Causality Along Subspaces: Inference*

Majid M. Al-Sadoon†

April 26, 2010

Abstract

This paper extends the analysis of Dufour et al. (2006) in the direction of the more general notions of causality introduced by Al-Sadoon (2009). In particular, we propose new tests for Granger non–causality along subspaces in VAR processes at various horizons. The methodology is illustrated by reexamining the Bernanke & Mihov (1998) data set for forecast horizons of 1–24 months. We find that monetary policy predicts output growth and inflation only along a single direction in output–growth–inflation space for most of the forecast horizons we consider. We also find that the variations of non–borrowed reserves and the federal funds rate along certain directions have no predictive power for output growth and inflation at most of the horizons considered.

JEL Classification: C12, C13, C32.

Keywords: Granger causality, subspace causality, vector autoregressions, reduced rank restrictions, Bootstrap.

*I would like to thank Lynda Khalaf and Sean Holly for helpful comments and suggestions. All remaining errors are my own.

†Contact details are available at: www.econ.cam.ac.uk/phd/mma48/.

1 Introduction

The concept of causality developed by Wiener (1956) and Granger (1969) is one of the most fundamental concepts of time series analysis (see e.g. the surveys of Geweke (1984) or Hamilton (1994)). The basic idea of Granger causality is that the cause should occur before the effect and so the cause should help predict the effect. The original idea confined "prediction" to one–step–ahead prediction; however, when Granger causality was extended to multivariate testing by Tjøstheim (1981) and Hsiao (1982) the question arose as to the possibility of indirect causality between variables. Dufour & Renault (1998) then introduced h–step causality, which was capable of describing both direct and indirect causality in a multivariate setting.

Recently, Al-Sadoon (2009) has shown that these cartesian notions of causality may not give a full characterization of the causal structure of the system. In particular, if a vector process Y causes the vector process X at horizon h, it may still be the case that the causal linkage is weak as either: (i) the effect of Y on X may be confined to a subspace of X–space or (ii) variations of Y in certain direction may have no effect on X. Thus, we must consider the subspaces along which causal linkages may lie. Al-Sadoon (2009) finds that the restrictions required for testing subspace causality are rank restrictions rather than the zero restrictions used for cartesian causality testing. This accords with T. W. Anderson's original contribution (Anderson, 1951) that the proper extension of zero univariate restrictions to the multivariate setting is rank restrictions rather than zero restrictions.

In this paper we present an estimation and inference procedure for testing subspace causality in finite order VAR models. We employ the method of (p, h) autoregressions 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 test statistic of Kleibergen & Paap (2006), which is a Wald test, and the rank estimation procedure is that proposed by Robin & Smith (2000), which tests sequentially increasing rank. The methodology is also extended to the $I(1)$ case by employing the results of Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996), of augmenting the regression equation by a redundant lag to achieve standard asymptotics.

The methodology above is applied to the Bernanke & Mihov (1998) data set, which was studied by Dufour et al. (2006). In particular, we look at the predictive effect of monetary policy (consisting of non–borrowed reserves and the federal funds rate) on output growth and inflation at horizons 1–24 months. We find that the federal funds rate predicts output growth and inflation only along a single very flat line in output–growth–inflation space for most of the forecast horizons we consider; therefore the predictive power of the federal funds rate is primarily for output growth rather than inflation. We also find that the variations of non–borrowed reserves and the federal funds rate along certain directions have no predictive power for output growth and inflation at most of the horizons considered; the slopes of these directions are consistently negative, indicating a tradeoff between the federal funds rate and non–borrowed reserves *viz–a–viz* output growth and inflation, thus one can interpret this result as a confirmation of the idea that the central bank may control either the money supply or interest rates but not both at the same time.

We should emphasize that we have deliberately kept the scope of this paper rather limited to extending the analysis of Dufour et al. (2006) in order to demonstrate the properties of subspace as opposed to cartesian causality. Our testing procedure and accompanying Matlab code is able to reproduce all of the results of Dufour et al. (2006). However, we do propose a list of further venues of research in the conclusion to this paper.

The paper is organized as follows. Section 2 motivates and reviews the idea of subspace causality. Section 3 discusses estimation and inference. Section 4 is an empirical illustration of the methodology. Section 5 concludes and section 6 is an appendix.

2 Causality in VAR Models

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

$$W(t) = \mu + \sum_{j=1}^{p} \pi_j W(t-j) + a(t), \quad t = p, \ldots, T \tag{2.1}$$

where $a(t)$ is a martingale difference sequence with respect to the information set generated by W, with $\mathbb{E}(a(t)a'(t)) = \Omega > 0$. The first p observations of W are assumed given. Although we considers only a constant deterministic term, our results apply more generally to any deterministic trend.

We will be interested in measuring predictability of the various components of W at various horizons based on its history. Iterating (2.1) over h periods we get,

$$W(t+h) = \mu_h + \sum_{j=1}^{p} \pi_j^{(h)} W(t+1-j) + \sum_{j=0}^{h-1} \psi_j a(t+h-j), \quad t = p-1, \ldots, T-h, \tag{2.2}$$

where $\mu_h = \sum_{j=0}^{h-1} \psi_j \mu$ and the coefficients matrices, derived by Dufour & Renault (1995), are given by,

$$\pi_j^{(1)} = \pi_j, \qquad \pi_j^{(h+1)} = \pi_{j+h} + \sum_{l=1}^{h} \pi_{h-l+1} \pi_j^{(l)} = \pi_{j+1}^{(h)} + \pi_1^{(h)} \pi_j, \qquad j, h \geq 1 \tag{2.3}$$

$$\psi_h = \pi_1^{(h)}, \qquad h \geq 1 \tag{2.4}$$

Therefore, the best linear predictor of $W(t+h)$ based on current and past values of W is given by,

$$P(W(t+h)|W[0,t]) = \mu_h + \sum_{j=1}^{p} \pi_j^{(h)} W(t+1-j), \quad t = p-1, \ldots, T-h, \tag{2.5}$$

We will require W to be partitioned 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. The coefficient matrices are then partitioned conformably with W as,

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

Dufour & Renault (1998) define h–step non–causality as follows: Y fails to cause X at a given horizon h if at every time t the forecast of $X(t+h)$ does not depend on current or past Y. We will denote this by $Y \nrightarrow_h X$. To see what the implications of non–causality are for the VAR(p) model, consider the X equations of the system (2.2),

$$X(t+h) = \mu_{Xh} + \sum_{j=1}^{p} \pi_{XYj}^{(h)} Y(t-j+1) + \pi_{XXj}^{(h)} X(t-j+1) + \pi_{XZj}^{(h)} Z(t-j+1) + \mathcal{X}(t+h), \quad t \geq p-1 \tag{2.6}$$

where $\mathcal{X}(t+h)$ is an MA$(h-1)$ of the innovations of W. Now the best linear predictor of $\sum_{j=1}^{p} \pi_{XXj}^{(h)} X(t-j+1) + \pi_{XZj}^{(h)} Z(t-j+1)$ is itself whether we exclude Y from the information set or not; on the other hand, the best linear predictor of $\mathcal{X}(t+h)$ is zero as the innovations of W are martingale differences. Thus, whether or not $Y \nrightarrow_h X$ depends on how the best linear predictor of $\sum_{j=1}^{p} \pi_{XYj}^{(h)} Y(t-j+1)$ varies as we exclude the history of Y.

Result 2.1 (Theorem 3.1 of Dufour & Renault (1998))**.** $Y \nrightarrow_h X$ if and only if, $\pi^{(h)}_{XYj} = 0$ for all $1 \leq j \leq p$.

Al-Sadoon (2009) has shown that the Dufour & Renault (1998) framework does not capture the full linear structure of dependence in multivariate time series; in particular if we fail to reject non–causality, it may still be the case that the causal linkages are weak and occur only along certain subspaces of the variations in X and Y. This is illustrated in the following example.

Example 2.1. Consider US monthly data on real GDP growth, GDP, inflation, P, the growth of non–borrowed reserves, NBR, and the percentage change in the federal funds rate, r for the period January 1965 to December 1996 (more on this data set in section 4). We would like to visualize the predictive effect of monetary policy $Y = (NBR, r)$ on $X = (GDP, P)$.[1] Now the forecast of X can include current and lagged Y's as predictors or it may not; the significance of the difference between the two forecasts is indicative of Granger causality as we have explained above. We constructed our forecasts using equation (2.6) with and without Y using a lag length of 16 and plotted the difference in the following graphs (more on this choice of estimation method in sections 3 and 4).

It is clear from Figure 2.1 that monetary policy predicts primarily output growth. However, it would be a mistake to conclude that monetary policy has little predictive value for inflation. There seems to be a tilt in the predictable variations of output and inflation at horizons 16–19, with monetary policy predicting output and inflation along a slightly positively sloped line in (GDP, P) space.

We can now ask, where is this predictive power coming from, non–borrowed reserves or the federal funds rate? To answer this question, we repeated the same exercise but this time looking at the effect of excluding only NBR in Figure 2.2 and then the effect of excluding only r in Figure 2.3. We see that non–borrowed reserves has less predictive power than the federal funds rate. We also continue to see the pattern of a slightly upward sloping line in (GDP, P) space along which monetary policy has predictive power.

In the above discussion we have considered the directions in X space along which Y has predictive power. We may ask alternatively, are there any directions in Y space along which there is no predictive power. Suppose that in place of the given monetary policy

[1]See Bernanke & Mihov (1998) and Christiano et al. (1999) for more on this choice of instruments.

Figure 2.1: Difference in Forecasts with and without Monetary Policy as Predictors.

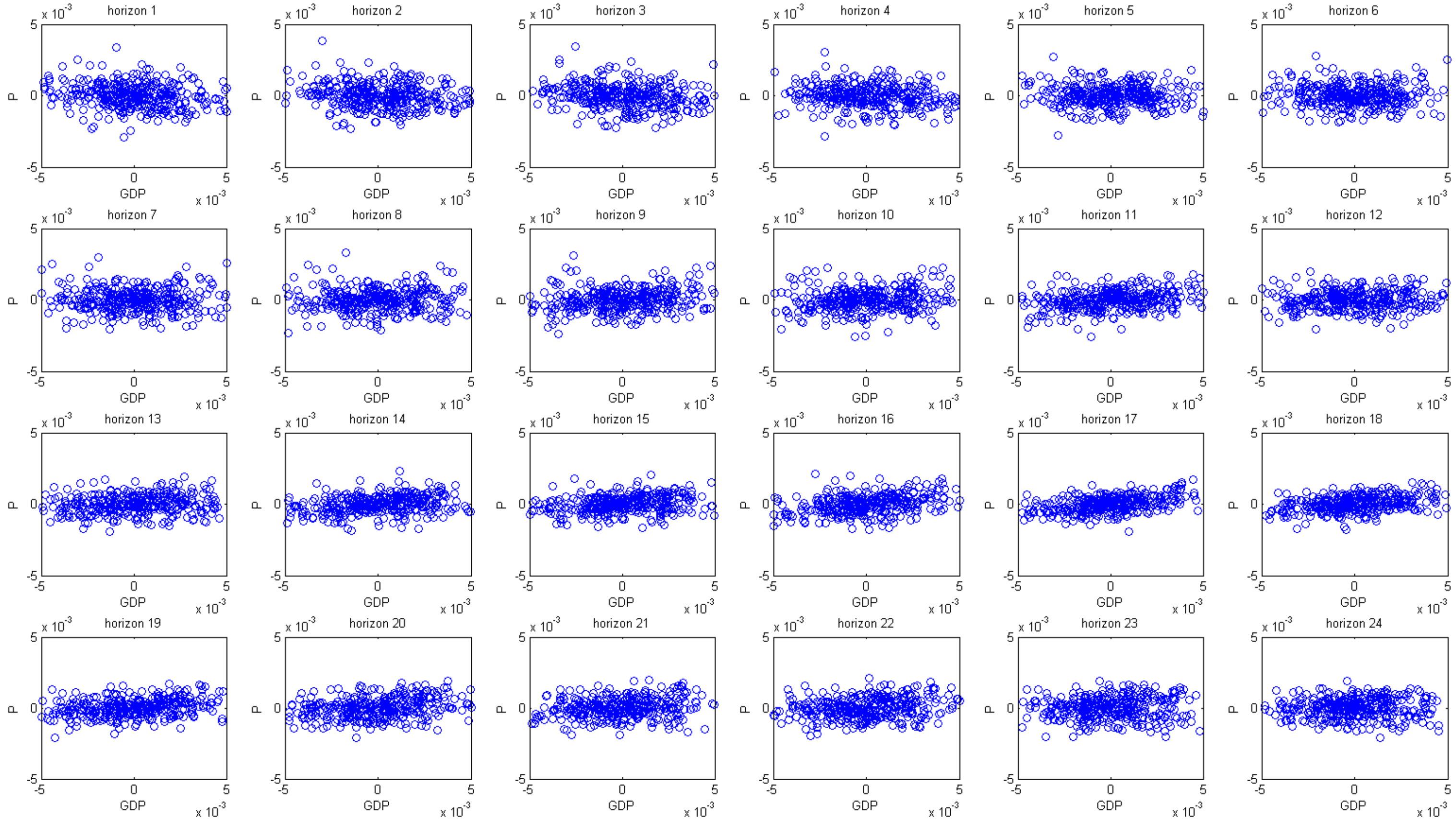


Figure 2.2: Difference in Forecasts with and without Non–Borrowed Reserves as a Predictor.

