journal of Econometrics*, *127*(2), 337–362.

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.

Dufour, J.-M. & Valery, P. (2016). Hypothesis tests when rank conditions fail: a smooth regularization approach. mimeo, HEC Montréal. Available at http://www.cireqmontreal.com/wp-content/uploads/2016/05/valery.pdf.

Dufour, J.-M. & Zhang, H. J. (2015). Short and long run second-order causality: Theory, measures and inference. Mimeo.

Duplinskiy, A. (2014). Is regularization necessary? a wald-type test under non-regular conditions. Research Memorandum 025, Maastricht University, Graduate School of Business and Economics (GSBE).

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

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

Garratt, T., Lee, K., Pesaran, H., & Shin, Y. (2006). *Global and National Macroeconometric Modelling: A Long Run Structural Approach*. Oxford, UK: Oxford University Press.

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.

Golub, G. H. & Van Loan, C. F. (1996). *Matrix Computations* (3 ed.). Classics in Applied Mathematics. Baltimore, USA: Johns Hopkins University Press.

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

Hall, A. R. (2005). *Generalized Method of Moments*. Advanced Texts in Econometrics. Oxford, UK: Oxford University Press.

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

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

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

Jarociński, M. & Maćkowiak, B. (2017). Granger causal priority and choice of variables in vector autoregressions. *The Review of Economics and Statistics*, *99*(2), 319–329.

Jorda, O. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, *95*(1), 161–182.

Juselius, K. (2006). *The Cointegrated VAR Model*. Advanced Texts in Econometrics. Oxford, UK: Oxford University Press.

Kiefer, N. M. & Vogelsang, T. J. (2002a). Heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation. *Econometrica*, *70*(5), 2093–2095.

Kiefer, N. M. & Vogelsang, T. J. (2002b). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory*, *18*(06), 1350–1366.

Kiefer, N. M. & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, *21*(06), 1130–1164.

Kiefer, N. M., Vogelsang, T. J., & Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, *68*(3), 695–714.

Kleibergen, F. & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97–126.

Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, *74*(1), 119–147.

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

Lutkepohl, H. & Burda, M. M. (1997). Modified wald tests under nonregular conditions. *Journal of Econometrics*, *78*(2), 315–332.

McQuarrie, A. D. R. & Tsai, C.-L. (1998). *Regression and Time Series Model Selection*. London, UK: World Scientific Co. Pte. Ltd.

Newey, W. K. & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, *28*(3), 777–87.

Okun, A. M. (1962). Potential gnp: Its measurement and significance. In *Proceedings of the Business and Economics Statistics Section of the American Statistical Association* (pp.

98–103). Washington, DC: American Statistical Association.

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

Reinsel, G. C. (2003). *Elements of Multivariate Time Series Analysis* (2 ed.). Springer Series in Statistics. New York, USA: Springer.

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

Robin, J.-M. & Smith, R. J. (2000). Tests of rank. *Econometric Theory*, *16*(02), 151–175.

Romer, C. D. & Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, *94*(4), 1055–1084.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics*, *8*(1), 147–164.

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

Song, X. & Taamouti, A. (2017). Measuring nonlinear granger causality in mean. *Forthcoming, Journal of Business & Economic Statistics*.

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

Toda, H. Y. & Phillips, P. C. B. (1993). Vector autoregressions and causality. *Econometrica*, *61*(6), 1367–1393.

Toda, H. Y. & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics*, *66*(1-2), 225–250.

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

White, H. (2001). *Asymptotic Theory for Econometricians: Revised Edition*. Economic Theory, Econometrics, and Mathematical Economics. Bingley, UK: Emerald Group Publishing.

White, H., Chalak, K., & Lu, X. (2011). Linking granger causality and the pearl causal model with settable systems. *Journal of Machine Learning Research Workshop and Conference Proceedings*, (12), 1–29.

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

White, H. & Pettenuzzo, D. (2014). Granger causality, exogeneity, cointegration, and economic policy analysis. *Journal of Econometrics*, *178*(P2), 316–330.

Wiener, N. (1956). The theory of prediction. In E. F. Beckenback (Ed.), *Modern Mathematics*

for Engineers, 1 chapter 8.

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