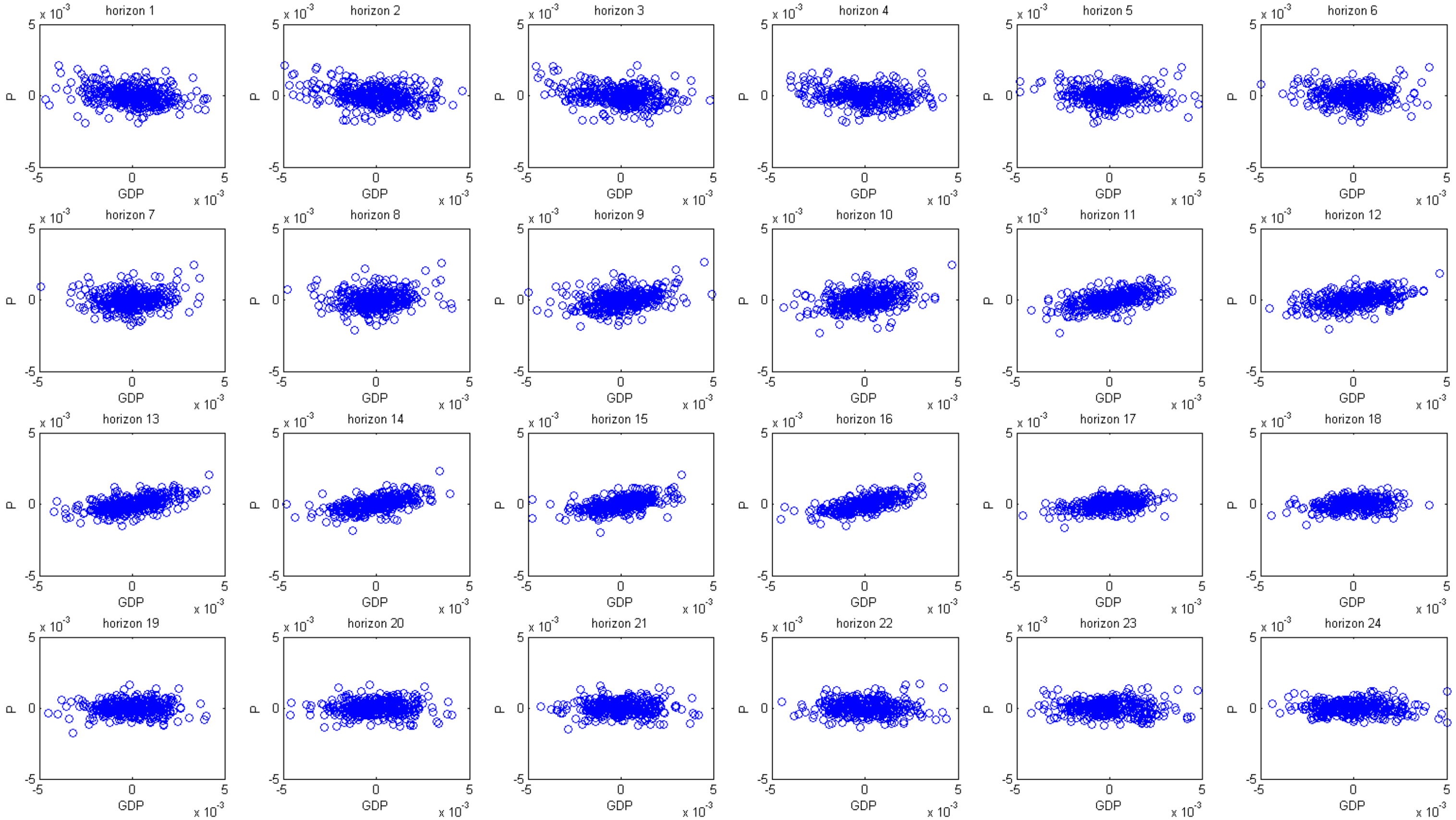
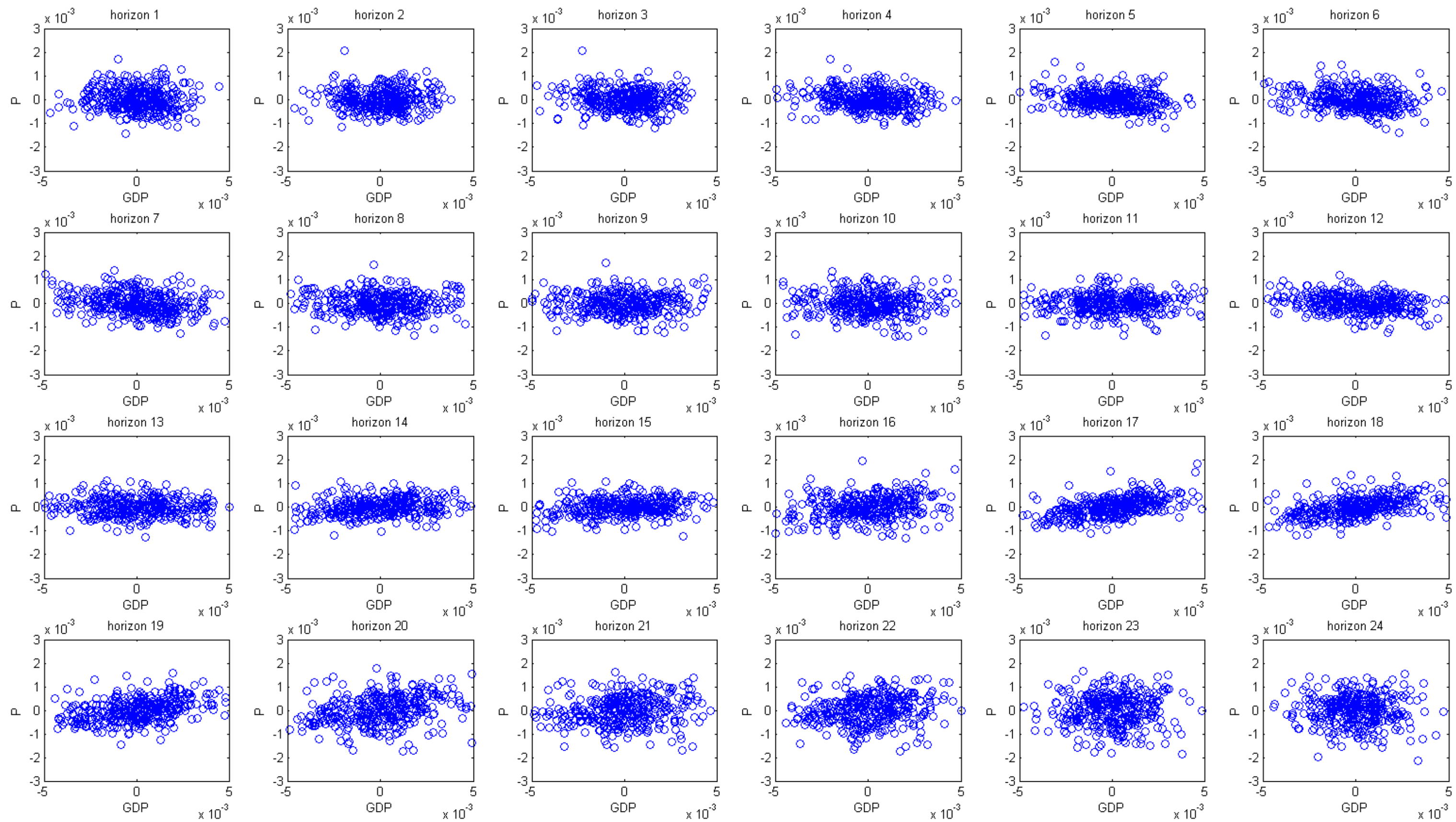


Figure 2.3: Difference in Forecasts with and without the Federal Funds Rate as a Predictor.



instruments, we were to construct two different instruments, $I_1(\theta) = \cos(\theta)NBR + \sin(\theta)r$ and $I_2(\theta) = -\sin(\theta)NBR + \cos(\theta)r$. Such a transformation amounts to rotating (NBR, r) space by θ degrees. We would like to see in which directions of variation, monetary policy has its greatest and least predictive power. In order to do that, we can invert the Wald test for the exclusion of $I_1(\theta)$ from the X equations for the range of $\theta \in [-\pi/2, \pi/2]$. This is illustrated in Figure 2.4, where the horizontal lines are the critical values of the test at 5% and 10% (we continued to use 16 lags).

It is clear from Figure 2.4 that variations of monetary policy in certain directions have statistically insignificant effects on output growth and inflation. In particular, at horizon 4 there is a range of positively sloped lines in (NBR, r) space along which monetary policy has no predictive value for output and inflation. This range exists for every horizon after 4.

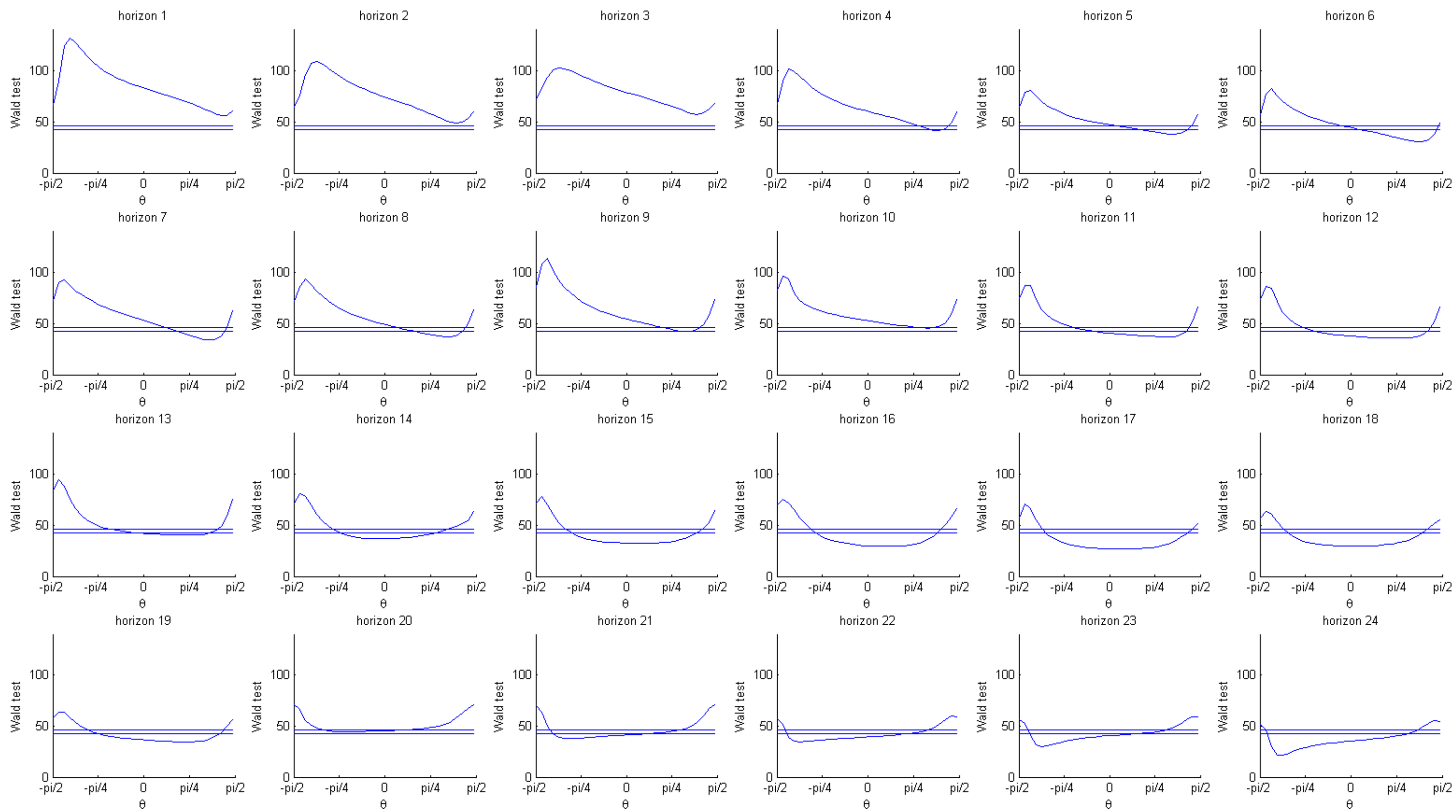
We may now ask, is this loss of predictive power along certain directions due to loss of predictive power for output growth or for inflation? We can see from Figure 2.5 and 2.6 that the picture is qualitatively similar although monetary policy fails to predict inflation very well at horizons higher than 3, which is a conclusion we have already arrived at from looking at Figure 2.1.

Of course the critical values used in the above test are based on asymptotic theory, which as is well known may over–reject in small samples (see e.g. Dufour & Khalaf (2002) and the references therein). Thus the regions of non–rejection found above are potentially much larger; a fact that we would have seen if we had bootstrapped the test at each $\theta \in [-\pi/2, \pi/2]$ rather than having relied on asymptotic theory.

It is clear from this example that neither the univariate analysis of causality of Granger (1969) and Dufour et al. (2006) nor the block–zero multivariate causality restrictions of Tjøstheim (1981) would have captured much of the structure we have uncovered. The univariate analysis would have failed to note the subspaces along which causality resides because it focuses on a small set of variations in the data rather than the full set of variations. On the other hand, the block–zero restriction tests would have indicated causality where the causal linkage is quite limited and confined to a subspace. In summary, univariate tests are too specific, while multivariate tests are too blunt. □

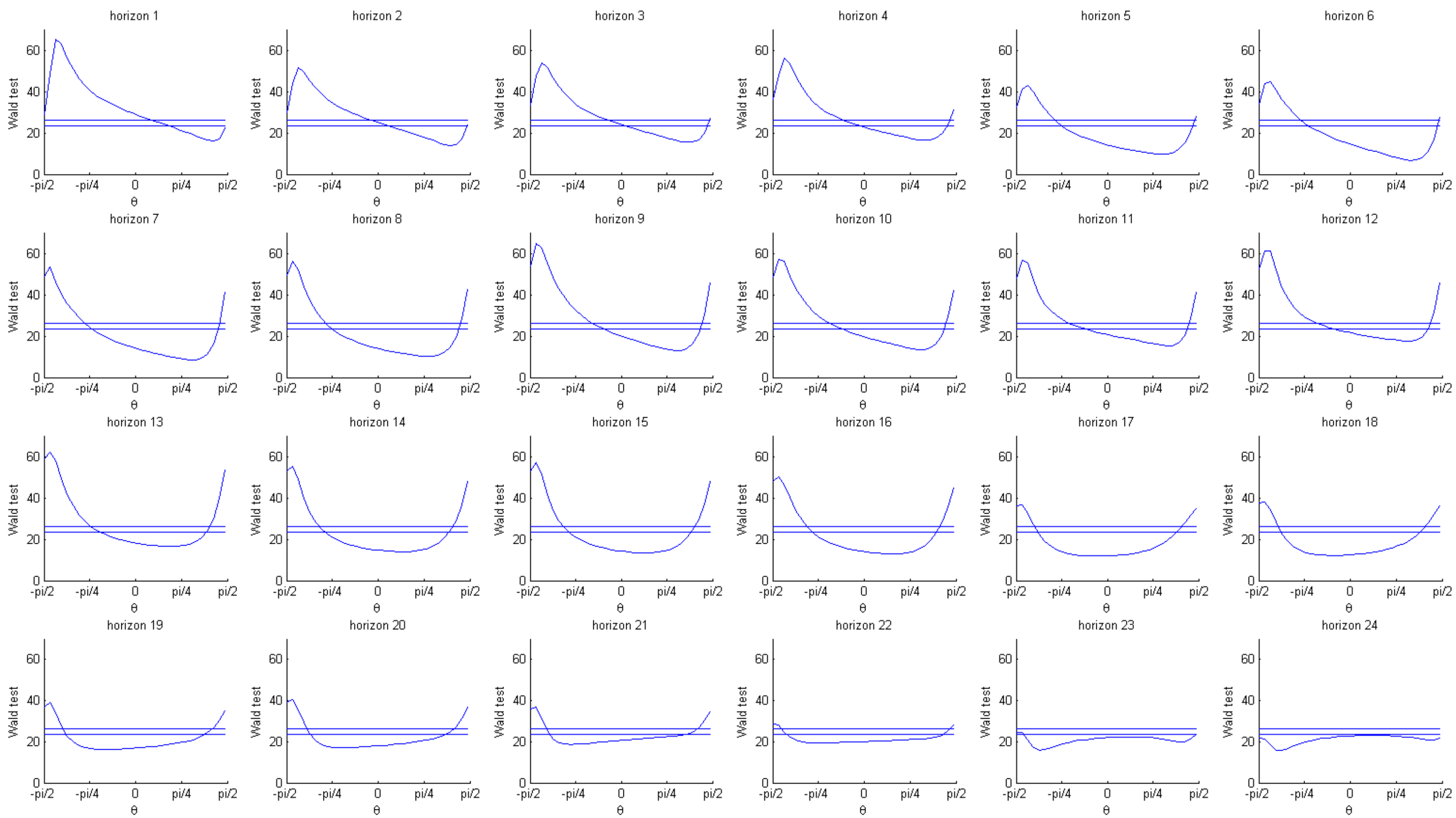
How do we capture the structure illustrated in the above example? One solution would be to conduct a principle components analysis on the difference of residuals in Figures 2.1,

Figure 2.4: Wald Test for the Exclusion of $I_1(\theta)$ from the Output Growth and Inflation Equations.[a]



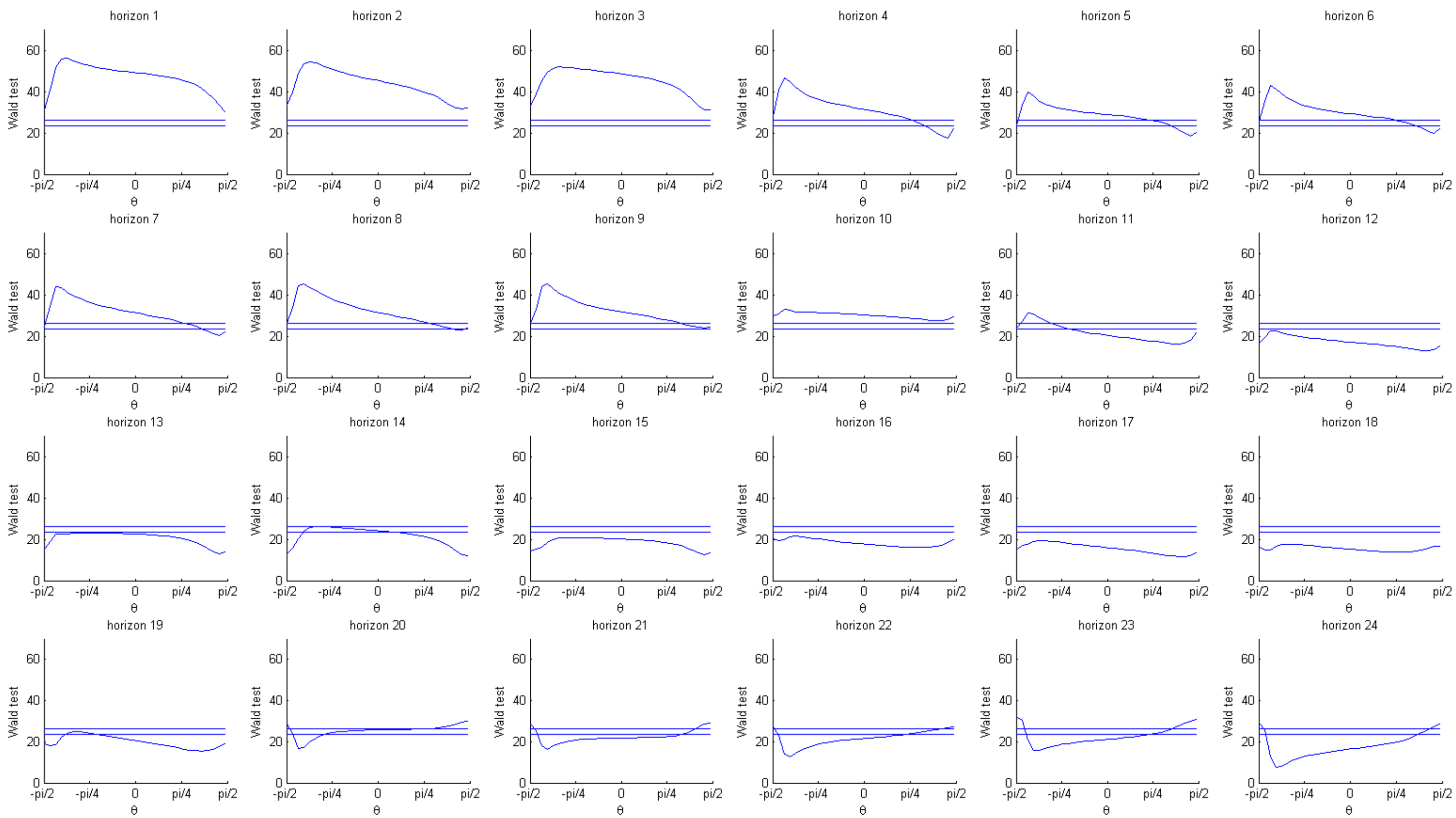
[a]Horizontal lines denote the 10% and 5% critical values of the Wald test.

Figure 2.5: Wald Test for the Exclusion of $I_1(\theta)$ from the Output Growth Equation.[a]



[a]Horizontal lines denote the 10% and 5% critical values of the Wald test.

Figure 2.6: Wald Test for the Exclusion of $I_1(\theta)$ from the Inflation Equation.[a]



[a]Horizontal lines denote the 10% and 5% critical values of the Wald test.

2.2, and 2.3 but that would isolate the main directions of variation without indicating which ones were insignificant and even then it would be inefficient in ignoring the fact that the data was generated from a regression model. Another solution would be to conduct a canonical correlations analysis similar to that presented in Appendix 6.1 in the context of a regression model but this has the disadvantage of being somewhat messy (Otter (1990) follows this approach in the context of forecasting). Yet a third route would invert the Wald tests of non–causality as we did Figures 2.4, 2.5, and 2.6; but this gives us set estimates rather than point estimates; of course, we could minimize the Wald statistic to obtain a point estimate but except for very special cases (i.e. when the estimator has a variance of Kronecker product form) it is not known what the distribution of this statistic is.

Al-Sadoon (2009) proposes a new notion of non–causality called subspace non–causality; we say that Y along subspace $\mathcal{V} \subseteq \mathbb{R}^{n_Y}$ fails to cause X along subspace $\mathcal{U} \subseteq \mathbb{R}^{n_X}$ at horizon $h \geq 1$ if eliminating the history of variations of Y along $\mathcal{V}$ from the information set does not change the h-step forecast of X in the direction of $\mathcal{U}$. We denote this by $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}$ and note that it is equivalent to $P_{\mathcal{V}}Y \nrightarrow_h P_{\mathcal{U}}X$, where $P_{\mathcal{U}}$ and $P_{\mathcal{V}}$ are the orthogonal projection matrices onto $\mathcal{U}$ and $\mathcal{V}$ respectively. The requisite restrictions for this sort of non–causality are as follows.

Result 2.2 (Theorem 4.1 of Al-Sadoon (2009))**.** $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}$ if and only if, $P_{\mathcal{U}}\pi^{(h)}_{XYj}P_{\mathcal{V}} = 0$ for all $1 \leq j \leq p$.

Now if $\mathcal{U}$ and $\mathcal{V}$ are known then testing for subspace non–causality is easily done by employing a Wald test as in Dufour et al. (2006). Typically, however, we will not know a priori along which subspaces non–causality occurs; thus restrictions of the form $P_{\mathcal{U}}\pi^{(h)}_{XYj}P_{\mathcal{V}} = 0$ for all $1 \leq j \leq p$ are actually rank restrictions, which brings us back to the original insight of Anderson (1951), who first proposed these tests: the appropriate extension of zero restrictions in univariate regressions is not zero block restrictions but reduced rank restrictions.

In this paper, we will not consider the full range of possible rank restrictions (the reader may consult Reinsel & Velu (1998) for a comprehensive review); we will only consider the cases where either $\mathcal{U}$ or $\mathcal{V}$ is fixed to span the full space of variation. Now define the maximal subspace $\mathcal{U}$ such that $Y \nrightarrow_h X|_{\mathcal{U}}$ to be $\mathcal{U}_h^{XY}$ and let U_h^{XY} to be a matrix of orthonormal columns which span $\mathcal{U}_h^{XY}$; likewise, the maximal subspace $\mathcal{V}$ such that $Y|_{\mathcal{V}} \nrightarrow_h X$ is denoted by $\mathcal{V}_h^{XY}$ and V_h^{XY} is the matrix of orthonormal columns which span $\mathcal{V}_h^{XY}$; similarly define

$\mathcal{U}_h^{WW}$, U_h^{WW}, $\mathcal{V}_h^{WW}$ and V_h^{WW}.[2] Then we can find U_h^{XY} by finding the appropriate U that satisfies,

$$U' \begin{bmatrix} \pi_{XY1}^{(h)} & \cdots & \pi_{XYp}^{(h)} \end{bmatrix} = 0 \tag{2.7}$$

Likewise, V_h^{XY} is the appropriate V that satisfies,

$$\begin{bmatrix} \pi_{XY1}^{(h)} \\ \vdots \\ \pi_{XYp}^{(h)} \end{bmatrix} V = 0 \tag{2.8}$$

3 Estimation and Inference

3.1 Estimation of Stationary VARs

When (2.1) is stationary, the parameters of the model can be estimated straightforwardly by OLS, from which we can construct the h-period coefficients using (2.3). Unfortunately, it is difficult to use standard asymptotics to test hypotheses on these estimates as they may have non-singular covariance matrices (see the example in section 3.6.4 of Lütkepohl (2006)). Lutkepohl & Burda (1997) have suggested a simulation technique for tackling this issue and (Lütkepohl, 2006, p. 108) has also suggested imposing the zero restrictions on the coefficients directly. Yet another route could employ the theory of regularized hypothesis tests *à la* Moore (1977), Andrews (1987), and Dufour & Valery (2009).

In this paper we will opt for simplicity, employing the technique of Dufour et al. (2006) which regresses $W(t+h)$ on current and up to $p-1$ lagged values of W to obtain the coefficient

[2] Lemma 3.3 of Al-Sadoon (2009) proves that these subspaces are unique.

matrices $\pi_1^{(h)}, \ldots, \pi_p^{(h)}$.[3] To see that this is possible, rewrite (2.2) as,

$$\mathsf{Y}_h = B_h \mathsf{X}_h + \mathsf{U}_h, \quad \mathsf{Y}_h = \begin{bmatrix} W(p+h-1) & \cdots & W(T) \end{bmatrix}, \quad B_h = \begin{bmatrix} \mu_h & \pi_1^{(h)} & \cdots & \pi_p^{(h)} \end{bmatrix}$$

$$\mathsf{X}_h = \begin{bmatrix} \mathsf{X}_h(p-1) & \mathsf{X}_h(p) & \cdots & \mathsf{X}_h(T) \end{bmatrix}, \quad \mathsf{U}_h = \begin{bmatrix} \mathsf{U}_h(p-1) & \mathsf{U}_h(p) & \cdots & \mathsf{U}_h(T) \end{bmatrix},$$

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

Then the OLS estimator of B_h is $\widetilde{B}_h = \mathsf{Y}_h \mathsf{X}_h' (\mathsf{X}_h \mathsf{X}_h')^{-1}$. This estimator is $\sqrt{T}$–consistent under fairly general regularity conditions (see p. 72 of Lütkepohl (2006) or section 6.2 of White (2001)). However, care must be exercised in testing hypotheses on $\widetilde{B}_h$ because the errors in the regression are not uncorrelated but have an MA$(h-1)$ structure. Therefore, its asymptotic covariance is not of the Kroncker product form and thus standard test statistics on $\widetilde{B}_h$ do not apply. To see this more clearly write,

$$\begin{aligned} \sqrt{T}(\mathrm{vec}(\widehat{B}_h) - \mathrm{vec}(B_h)) &= \left(\left(\frac{\mathsf{X}_h \mathsf{X}_h'}{T} \right)^{-1} \otimes I_n \right) \mathrm{vec} \left(\frac{\mathsf{U}_h \mathsf{X}_h'}{\sqrt{T}} \right) \\ &= \left(\left(\frac{\mathsf{X}_h \mathsf{X}_h'}{T} \right)^{-1} \otimes I_n \right) \frac{1}{\sqrt{T}} \sum_{t=p-1}^{T} \mathsf{X}_h(t) \otimes \mathsf{U}_h(t) \end{aligned}$$

Now since $\mathsf{U}_h(t)$ is an MA$(h-1)$ process, the summands $\mathsf{X}_h(t) \otimes \mathsf{U}_h(t)$ 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 (White, 2001, section 6.3) we have that,

$$\frac{1}{\sqrt{T}} \sum_{t=p-1}^{T} \mathsf{X}_h(t) \otimes \mathsf{U}_h(t) \to N(0, \Psi_h),$$

where,

$$\Psi_h = \sum_{j=-h+1}^{h-1} \mathrm{cov}(\mathsf{X}_h(t) \otimes \mathsf{U}_h(t), \mathsf{X}_h(t-j) \otimes \mathsf{U}_h(t-j))$$

Dufour et al. (2006) show that this matrix is invertible under the assumptions we have made. The asymptotic distribution of our estimator is then,

$$\sqrt{T}(\mathrm{vec}(\widetilde{B}_h) - \mathrm{vec}(B_h)) \to N\left(0, (\mathbb{E}(\mathsf{X}_h(t)\mathsf{X}_h'(t))^{-1} \otimes I_n)\Psi_h(\mathbb{E}(\mathsf{X}_h(t)\mathsf{X}_h'(t))^{-1} \otimes I_n)\right)$$

[3]This regression model is also used in the direct method for forecasting VARs (see e.g. Bhansali (2002))

The asymptotic variance of $\widetilde{B}_h$ is invertible because Ψ_h is invertible.

Now, as is well known from the HAC literature, sample analogues can be substituted in for $\mathbb{E}(\mathsf{X}_h(t)\mathsf{X}'_h(t))$ but not for Ψ_h because the sample analogue is not guaranteed to be positive definite. We may, however, use a Bartlett–Newey–West estimator of the form,

$$\widetilde{\Psi}_h = \sum_{j=0}^{m(T)-1} \left(1 - \frac{|j|}{m(T)}\right) \widetilde{\text{cov}}(\mathsf{X}_h(t) \otimes \widetilde{\mathsf{U}}_h(t), \mathsf{X}_h(t-j) \otimes \widetilde{\mathsf{U}}_h(t-j))$$

where,

$$\widetilde{\mathsf{U}}_h(t) = W(t+h) - \widetilde{B}_h \mathsf{X}_h(t),$$

and $m(T)/T^{\frac{1}{4}} \to 0$, where $m(T)$ is commonly known as the bandwidth of the estimator (Newey & West, 1987). Hall (2005) and Cushing & McGravey (1999) review the literature on the optimal choice of $m(T)$; they also review the literature on parametric estimators of Ψ_h (see also den Haan & Levin (1997)).[4] However, we will be concerned in this paper with the Newey–West estimator to maintain continuity with the previous literature, Dufour & Jouini (2005) and Dufour et al. (2006).

3.2 Inference in Stationary VARs

Having derived our estimator $\widetilde{B}_h$ and its asymptotic distribution, we can begin to discuss hypothesis testing. Causality tests are carried out 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,

$$\begin{bmatrix} \pi_{XY1}^{(h)} & \cdots & \pi_{XYp}^{(h)} \end{bmatrix} = L' B_h \begin{bmatrix} 0_{1 \times n_Y p} \\ I_p \otimes R \end{bmatrix}$$

and,

$$\begin{bmatrix} \pi_{XY1}^{(h)} \\ \vdots \\ \pi_{XYp}^{(h)} \end{bmatrix} = \sum_{i=1}^{p} (e_i \otimes L') B_h \begin{bmatrix} 0_{1 \times n_Y} \\ (e_i \otimes I_n) R \end{bmatrix},$$

[4] It would be very interesting to see how the Kiefer & Vogelsang (2005) analysis fairs in our setting. They set the bandwidth to T with the result that their estimate of Ψ_h is inconsistent; however, their test statistic is still asymptotically pivotal, albeit with non–standard distribution. Unfortunately, that is outside the scope of this paper and is left to future research.

where $e_i \in \mathbb{R}^p$ is the i-th standard basis vector. This form will allow us to easily write down the asymptotic distribution of the relevant coefficients in (2.7) and (2.8) as they relate to the asymptotic distribution of B_h since the general form of the submatrices we look at is $C = \sum_{i=1}^{p} L_i B_h R_i$ where L_i and R_i are left and right selection matrices. In all cases, under the assumptions we have made in this section the asymptotic covariance matrix of $\widetilde{C} = \sum_{i=1}^{p} L_i \widetilde{B}_h R_i$ is invertible because the asymptotic covariance of $\widetilde{C}$ is invertible and $\sum_{i=1}^{p} R_i' \otimes L_i$ is of full rank.[5]

Now the question is, how do we test rank restrictions of the form (2.7) and (2.8)? Various options abound, the original analysis of Anderson (1951) can be applied to the regression models (2.2) or (2.6) but because Ψ_h is not of the Kronecker form, the asymptotic distribution of Anderson's test statistic is not standard. Robin & Smith (2000) show that it is in general a weighted sum of independent $\chi^2(1)$ random variables with weights that depend on the parameters of the model; they show that when the weights are estimated consistently, the test has the correct size asymptotically. However, the potential for nuisance parameters in the asymptotic distribution makes this option less attractive than the alternative below. Cragg & Donald (1997) propose an estimator of rank that does not require Ψ_h to be of Kronecker product form and is asymptotically nuisance parameter free but it is burdensome to compute numerically. Camba-Mendez & Kapetanios (2009) surveys a number of other possible rank tests. We have opted instead to use the method of Kleibergen & Paap (2006) because it is simple to implement, it is asymptotically nuisance–parameter–free, and it is a Wald test which encompasses the test of Dufour et al. (2006). We shall now give a very simple, albeit heuristic, derivation of this test.

Suppose we would like to test the hypothesis, $H_0(q)$: $\text{rank}(C) = q$ for some $0 \leq q < \min\{m, l\}$. Suppose moreover that $\sqrt{T}(\text{vec}(\widetilde{C}) - \text{vec}(C)) \xrightarrow{d} N(0, \Theta)$, where Θ is constructed from the asymptotic variance of $\widetilde{B}_h$ and is positive definite as we have shown above. Let $C = USV'$ be a singular value decomposition; our test will make use of the sample analogue of this decomposition, suitably identified.[6] Now partition the singular value decomposition

[5]The fact that this sum of Kronecker products has full rank follows by noting that for the selection matrices we have chosen to work with, the mapping $B_h \mapsto C$ is always surjective.

[6]Interestingly, it has long been known in the numerical analysis literature that the singular values of a matrix indicate its "effective" rank (see the concept of ill–conditioning in Press et al. (1992) and the historical survey by Stewart (1993)) and yet the concept was not actually utilized in econometrics until Kleibergen & Paap (2006).

as,

$$U = \begin{bmatrix} U_{\cdot 1} & U_{\cdot 2} \end{bmatrix}, \qquad S = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}, \qquad , V = \begin{bmatrix} V_{\cdot 1} & V_{\cdot 2} \end{bmatrix},$$

so that $S_1 \in \mathbb{R}^{q \times q}$ and $C = U_{\cdot 1} S_1 V_{\cdot 1}' + U_{\cdot 2} S_2 V_{\cdot 2}'$. Then under $H_0(q)$, $U_{\cdot 2}' C V_{\cdot 2} = 0$ and so,

$$\sqrt{T}(V_{\cdot 2} \otimes U_{\cdot 2})' \text{vec}(\widetilde{C}) \xrightarrow{d} N(0, (V_{\cdot 2} \otimes U_{\cdot 2})' \Theta (V_{\cdot 2} \otimes U_{\cdot 2}))$$

The asymptotic covariance matrix $(V_{\cdot 2} \otimes U_{\cdot 2})' \Theta (V_{\cdot 2} \otimes U_{\cdot 2}) \in \mathbb{R}^{(l-q)(m-q) \times (l-q)(m-q)}$ is invertible because Θ is invertible. Therefore,

$$T \text{vec}'(\widetilde{C})(V_{\cdot 2} \otimes U_{\cdot 2}) \{(V_{\cdot 2} \otimes U_{\cdot 2})' \Theta (V_{\cdot 2} \otimes U_{\cdot 2})\}^{-1} (V_{\cdot 2} \otimes U_{\cdot 2})' \text{vec}(\widetilde{C}) \xrightarrow{d} \chi^2((m-q)(l-q))$$

Finally, replacing $V_{\cdot 2}$, $U_{\cdot 2}$, and Θ by consistent estimators achieves the same asymptotic distribution (this, of course, is not true without further justification, which we omit). The only thing to worry about then is to identify $V_{\cdot 2}$ and $U_{\cdot 2}$ in such a way that their dependence on C is continuous, what will ensure consistency, and Kleibergen & Paap (2006) provide one of many possible identification schemes. Therefore the test statistic we will utilize is,

$$\text{rk}(q) = T \text{vec}'(\widetilde{C})(\widetilde{V}_{\cdot 2} \otimes \widetilde{U}_{\cdot 2}) \{(\widetilde{V}_{\cdot 2} \otimes \widetilde{U}_{\cdot 2})' \widetilde{\Theta} (\widetilde{V}_{\cdot 2} \otimes \widetilde{U}_{\cdot 2})\}^{-1} (\widetilde{V}_{\cdot 2} \otimes \widetilde{U}_{\cdot 2})' \text{vec}(\widetilde{C}) \tag{3.1}$$

Note that under $H_0(0)$, rk(0) is simply the Wald test for testing zero block restrictions. Note also that the test statistic does not depend on the particular identification of $U_{\cdot 2}$ and $V_{\cdot 2}$.

How does this test statistic fair against the alternative $H_1(q) : \text{rank}(C) > q$? Under the alternative, $U_{\cdot 2}' C V_{\cdot 2} = S_2 \neq 0$ and following the same logic as above,

$$T\{\text{vec}(\widetilde{C})(\widetilde{V}_{\cdot 2} \otimes \widetilde{U}_{\cdot 2}) - \text{vec}(S_2)\}' \{(\widetilde{V}_{\cdot 2} \otimes \widetilde{U}_{\cdot 2})' \widetilde{\Theta} (\widetilde{V}_{\cdot 2} \otimes \widetilde{U}_{\cdot 2})\}^{-1} \{(\widetilde{V}_{\cdot 2} \otimes \widetilde{U}_{\cdot 2})' \text{vec}(\widetilde{C}) - \text{vec}(S_2)\} = O_P(1),$$

which implies that $\text{rk}(q) - T \text{vec}'(S_2)\{(V_{\cdot 2} \otimes U_{\cdot 2})' \Theta (V_{\cdot 2} \otimes U_{\cdot 2})\}^{-1} \text{vec}(S_2) = O_P(1)$ and therefore $\text{rk}(q) \xrightarrow{p} \infty$. Thus the Kleibergen & Paap (2006) test asymptotically rejects the null under the alternative hypothesis. It is easy to check that rk(q) has the usual properties of the Wald test under the local alternative $H_1 : U_{\cdot 2}' C V_{\cdot 2} = S_2 / \sqrt{T}$.

The above results suggest an algorithm for estimating the true rank of C *à la* Robin & Smith (2000). For a given size α, calculate rk(q) for $q = 0, \ldots, \min\{m, l\} - 1$ and stop at the first q for which $\text{rk}(q) > \chi^2_{1-\alpha}((l-q)(m-q))$, where $\chi^2_{1-\alpha}((l-q)(m-q))$ is the $1 - \alpha$ quantile of the $\chi^2((l-q)(m-q))$ distribution. Asymptotically, this estimate of the rank of C is never smaller than the true rank but as pointed out by Robin and Smith it may be larger because at the true rank of C the test may still reject with a positive probability of type I error. Therefore the algorithm is not weakly consistent for the true rank of C unless the true rank is equal to $\min\{m, l\}$. Robin and Smith propose that the test size used in the algorithm above should be made to depend on q and T so that $\alpha_{qT} = o(1)$ and $-T \log \alpha_{qT} = o(1)$. They prove that such a scheme yields a consistent estimator of the rank of C; it is easy to show that their result continues to hold when using the Kleibergen & Paap (2006) test with minimal modification to the original proof.

Having estimated the rank of C, the associated matrices $\widetilde{U}_{\cdot 2}$ and $\widetilde{V}_{\cdot 2}$ then span the left and right null spaces of $\widetilde{C}$ respectively and are $\sqrt{T}$–consistent for their population counterparts as proven by Kleibergen & Paap (2006). In our case, if C is as in (2.7) then $\widetilde{U}_h^{XY} = \widetilde{U}_{\cdot 2}$ while if C is as in (2.8) then $\widetilde{V}_h^{XY} = \widetilde{V}_{\cdot 2}$.

Of course it is well known that hypothesis tests based on asymptotic theory may give poor results in finite samples (see e.g. Dufour & Khalaf (2002) and Camba-Mendez et al. (2003)). Therefore, we will use bootstrap or Monte Carlo testing methods which give better size control in finite samples. It is important to note that we would have had to use these methods no matter what rank test we picked initially; the advantage of the Kleibergen & Paap (2006) test, however, is that it is asymptotically pivotal (i.e. asymptotically nuisance–parameter–free) and this imparts further efficiency to the simulation–based test (see Hall (1992) for more on the importance of asymptotic pivots). The general form of the testing algorithm follows Dufour et al. (2006),

Algorithm 3.1. For a given horizon h and size α,

(i) Compute $\widetilde{B}_1$ and $\widetilde{\Omega}$ from (2.1) by OLS.

(ii) If $h > 1$, compute $\widetilde{B}_h$ and its asymptotic covariance matrix from (2.2) or (2.6) by OLS. Also compute $\widetilde{\psi}_i$ for $i = 0, \ldots, h-1$ using (2.4).

(iii) Construct $\widetilde{C}$ and its asymptotic covariance from $\widetilde{B}_h$ and it asymptotic covariance matrix.

(iv) For $q = 0, \ldots, \min\{m, l\} - 1$ and until $\widehat{p}_N(q) > \alpha$,

   (a) Compute the rank statistics, $\mathrm{rk}(q)$ and a rank restricted $\widetilde{B}_h^q$ consistent for B under $H_0(q)$ (see the discussion below).

   (b) For $i = 1, \ldots, N$,

i. Construct a bootstrap sample of T observations using $\widetilde{B}_h^q$, $\{\widetilde{\psi}_i\}_{i=0}^{h-1}$, and $\widetilde{\Omega}$ in either equation (2.2) or (2.6) (see the discussion below).

ii. Compute (3.1) for the bootstrapped sample and denote it by $\text{rk}_i^b(q)$.

(c) Compute the bootstraped p–value, $\widehat{p}_N(q) = \frac{1+\#\{i\colon \text{rk}_i^b(q)>\text{rk}(q)\}}{N+1}$

(v) The rank of C is estimated as q and its left and right null spaces are estimated respectively as $\text{span}(U_{\cdot 2})$ and $\text{span}(V_{\cdot 2})$, where $U_{\cdot 2} \in \mathbb{R}^{m\times(m-q)}$ and $V_{\cdot 2} \in \mathbb{R}^{l\times(l-q)}$ are as constructed above.

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 $\widetilde{\Omega}$ then generating the sample from either equation (2.2) or (2.6) using $\widetilde{B}_h^q$, $\{\widetilde{\psi}_i\}_{i=0}^{h-1}$ 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 (i). The researcher may also choose to simulate more than T data point to allow "burn–in," what will ensure the data's stationarity. In our results, we have simulated $2T$ data points and discarded the first half.

The second point is that the construction of $\widetilde{B}_h^q$ can be carried out in a number of ways. We could simply replace the estimate for C by the estimate for $U_{\cdot 1}S_1V_{\cdot 1}'$ in $\widetilde{B}_h$. In our work, however, we have chosen to use the restricted OLS estimator with the restrictions $\widetilde{U}_{\cdot 2}'C = 0$ when testing for U_h^{XY} and $C\widetilde{V}_{\cdot 2} = 0$ when testing for V_h^{XY} (i.e. restricting B_h as if the right and left null spaces were known). This is written as,

$$\text{vec}(\widetilde{B}_h^q) = \Big(I_{n(np+1)} - \{(\mathsf{X}_h\mathsf{X}_h')^{-1} \otimes \widetilde{\Omega}_h\}D_h^{q\prime}\{D_h^q((\mathsf{X}_h\mathsf{X}_h')^{-1} \otimes \widetilde{\Omega}_h)D_h^{q\prime}\}^{-1}D_h^q\Big)\,\text{vec}(\widetilde{B}_h), \quad (3.2)$$

where $D_h^q = \sum_{i=1}^p (R_i' \otimes \widetilde{U}_{\cdot 2}'L_i) = (I_{n_Y p} \otimes \widetilde{U}_{\cdot 2})' \sum_{i=1}^p (R_i' \otimes L_i)$ when testing for U_h^{XY} and $D_h^q = (\widetilde{V}_{\cdot 2} \otimes I_{n_X p})' \sum_{i=1}^p (R_i' \otimes L_i)$ when testing for V_h^{XY} and $\widetilde{\Omega}_h = \frac{1}{T}\sum_{t=1}^T \widetilde{\mathsf{U}}_h(t)\widetilde{\mathsf{U}}_h'(t)$. Note that this restricted estimator also does not depend on the particular identification of $U_{\cdot 2}$ and $V_{\cdot 2}$. Whatever choice of $\widetilde{B}_h^q$, the only condition that must be satisfied is that it be consistent under $H_0(q)$ (see e.g. Davison & Hinkley (1997) or Godfrey (2009)).

3.3 Estimation and Inference for Non–Stationary VARs

Suppose now that (2.1) is allowed to be $I(1)$. In that case some components of $\widetilde{B}_h$ will be super-consistent and it will not have an invertible asymptotic covariance matrix; as a result the Kleibergen & Paap (2006) test, which requires an invertible covariance matrix, may be invalid. Toda & Phillips (1993) gives a detailed analysis of the problem and Lütkepohl (2006) provides a text–book analysis in section 7.6.

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

$$W(t) = \mu + \sum_{j=1}^{p+1} \pi_j W(t-j) + a(t), \quad t = p, \ldots, T, \tag{3.3}$$

instead of (2.1) then the estimates of the coefficient matrices π_j, $1 \leq j \leq p$ 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 (2.2) where it is not difficult to show that in the regression,

$$W(t+h) = \mu_h + \sum_{j=1}^{p+1} \pi_j^{(h)} W(t+1-j) + \mathsf{U}_h(t), \tag{3.4}$$

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 section 3.2 to infer the rank and null spaces of C.

4 Empirical Illustration

Sims (1972) began a long tradition of studying granger causality in monetary economics (see the surveys in Blanchard (1990) and chapter 1 of Walsh (2003)). Here we extend the work of Dufour & Jouini (2005) and Dufour et al. (2006), which focuses on the Bernanke & Mihov (1998) data set. In particular, we consider Bernanke and Mihov's monthly estimates (constructed by state space methods) of real GDP and the GDP deflator as well as their monthly data on non–borrowed reserves and the federal funds rates.

In the first instance, we took logs and first differences to induce stationarity. The lag length was chosen to be 16 following Dufour et al. (2006) who find that it minimizes the

Akaike information criterion. As is well known from the model selection literature, the AIC is inconsistent for the lag length although it may offer a better fit than consistent criteria like the Baysian or Schwartz information criterion in finite samples (see e.g. McQuarrie & Tsai (1998) or chapter 4 of Lütkepohl (2006)). The bandwidth used in the Bartlett–Newey–West estimator was fixed at $m(T) = h - 1$; this is perhaps too small, but doubling it produced no significant difference to the results and so we will continue to follow Dufour et al. (2006) who proposed this value. Finally, we chose the parametric (or local Monte Carlo) simulation tests instead of the non–parametric ones using $N = 999$ bootstrapped samples and 5% critical values.[7]

To show that this procedure has good size control as well as good power properties we conducted the following experiment. First, we estimated $\widetilde{B}_{12}$. Then we estimated, $\widetilde{B}^1_{12}$, such that $r \nrightarrow_{12} (GDP, P)|_{\mathcal{U}}$, where $\mathcal{U}$ is the subspace associated with the smallest singular value of the restricted matrix. This restriction may be imposed as, $D^1_{12}\text{vec}(\widetilde{B}^1_{12}) = 0$ for some restriction matrix D^1_{12}. Next, we perturbed the restricted model to $\widetilde{B}^{1,\theta}_{12}$ so that, $D^1_{12}\text{vec}(\widetilde{B}^{1,\theta}_{12}) = \theta e_1$, where e_1 is the first standard basis vector and θ is a small number. We will refer to this alternative as $H^\theta_0(1)$. We then simulated 100 data sets using $\widetilde{B}^{1,\theta}_{12}$, for each $\theta = 0, 0.001, \ldots, 0.01$ (larger values of θ led to explosive models). Table gives the rejection rates of $H_0(0)$ and $H_0(1)$.

Table 4.1: Rejection Frequencies for Different Values of θ at Significance 5%.

$\theta =$	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009	0.010
Reject $H_0(0)$	0.78	0.81	0.85	0.85	0.87	0.87	0.89	0.9	0.92	0.95	0.93
Reject $H_0(1)$	0.05	0.08	0.08	0.13	0.16	0.18	0.31	0.37	0.45	0.49	0.55
Reject $H_0(0)$ and accept $H_0(1)$	0.74	0.73	0.77	0.72	0.71	0.69	0.58	0.53	0.48	0.47	0.4
Reject $H_0(0)$ and reject $H_0(1)$	0.04	0.08	0.08	0.13	0.16	0.18	0.31	0.37	0.44	0.48	0.53

The test achieves the correct size of 0.05 but its power to reject $H_0(0)$ is not terribly strong as 22% of the $H_0(0)$ tests were accepted. Thus the testing procedure may underestimate the correct subspaces causality rank; as we have shown above and discussed at length in Camba-Mendez et al. (2003) this is purely a small sample phenomenon. The testing producer does, however, exhibit power against $H^\theta_0(1)$ particulary in the $H_0(0)$ tests.

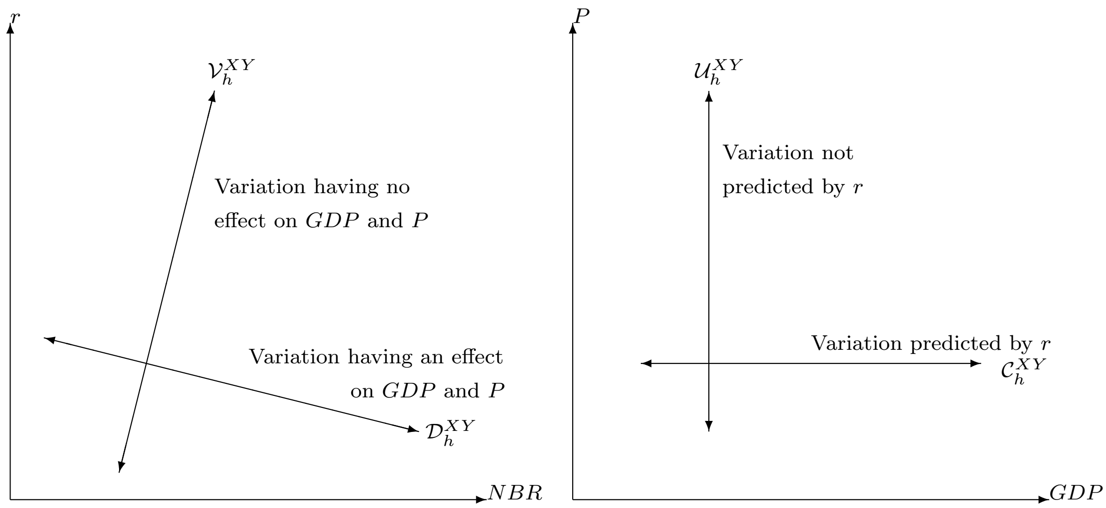
Having proven that the testing procedure performs well, we proceeded to apply it to the full data set. The results are summarized in Tables 6.1 and 6.2. The first four lines are almost

[7]We should note that the Matlab code accompanying this paper allows the user to easily vary all of these parameters.

identical to the analogous ones in Dufour et al. (2006); the small variations from their results are due to simulation error. Our test statistics were also identical to theirs because as we have noted above, the rk(0) statistic is precisely the Wald test used by Dufour et al. (2006). The notable effects in the univariate tests above are that r has predictive power for GDP over the range from 7 months ahead to 18 months ahead while NBR is predictive power for P for three months ahead and has no predictive power beyond that.

From the tests $(NBR, r) \nrightarrow_h (GDP, P)$ we can see that monetary policy has predictive power for output and inflation in the forecast horizons 1–4 and 7–9. From the tests $(NBR, r) \nrightarrow_h (GDP, P)|_{\mathcal{U}}$, we conclude that there is some evidence of subspace non–causality but it is not terribly strong as the subspace non–causality hypothesis is barely accepted. From the tests $(NBR, r)|_{\mathcal{V}} \nrightarrow_h (GDP, P)$, on the other hand, we can see substantial evidence of subspace non–causality; there appears to be a statistically significant relationship between NBR and r along which policy has no predictive power for the other macroeconomic variables. From the associated V_h^{XY} we can see that increases in NBR that are offset by increases r have no effect on GDP and P over all the horizons where monetary policy has effect; thus the line along $(V_h^{XY})_\perp$ can be interpreted as measuring the statistical tradeoff between policy instruments *viz–a–viz* output and inflation. This is illustrated in the first graph of Figure 4.1.

Figure 4.1: Subspace Causality in Monetary Policy



Next we look at the predictive power of the individual policy variables for output and inflation. The tests, $NBR \nrightarrow_h (GDP, P)$ show that NBR leads GDP and P for horizons 1–3 months. There is evidence of subspace non–causality as can be seen from the $NBR \nrightarrow_h$

$(GDP, P)|_{\mathcal{U}}$ tests but it is not very strong. As for the effect of r on GDP and P we can see that variations of the interest rate have tended to precede variations of output and inflation along an ever so slightly positively sloped line. While this may be consistent with there being a statistical Phillips curve (Stock & Watson, 1999), the slope is too flat to be conclusive and is actually negative at horizon 7; therefore we conclude that most of the predictive power for r is for GDP rather than P.[8] See Figure 4.1.

Finally, we consider the effect of the policy variables together on output and inflation individually. Again we find substantial scope for subspace non–causality; in particular, the policy tradeoff observed above emerges much more clearly in these tests, see Figure 4.1. In order to check for robustness to the assumption of stationarity, the analysis was repeated for the augmented–lag regression model (3.4) with no significant difference to the results (see Tables 6.3 and 6.4).

One area of concern in the results is the non–monotonicity of the hypothesis tests. For example, we reject the hypothesis $r \nrightarrow_{12} GDP$ but accept the hypothesis $(NBR, r) \nrightarrow_{12} (GDP, P)$; there are various other examples in the table as well. One should recall, however, that because these are simulated p–values there is no a priori reason why they should be monotonic in small samples; they are guaranteed to be monotonic only asymptotically. Increasing N may not alleviate the problem because the bootstrap samples are based on $\widetilde{B}_h^q$, which may have very different dynamics than B_h in small samples.[9] Additionally we repeated the analysis using the modified Wald test of Lutkepohl & Burda (1997), which is based on a generalized inverse, but the non–monotonicity was still present.[10] We conclude that there is need for more Monte Carlo testing of the procedure we have employed in order to ascertain why the procedure fails to correctly detect rank. For the time being, the procedure proposed should always be coupled with univariate testing.

[8]Indeed, we ran the tests on quarterly data and found that the slopes were even less consistently positive.

[9]It can happen, although it did not happen once in all our tests, that $\widehat{B}_h^q$ in unstable although $\widehat{B}_h$ is stable. In that case, the Matlab code allows the user to check the stability of $\widehat{B}_h^q$ and reconstruct it in the various consistent ways we have discussed. If all else fails, one may "shrink" $\widehat{B}_h^q$ just enough to allow for stability.

[10]The results are available from the author upon request. The Matlab code accompanying this paper allows the user to perform all of the above tests easily.

5 Conclusion

In this paper, we have presented graphical and econometric estimation and inference procedures for testing subspace causality in finite order stationary and non–stationary VAR models. We then applied the methodology to the Bernanke & Mihov (1998) data set and found evidence of subspace non–causality; in particular, (i) the federal funds rate predicts output and inflation along a subspace and (ii) variations of non–borrowed reserves and the federal funds rate along certain directions have no predictive power for output growth and inflation at most of the horizons considered. In the rest of this section we will focus on avenues for future research.

First, having presented a method of estimating subspaces of non–causality it would be useful to have a methodology to test hypotheses on these subspaces. For example, theory suggests that output growth and inflation move proportionally in response to aggregate demand disturbances, this can be formulated as $H_0 : \mathcal{U}_h^{XY} \in \mathfrak{U}$, where $\mathfrak{U}$ consists of all positively sloped lines in output–growth–inflation space. In this regard, it would also be useful to have a test of structural stability that allows us, for example, to test whether the trade–off between non–borrowed reserves and the federal funds rate has changed throughout the sample period.

Second, many of the estimated p–values were not monotonic with respect to the severity of the restrictions. Preliminary Monte Carlo testing has indicated that indeed this problem is pervasive. There is therefore a need for more Monte Carlo testing of the procedure to understand its behavior in small samples. In this regard, it would be interesting to see how the test behaves as we vary the HAC estimator or by using the new methods proposed by Kiefer & Vogelsang (2005).

Finally, although the procedure outlined in this paper can easily be extended to test causality *up to* horizon h, rather than just *at* horizon h,[11] there is still need for a simple long run causality test. Bruneau & Jondeau (1999) proposed such a test for cointegrated VARs but, as Yamamoto & Kurozumi (2006) have noted, the test may be misleading if the covariance matrix of the test is singular due to the super–consistency of the estimates; Yamamoto & Kurozumi (2006) propose estimating the rank of the covariance matrix of the statistic then using a generalized inverse. Clearly, a more automated solution is desirable.

[11] Simply stack h equations of the form (2.2) for horizons ranging over 1 to h and estimate by OLS; the residual is then an MA($h-1$) process and testing can proceed as usual using a HAC estimator.

6 Appendix

6.1 Canonical Correlations and Subspace Causality in Regression Models

In this section we show how to estimate the subspaces of non–causality using partial canonical correlation analysis. A detailed discussion of partial canonical correlations can be found in Reinsel (2003) and Reinsel & Velu (1998).

Suppose we have zero–mean random vectors $X \in \mathbb{R}^n$, $Z \in \mathbb{R}^k$, $Y_i \in \mathbb{R}^m$ for $i = 1, \dots, p$, $Y = (Y_1', \dots, Y_p')'$, and let the variance matrix of $(X', Y', Z')'$ be,

$$\Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} & \Sigma_{XZ} \\ \Sigma_{YX} & \Sigma_{YY} & \Sigma_{YZ} \\ \Sigma_{ZX} & \Sigma_{ZY} & \Sigma_{ZZ} \end{bmatrix} = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY_1} & \cdots & \Sigma_{XY_p} & \Sigma_{XZ} \\ \Sigma_{Y_1X} & \Sigma_{Y_1Y_1} & \cdots & \Sigma_{Y_1Y_p} & \Sigma_{Y_1Z} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Sigma_{Y_pX} & \Sigma_{Y_pY_1} & \cdots & \Sigma_{Y_pY_p} & \Sigma_{Y_pZ} \\ \Sigma_{ZX} & \Sigma_{ZY_1} & \cdots & \Sigma_{ZY_p} & \Sigma_{ZZ} \end{bmatrix}$$

The best linear predictor of X in terms of Y and Z is then $\Sigma_{XY\cdot Z}\Sigma^{\dagger}_{YY\cdot Z}Y + \Sigma_{XZ\cdot Y}\Sigma^{\dagger}_{ZZ\cdot Y}Z$, where $\Sigma^{\dagger}_{YY\cdot Z}$ is the Moore–Penrose inverse of $\Sigma_{YY\cdot Z}$ (see Horn & Johnson (1985)) and the partial covariance matrices are given by,

$$\Sigma_{XY\cdot Z} = \Sigma_{XY} - \Sigma_{XZ}\Sigma^{\dagger}_{ZZ}\Sigma_{ZY} \qquad \Sigma_{XZ\cdot Y} = \Sigma_{XZ} - \Sigma_{XY}\Sigma^{\dagger}_{YY}\Sigma_{YZ}$$
$$\Sigma_{YY\cdot Z} = \Sigma_{YY} - \Sigma_{YZ}\Sigma^{\dagger}_{ZZ}\Sigma_{ZY} \qquad \Sigma_{ZZ\cdot Y} = \Sigma_{ZZ} - \Sigma_{ZY}\Sigma^{\dagger}_{YY}\Sigma_{YZ}$$

Suppose now we are interested in directions along which X and Y have the strongest correlation after conditioning on Z. Thus we are interested in,

$$\begin{aligned} \rho^1_{XY\cdot Z} &= \sup\{|\mathrm{corr}(U, V)| : U \in \mathrm{sp}\{X, Z\} - \mathrm{sp}\{Z\}, V \in \mathrm{sp}\{Y, Z\} - \mathrm{sp}\{Z\}\} \\ &= \sup\{|\langle U, V\rangle| : U \in \mathrm{sp}\{X, Z\} - \mathrm{sp}\{Z\}, V \in \mathrm{sp}\{Y, Z\} - \mathrm{sp}\{Z\}, \|U\| = \|V\| = 1\} \end{aligned}$$

Now the generic element of $\mathrm{sp}\{X, Z\} - \mathrm{sp}\{Z\}$ is $x'P(X|\mathrm{sp}\{X, Z\} - \mathrm{sp}\{Z\}) = x'(X - \Sigma_{XZ}\Sigma^{\dagger}_{ZZ}Z)$ for some $x \in \mathbb{R}^n$. On the other hand, the generic element of $\mathrm{sp}\{Y, Z\} - \mathrm{sp}\{Z\}$ is $y'P(Y|\mathrm{sp}\{Y, Z\} - \mathrm{sp}\{Z\}) = y'(Y - \Sigma_{YZ}\Sigma^{\dagger}_{ZZ}Z)$ with $y \in \mathbb{R}^{mp}$. Thus we have,

$$\rho^1_{XY\cdot Z} = \max\{x'\Sigma_{XY\cdot Z}y : x \in \mathbb{R}^n, y \in \mathbb{R}^{mp}, x'\Sigma_{XX\cdot Z}x = y'\Sigma_{YY\cdot Z}y = 1\}$$

Thus we arrive at the usual canonical correlations equations (see e.g. Anderson (2003)) with the covariance matrices replaced by the partial covariance matrices that factor out the linear effect of Z. The solutions x_1 and y_1 to the above maximization problem are then used to find the canonical variates, $U_1 = x_1'(X - \Sigma_{XZ}\Sigma_{ZZ}^{\dagger}Z)$ and $V_1 = y_1'(Y - \Sigma_{YZ}\Sigma_{ZZ}^{\dagger}Z)$ so that finally, $\rho_{XY\cdot Z}^1 = |\langle U_1, V_1\rangle|$. The procedure is then carried on recursively for $i \geq 1$ as,

$$\begin{aligned}\rho_{XY\cdot Z}^{i+1} = \sup\{|\mathrm{corr}(U,V)| : U \in \mathrm{sp}\{X,Z\} - \mathrm{sp}\{U_1,\dots,U_i,Z\},\\ V \in \mathrm{sp}\{Y,Z\} - \mathrm{sp}\{V_1,\dots,V_i,Z\}\},\end{aligned}$$

which similarly reduces to,

$$\begin{aligned}\rho_{XY\cdot Z}^{i+1} = \max\{x'\Sigma_{XY\cdot Z}y : x \in \mathbb{R}^n, y \in \mathbb{R}^{mp}, x'\Sigma_{XX\cdot Z}x = y'\Sigma_{YY\cdot Z}y = 1,\\ x'\Sigma_{XX\cdot Z}x_j = y'\Sigma_{YY\cdot Z}y_j = 0, j = 1,\dots,i\}\end{aligned}$$

These can be solved to obtain further canonical correlations until the algorithm terminates after $\min\{n, mp\}$ steps. As shown in Anderson (2003), the solution to the algorithm can be represented as the set of all solutions to,

$$\begin{bmatrix} -\lambda_i\Sigma_{XX\cdot Z} & \Sigma_{XY\cdot Z} \\ \Sigma_{YX\cdot Z} & -\lambda_i\Sigma_{YY\cdot Z}\end{bmatrix}\begin{bmatrix} x_i \\ y_i\end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad x_i'\Sigma_{XX\cdot Z}x_j = \delta_{ij}, \qquad y_i'\Sigma_{YY\cdot Z}y_j = \delta_{ij},$$

$$U_i = x_i'(X - \Sigma_{XZ}\Sigma_{ZZ}^{\dagger}Z), \quad V_i = y_i'(Y - \Sigma_{YZ}\Sigma_{ZZ}^{\dagger}Z), \quad \rho_{XY\cdot Z}^i = \lambda_i = |\langle U_i, V_i\rangle|, \tag{6.1}$$

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

Clearly the canonical variates associate with canonical correlations of zero define directions of uncorrelated conditional variation between X and Y. That is, $\{x_i : \rho_{XY\cdot Z}^i = 0, i = 1,\dots,\min\{n, mp\}\}$ are the directions along which variations in X are not attributable to variations in Y after controlling for Z. This can easily be seen from equation (6.1) where if $\lambda_i = 0$ then $x_i'\Sigma_{XY\cdot Z} = 0$. Thus the subspace, $\mathcal{U} = \mathrm{span}\{x_i : \rho_{XY\cdot Z}^i = 0, i = 1,\dots,\min\{n, mp\}\}$ is the regression analogue to the subspace of non–causality, $\mathcal{U}_h^{XY}$.

Suppose that instead we are interested in the directions along which X and the components of Y have the strongest correlation after conditioning on Z. In order to study this correlation, we need a device that allows us to look at the correlation of X with each component of Y individually. Thus we will consider correlations between $\mathcal{X} = \sum_{i=1}^p \phi_i E_i X$, where $E_i = [0 \ \cdots \ I_n^{\text{(ith place)}} \ \cdots 0]' \in \mathbb{R}^{np\times n}$ and $\mathcal{Y} = \sum_{j=1} \phi_j Y_j$, where the random vector $\phi = (\phi_1, \ldots, \phi_p)'$ is independent of, X, Y, and Z, and satisfies $\mathbb{E}(\phi\phi') = I_p$. We will also need $\mathcal{Z} = \sum_{i=1}^p \phi_i F_i Z$, where $F_i = [0 \ \cdots \ I_k^{\text{(ith place)}} \ \cdots 0]' \in \mathbb{R}^{kp\times k}$. This construction makes sense because the covariance between $\mathcal{X}$ and $\mathcal{Y}$ is,

$$\Sigma_{\mathcal{X}\mathcal{Y}} = \begin{bmatrix} \Sigma_{XY_1} \\ \vdots \\ \Sigma_{XY_p} \end{bmatrix},$$

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

$$\Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = \begin{bmatrix} \Sigma_{XY_1\cdot Z} \\ \vdots \\ \Sigma_{XY_p\cdot Z} \end{bmatrix} = \begin{bmatrix} \Sigma_{XY_1} - \Sigma_{XZ}\Sigma_{ZZ}^{\dagger}\Sigma_{ZY_1} \\ \vdots \\ \Sigma_{XY_p} - \Sigma_{XZ}\Sigma_{ZZ}^{\dagger}\Sigma_{ZY_p} \end{bmatrix}$$

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

Now define the first canonical correlation as,

$$\theta_{XY\cdot Z}^1 = \sup\{|\text{corr}(U,V)| : U \in \text{sp}\{\mathcal{X}, Z\} - \text{sp}\{Z\}, V \in \text{sp}\{\mathcal{Y}, Z\} - \text{sp}\{Z\}\},$$

and following the same reasoning as above we find that this can be rewritten as,

$$\theta_{XY\cdot Z}^1 = \max\left\{ x'\Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}}y = \sum_{i=1}^p x_i'\Sigma_{XY_i\cdot Z}y : x = (x_1', \ldots, x_p')' \in \mathbb{R}^{np}, y \in \mathbb{R}^m, \right.$$
$$\left. x'\Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}}x = y'\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}}y = 1 \right\}$$

To solve this we formulate the criterion function as, $\mathscr{L} = x'\Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}}y + \frac{1}{2}\lambda(1 - x'\Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}}x) + \frac{1}{2}\mu(1 - y'\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}}y)$ and find the first order conditions,

$$\frac{\partial\mathscr{L}}{\partial x} = y'\Sigma_{\mathcal{Y}\mathcal{X}\cdot\mathcal{Z}} - \lambda x'\Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}} = 0$$
$$\frac{\partial\mathscr{L}}{\partial y} = x'\Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} - \mu y'\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}} = 0$$

Multiplying the first equation on the right by x and the second equation by y, we get, $\lambda = \mu = x'\Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}}y$. Thus our problem boils down the following generalized eigenvalue problem,

$$\begin{bmatrix} -\lambda\Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}} & \Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} \\ \Sigma_{\mathcal{Y}\mathcal{X}\cdot\mathcal{Z}} & -\lambda\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Having found the first canonical correlation, $\theta^1_{XY\cdot Z}$, the optimal vectors $x = x_1$ and $y = y_1$, and the associated canonical variates $U_1 = x_1'(\mathcal{X} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\mathcal{Z})$ and $V_1 = y_1'(\mathcal{Y} - \Sigma_{\mathcal{Y}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\mathcal{Z})$, we proceed recursively for $i \geq 1$ as,

$$\theta^{i+1}_{XY\cdot Z} = \sup\{|\mathrm{corr}(U,V)| : U \in \overline{\mathrm{sp}}\{\mathcal{X},\mathcal{Z}\} - \overline{\mathrm{sp}}\{U_1,\ldots,U_i,\mathcal{Z}\},$$
$$V \in \overline{\mathrm{sp}}\{\mathcal{Y},\mathcal{Z}\} - \overline{\mathrm{sp}}\{V_1,\ldots,V_i,\mathcal{Z}\}\}$$

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

$$\begin{bmatrix} -\lambda_i\Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}} & \Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} \\ \Sigma_{\mathcal{Y}\mathcal{X}\cdot\mathcal{Z}} & -\lambda_i\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}} \end{bmatrix} \begin{bmatrix} x^i \\ y^i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$x^{i\prime}\Sigma_{\mathcal{X}\mathcal{X}\cdot\mathcal{Z}}x^j = \delta_{ij}, \qquad y^{i\prime}\Sigma_{\mathcal{Y}\mathcal{Y}\cdot\mathcal{Z}}y^j = \delta_{ij},$$
$$U_i = x^{i\prime}(\mathcal{X} - \Sigma_{\mathcal{X}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\mathcal{Z}), \quad V_i = y^{i\prime}(\mathcal{Y} - \Sigma_{\mathcal{Y}\mathcal{Z}}\Sigma^{\dagger}_{\mathcal{Z}\mathcal{Z}}\mathcal{Z}), \quad \theta^i_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}} = \lambda_i = |\langle U_i, V_i\rangle|, \qquad (6.2)$$

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 X and the components of Y. That is, $\{y^i : \theta^i_{XY\cdot Z} = 0, i = 1,\ldots,\min\{np,m\}\}$ are the directions along which variations in X are not attributable to the joint variations of the components of Y after controlling for Z. This can easily be seen from equation (6.2) where if $\lambda_i = 0$ then $\Sigma_{\mathcal{X}\mathcal{Y}\cdot\mathcal{Z}}y^i = 0$, which is equivalent to $\Sigma_{XY_j\cdot Z}y^i = 0$ for $j = 1,\ldots,p$. Thus the subspace, $\mathcal{V} = \mathrm{span}\{y_i : \rho^i_{XY\cdot Z} = 0, i = 1,\ldots,\min\{np,m\}\}$ is the regression analogue to the subspace of non–causality, $\mathcal{V}^{XY}_h$.

6.2 Tests for Subspace Non–causality

Table 6.1: Causality Test Simulated p-values and Non–causal Directions for the Horizons 1-12.

h	1	2	3	4	5	6	7	8	9	10	11	12
$NBR \nrightarrow_h GDP$	0.199	0.273	0.282	0.394	0.833	0.736	0.815	0.710	0.357	0.277	0.375	0.333
$NBR \nrightarrow_h P$	**0.003**	**0.015**	**0.005**	0.136	0.168	0.168	0.118	0.154	0.219	0.200	0.504	0.582
$r \nrightarrow_h GDP$	0.238	0.172	0.054	0.073	0.080	0.077	**0.004**	**0.004**	**0.001**	**0.001**	**0.003**	**0.001**
$r \nrightarrow_h P$	0.111	0.121	0.139	0.360	0.414	0.398	0.327	0.358	0.202	0.142	0.426	0.757
$(NBR, r) \nrightarrow_h (GDP, P)$	**0.001**	**0.001**	**0.001**	**0.007**	0.059	0.090	**0.016**	**0.019**	**0.025**	0.197	0.784	0.509
$(NBR, r) \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$	**0.008**	**0.013**	**0.017**	0.051			0.070	0.067	0.255			
U_h^{XY}				$\begin{bmatrix}0.0952\\0.9955\end{bmatrix}$			$\begin{bmatrix}-0.0642\\0.9979\end{bmatrix}$	$\begin{bmatrix}-0.0979\\0.9952\end{bmatrix}$	$\begin{bmatrix}-0.1550\\0.9879\end{bmatrix}$			
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h (GDP, P)$	0.274	0.258	0.137	0.192			0.445	0.463	0.212			
V_h^{XY}	$\begin{bmatrix}0.0576\\0.9983\end{bmatrix}$	$\begin{bmatrix}0.0604\\0.9982\end{bmatrix}$	$\begin{bmatrix}0.0412\\0.9992\end{bmatrix}$	$\begin{bmatrix}0.0327\\0.9995\end{bmatrix}$			$\begin{bmatrix}0.1532\\0.9882\end{bmatrix}$	$\begin{bmatrix}0.1631\\0.9866\end{bmatrix}$	$\begin{bmatrix}0.1681\\0.9858\end{bmatrix}$			
$NBR \nrightarrow_h (GDP, P)$	**0.008**	**0.035**	**0.040**	0.148	0.304	0.332	0.144	0.166	0.268	0.362	0.569	0.675
$NBR \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$	0.056	0.080	0.180									
U_h^{XY}	$\begin{bmatrix}0.1489\\0.9888\end{bmatrix}$	$\begin{bmatrix}0.1206\\0.9927\end{bmatrix}$	$\begin{bmatrix}0.1282\\0.9918\end{bmatrix}$									
$r \nrightarrow_h (GDP, P)$	0.107	0.113	0.076	0.141	0.079	0.204	**0.022**	**0.023**	**0.007**	**0.003**	**0.010**	**0.029**
$r \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$							0.196	0.309	0.213	0.086	0.353	0.734
U_h^{XY}							$\begin{bmatrix}0.0147\\0.9999\end{bmatrix}$	$\begin{bmatrix}-0.0278\\0.9996\end{bmatrix}$	$\begin{bmatrix}-0.0427\\0.9991\end{bmatrix}$	$\begin{bmatrix}-0.0231\\0.9997\end{bmatrix}$	$\begin{bmatrix}-0.0408\\0.9992\end{bmatrix}$	$\begin{bmatrix}0.0154\\0.9999\end{bmatrix}$
$(NBR, r) \nrightarrow_h GDP$	**0.021**	**0.030**	**0.002**	**0.019**	0.066	0.089	**0.014**	**0.004**	**0.003**	**0.002**	**0.014**	**0.023**
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h GDP$	0.497	0.455	0.248	0.113			0.650	0.565	0.511	0.422	0.308	0.075
V_h^{XY}	$\begin{bmatrix}0.0676\\0.9977\end{bmatrix}$	$\begin{bmatrix}0.0689\\0.9976\end{bmatrix}$	$\begin{bmatrix}0.0459\\0.9989\end{bmatrix}$	$\begin{bmatrix}0.0325\\0.9995\end{bmatrix}$			$\begin{bmatrix}0.2054\\0.9787\end{bmatrix}$	$\begin{bmatrix}0.2204\\0.9754\end{bmatrix}$	$\begin{bmatrix}0.2155\\0.9765\end{bmatrix}$	$\begin{bmatrix}0.2087\\0.9780\end{bmatrix}$	$\begin{bmatrix}0.1888\\0.9820\end{bmatrix}$	$\begin{bmatrix}0.1316\\0.9913\end{bmatrix}$
$(NBR, r) \nrightarrow_h P$	**0.001**	**0.001**	**0.001**	**0.043**	0.069	0.066	0.073	**0.037**	0.103	0.288	0.735	0.914
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h P$	0.103	0.108	0.137	0.405				0.254				
V_h^{XY}	$\begin{bmatrix}-0.0015\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0070\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0086\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0336\\0.9994\end{bmatrix}$				$\begin{bmatrix}-0.0165\\0.9999\end{bmatrix}$				

Table 6.2: Causality Test Simulated p-values and Non–causal Directions for the Horizons 13-24.

h	13	14	15	16	17	18	19	20	21	22	23	24
$NBR \nrightarrow_h GDP$	0.504	0.717	0.572	0.773	0.941	0.907	0.753	0.759	0.578	0.513	0.341	0.237
$NBR \nrightarrow_h P$	0.304	0.346	0.280	0.444	0.814	0.875	0.464	0.144	0.262	0.196	0.309	0.568
$r \nrightarrow_h GDP$	**0.002**	**0.016**	**0.014**	**0.025**	**0.047**	**0.033**	0.056	0.057	0.201	0.107	0.210	0.263
$r \nrightarrow_h P$	0.616	0.361	0.365	0.430	0.698	0.492	0.428	0.264	0.185	0.257	0.138	0.283
$(NBR, r) \nrightarrow_h (GDP, P)$	0.090	0.108	0.308	0.417	0.470	0.702	0.661	0.313	0.555	0.681	0.886	0.863
$(NBR, r) \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$												
U_h^{XY}												
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h (GDP, P)$												
V_h^{XY}												
$NBR \nrightarrow_h (GDP, P)$	0.558	0.783	0.541	0.559	0.893	0.724	0.560	0.464	0.561	0.770	0.793	0.689
$NBR \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$												
U_h^{XY}												
$r \nrightarrow_h (GDP, P)$	**0.004**	**0.009**	**0.025**	**0.024**	0.142	0.094	0.265	0.076	0.119	0.233	0.261	0.206
$r \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$	0.463	0.248	0.414	0.723								
U_h^{XY}	$\begin{bmatrix} -0.0115 \\ 0.9999 \end{bmatrix}$	$\begin{bmatrix} -0.0320 \\ 0.9995 \end{bmatrix}$	$\begin{bmatrix} -0.0365 \\ 0.9993 \end{bmatrix}$	$\begin{bmatrix} -0.0703 \\ 0.9975 \end{bmatrix}$								
$(NBR, r) \nrightarrow_h GDP$	**0.003**	**0.046**	**0.031**	0.085	0.192	0.247	**0.039**	**0.046**	0.070	**0.032**	0.175	0.127
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h GDP$	0.056	0.070	0.079				**0.035**	**0.038**		0.111		
V_h^{XY}	$\begin{bmatrix} 0.1639 \\ 0.9865 \end{bmatrix}$	$\begin{bmatrix} 0.1305 \\ 0.9914 \end{bmatrix}$	$\begin{bmatrix} 0.1507 \\ 0.9886 \end{bmatrix}$							$\begin{bmatrix} 0.0431 \\ 0.9991 \end{bmatrix}$		
$(NBR, r) \nrightarrow_h P$	0.627	0.602	0.732	0.872	0.932	0.953	0.789	0.233	0.328	0.818	0.680	0.897
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h P$												
V_h^{XY}												

Table 6.3: Causality Test Simulated p-values and Non–causal Directions for the Horizons 1-12 with Lag–Augmented Regression.

h	1	2	3	4	5	6	7	8	9	10	11	12
$NBR \nrightarrow_h GDP$	0.228	0.340	0.295	0.386	0.867	0.745	0.826	0.772	0.325	0.313	0.352	0.302
$NBR \nrightarrow_h P$	**0.003**	**0.013**	**0.006**	0.159	0.119	0.068	0.164	0.165	0.267	0.298	0.498	0.607
$r \nrightarrow_h GDP$	0.233	0.163	0.062	0.099	0.121	0.155	**0.001**	**0.001**	**0.001**	**0.002**	**0.001**	**0.001**
$r \nrightarrow_h P$	0.094	0.126	0.338	0.386	0.591	0.419	0.400	0.2730	0.121	0.205	0.436	0.784
$(NBR, r) \nrightarrow_h (GDP, P)$	**0.001**	**0.001**	**0.001**	**0.034**	0.063	0.060	**0.017**	**0.020**	0.060	0.296	0.718	0.1160
$(NBR, r) \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$	**0.014**	**0.010**	**0.024**	0.058			**0.049**	0.209				
U_h^{XY}				$\begin{bmatrix}0.0873\\0.9962\end{bmatrix}$				$\begin{bmatrix}-0.1552\\0.9879\end{bmatrix}$				
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h (GDP, P)$	0.234	0.232	0.261	0.453			0.469	0.476				
V_h^{XY}	$\begin{bmatrix}0.0627\\0.9980\end{bmatrix}$	$\begin{bmatrix}0.0507\\0.9987\end{bmatrix}$	$\begin{bmatrix}0.0434\\0.9991\end{bmatrix}$	$\begin{bmatrix}0.0549\\0.9985\end{bmatrix}$			$\begin{bmatrix}0.1592\\0.9872\end{bmatrix}$	$\begin{bmatrix}0.2001\\0.9798\end{bmatrix}$				
$NBR \nrightarrow_h (GDP, P)$	**0.004**	0.056	**0.020**	0.159	0.220	0.262	0.152	0.2310	0.304	0.501	0.581	0.504
$NBR \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$	0.097		0.200									
U_h^{XY}	$\begin{bmatrix}0.1690\\0.9856\end{bmatrix}$		$\begin{bmatrix}0.1404\\0.9901\end{bmatrix}$									
$r \nrightarrow_h (GDP, P)$	0.086	0.110	0.167	0.235	0.130	0.303	**0.006**	**0.006**	**0.003**	**0.004**	**0.009**	**0.016**
$r \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$							0.313	0.264	0.106	0.168	0.337	0.711
U_h^{XY}							$\begin{bmatrix}-0.0062\\1.0000\end{bmatrix}$	$\begin{bmatrix}-0.0355\\0.9994\end{bmatrix}$	$\begin{bmatrix}-0.0258\\0.9997\end{bmatrix}$	$\begin{bmatrix}-0.0250\\0.9997\end{bmatrix}$	$\begin{bmatrix}-0.0255\\0.9997\end{bmatrix}$	$\begin{bmatrix}0.0001\\1.0000\end{bmatrix}$
$(NBR, r) \nrightarrow_h GDP$	**0.027**	**0.033**	**0.004**	**0.029**	0.113	0.101	**0.003**	**0.001**	**0.003**	**0.005**	**0.016**	**0.004**
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h GDP$	0.528	0.353	0.223	0.320			0.606	0.707	0.313	0.345	0.274	**0.012**
V_h^{XY}	$\begin{bmatrix}0.0726\\0.9974\end{bmatrix}$	$\begin{bmatrix}0.0572\\0.9984\end{bmatrix}$	$\begin{bmatrix}0.0477\\0.9989\end{bmatrix}$	$\begin{bmatrix}0.0562\\0.9984\end{bmatrix}$			$\begin{bmatrix}0.2283\\0.9736\end{bmatrix}$	$\begin{bmatrix}0.2732\\0.9620\end{bmatrix}$	$\begin{bmatrix}0.1896\\0.9819\end{bmatrix}$	$\begin{bmatrix}0.2017\\0.9795\end{bmatrix}$	$\begin{bmatrix}0.1903\\0.9817\end{bmatrix}$	
$(NBR, r) \nrightarrow_h P$	**0.001**	**0.006**	**0.003**	0.059	0.064	0.086	0.080	0.054	0.132	0.426	0.685	0.898
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h P$	0.124	0.133	0.306									
V_h^{XY}	$\begin{bmatrix}0.0036\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0091\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0132\\0.9999\end{bmatrix}$									

Table 6.4: Causality Test Simulated p-values and Non–causal Directions for the Horizons 13-24 with Lag–Augmented Regression.

h	13	14	15	16	17	18	19	20	21	22	23	24
$NBR \nrightarrow_h GDP$	0.479	0.815	0.595	0.812	0.934	0.893	0.724	0.694	0.549	0.325	0.308	0.139
$NBR \nrightarrow_h P$	0.346	0.379	0.508	0.452	0.845	0.797	0.352	0.228	0.267	0.143	0.379	0.478
$r \nrightarrow_h GDP$	**0.004**	**0.010**	**0.012**	**0.029**	**0.040**	0.059	**0.027**	0.137	0.164	0.144	0.153	0.224
$r \nrightarrow_h P$	0.499	0.366	0.548	0.4510	0.701	0.438	0.372	0.364	0.246	0.308	0.198	0.254
$(NBR, r) \nrightarrow_h (GDP, P)$	0.074	0.175	0.460	0.3380	0.399	0.693	0.447	0.318	0.585	0.619	0.889	0.882
$(NBR, r) \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$												
U_h^{XY}												
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h (GDP, P)$												
V_h^{XY}												
$NBR \nrightarrow_h (GDP, P)$	0.734	0.802	0.558	0.799	0.859	0.522	0.510	0.676	0.587	0.644	0.740	0.637
$NBR \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$												
U_h^{XY}												
$r \nrightarrow_h (GDP, P)$	**0.004**	**0.016**	**0.036**	**0.018**	0.066	0.236	0.126	0.214	0.226	0.087	0.231	0.263
$r \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$	0.376	0.308	0.791	0.551								
U_h^{XY}	$\begin{bmatrix}-0.0128\\0.9999\end{bmatrix}$	$\begin{bmatrix}-0.0276\\0.9996\end{bmatrix}$	$\begin{bmatrix}-0.0570\\0.9984\end{bmatrix}$	$\begin{bmatrix}-0.0558\\0.9984\end{bmatrix}$								
$(NBR, r) \nrightarrow_h GDP$	**0.002**	0.119	0.080	0.204	0.355	0.137	**0.020**	**0.035**	**0.047**	**0.0210**	0.120	0.119
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h GDP$	0.167						**0.042**	0.121	0.067	0.108		
V_h^{XY}	$\begin{bmatrix}0.2115\\0.9774\end{bmatrix}$							$\begin{bmatrix}0.0229\\0.9997\end{bmatrix}$	$\begin{bmatrix}-0.0380\\0.9993\end{bmatrix}$	$\begin{bmatrix}0.0244\\0.9997\end{bmatrix}$		
$(NBR, r) \nrightarrow_h P$	0.751	0.634	0.867	0.930	0.964	0.859	0.635	0.427	0.424	0.787	0.729	0.802
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h P$												
V_h^{XY}												

References

Al-Sadoon, M. (2009). Causality along subspaces: Theory. Cambridge Working Papers in Economics 0919, Faculty of Economics, University of Cambridge.

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.

Andrews, D. W. K. (1987). Asymptotic results for generalized wald tests. *Econometric Theory*, *3*(3), 348–358.

Bernanke, B. S. & Mihov, I. (1998). Measuring monetary policy. *The Quarterly Journal of Economics*, *113*(3), 869–902.

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.

Blanchard, O. J. (1990). Why does money affect output? a survey. In B. M. Friedman & F. H. Hahn (Eds.), *Handbook of Monetary Economics*, volume 2 of *Handbook of Monetary Economics* chapter 15, (pp. 779–835). Elsevier.

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. (2009). Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling. *Econometric Reviews*, *28*(6), 581–611.

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.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? In J. B. Taylor & M. Woodford (Eds.), *Handbook* *of Macroeconomics*, volume 1 of *Handbook of Macroeconomics* chapter 2, (pp. 65–148). Elsevier.

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.

Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and their Applications*. Cambridge, United Kingdom: 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. & Jouini, T. (2005). Finite-sample simulation-based inference in var models with applications to order selection and causality testing. Technical report.

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. & Valery, P. (2009). Hypothesis tests when rank conditions fail: a smooth regularization approach. mimeo. Available at http://neumann.hec.ca/pages/pascale.valery/Dufour_Valery_RegInfer_2009_09_10.pdf.

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.

Godfrey, L. (2009). *Bootstrap Tests for Regression Models*. New York, NY, USA: Palgrave Macmillan.

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 Tests in Econometrics. Oxford, UK: Oxford University Press.

Hall, P. (1992). *Bootstrap and Edgeworth Expansion*. Springer Series in Statistics. New York, USA: Springer.

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

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

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

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

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

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: Wold Scientific Co. Pte. Ltd.

Moore, D. S. (1977). Generalized inverses, wald's method, and the construction of chi-squared tests of fit. *Journal of the American Statistical Association*, *72*(357), 131–137.

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

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

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. (1992). *Numerical Recipes in C: the art of scientific computing* (2 ed.). 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.

Sims, C. A. (1972). Money, income, and causality. *American Economic Review*, *62*(4), 540–52.

Stewart, G. W. (1993). On the early history of the singular value decomposition. *SIAM Review*, *35*(4), 551–566.

Stock, J. H. & Watson, M. W. (1999). Business cycle fluctuations in us macroeconomic time series. In J. B. Taylor & M. Woodford (Eds.), *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics* chapter 1, (pp. 4–64). Elsevier.

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.

Walsh, C. E. (2003). *Monetary Theory and Policy*. Cambridge, MA, USA: MIT Press.

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

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.