Testing for multistep causality

Jörg Breitung*  
University of Bonn  
Institute of Econometrics  
Adenauerallee 24-42  
53113 Bonn  
Germany

Bertrand Candelon  
Department of Economics  
Maastricht University  
P.O. Box 616  
6200 MD Maastricht  
The Netherlands

December 2007

Keyword: Causality tests, vector autoregression, Wald test, local power

Abstract

In this paper we study the small sample properties of two alternative tests for multistep causality as proposed by Lütkepohl and Burda (Journal of Econometrics, 1997) and Dufour, Pelletier, Renault (Journal of Econometrics, 2006). We show that under a sequence of local alternatives the former test is uniformly more powerful than the computationally simpler test suggested in the latter paper. We also consider a modification of the test based on a GLS regression. This test may improve the power of the original tests dramatically. Furthermore, a simple test for long-run causality is proposed that is based on a semi-parametric estimator of the long-run covariance.

*We wish to thank Gianluca Cubbada, Jean-Marie Dufour, Helmut Lütkepohl, Tommaso Pioretti and the participants of the seminar at Ente Enaudi, Rome, the Fourth Macroeconomic Workshop, Halle, and the European Meeting of the Econometric Society in Budapest for helpful comments and suggestions.
1 Introduction

The original definition of causality suggested by Wiener (1956) and Granger (1969) focuses on one-step ahead predictions. If the past of the variable $y_2$ helps to predict the variable $y_1$ one period ahead (conditional on the past of the additional set of variables comprised in the vector $y_3$) the variable $y_2$ is called a (prima facie) cause of $y_1$. As noted by Dufour and Renault (1998) this concept may be too restrictive if the causal link operates across several time periods and intermediate variables. In these cases it is natural to generalize the original causality concept to multistep causality defined as the potential of the variable $y_2$ to predict $y_1$ $h$ step ahead. Furthermore this methodology is able to distinguish short run from long run causality (cf. Dufour and Renault 1991). An alternative concept for defining short and long run causality in the frequency domain was suggested by Granger (1969), Hosoya (1991), and Granger and Lin (1995). Empirical test procedures for causality at some given frequencies were suggested by Geweke (1982), Yao and Hosoya (2000), Hosoya (2001) and Breitung and Candelon (2006).

Lütkepohl and Burda (henceforth LB) (1997) considered Wald tests for multi-step causality based on a vector autoregressive (VAR) framework. An important problem with their test is that the hypothesis of non-causality at horizon $h > 1$ implies highly nonlinear restrictions on the VAR parameters. Therefore, the nonlinear Wald statistic is difficult to compute if the VAR system has more than three variables or more than a single lag. Moreover, as pointed out by LB (1997), the Wald test may fail to have an asymptotic $\chi^2$ distribution. To overcome this difficulty, LB (1997) suggest modifications of the test that yield asymptotically $\chi^2$ distributed test statistics.

To sidestep the difficulties due to the test of nonlinear hypotheses, Hill (2007) suggests a recursive parametric representation of causal chains that can be tested by using a sequential test procedure of linear hypotheses. For illustration consider a three-variate system such that $y_{1,t}$, $y_{2,t}$ and $y_{3,t}$ are scalar time series. There is no causal link from variable $y_2$ to $y_1$ up to horizon 2, if (i) $y_2$ is not a (one-step ahead) cause of $y_1$ and (ii) $y_1$ is not a (one-step ahead) cause of $y_3$ or $y_2$ is not
a (one-step ahead) cause of $y_1$. Therefore, Hill (2007) proposed using a sequence of tests to investigate the causal link with $h \geq 2$. An important drawback of this approach is that Bonferroni-type bounds have to be employed to control the overall size of the test. Furthermore, the number of tests grows rapidly with increasing time horizon and the number of variables in the system.

In this paper we investigate the asymptotic properties of the nonlinear (LB) and linear (DPR) test procedures. It is shown that the LB test is uniformly more powerful than the DPR test. Depending on the parameters of the data generating process, the power loss may become arbitrarily large. In the case where the LB test fails to possess an asymptotic $\chi^2$ distribution, it is shown to be conservative. To improve the power of the test statistics we also consider tests based on a GLS regression of the linear $h$-step ahead representation considered by DPR. Another contribution of the paper is to suggest a simple test for long-run causality, that is, a test for causality at $h \to \infty$ in a cointegrated VAR.

The paper is organized as follows. In Section 2 we review the test statistics proposed by LB (1997) and DPR (2006). The asymptotic properties of the original tests are studied in Section 3 and in section 4 we investigate the asymptotic and finite sample properties of related tests based on generalized least-squares (GLS). In section 7 a test statistic for long-run causality is suggested that is based on a semi-parametric estimator of the long-run covariance between the variables. Section 8 concludes.

2 Test procedures

Let $y_t = \begin{bmatrix} y_{1,t}' & y_{2,t}' & y_{3,t}' \end{bmatrix}'$ be a $k = k_1 + k_2 + k_3$ dimensional vector of stationary time series. We say that $y_{2,t}$ causes $y_{1,t}$ at horizon $h$ if $(y_{2,\tau} : \tau \leq t)$ helps to predict $y_{1,t+h}$ in the sense that

$$\text{Var}(y_{1,t+h}|y_t, y_{t-1}, \ldots) < \text{Var}(y_{1,t+h}|y_t^0, y_{t-1}^0, \ldots),$$

where $y_t^0 = [y_{1,t}', y_{3,t}']'$. Assume that $y_t$ admits a stationary VAR($p$) representation of the form

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t,$$  \hspace{1cm} (1)
where $\varepsilon_t$ is white noise with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t\varepsilon'_t) = \Sigma$. The model can be re-written in companion form as

$$z_t = Az_{t-1} + \nu_t,$$

where

$$z_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad A = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_K & 0 & \cdots & 0 & 0 \\ 0 & I_K & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{bmatrix}, \quad \nu_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$ 

The null hypothesis that $y_{2,t}$ does not cause $y_{1,t}$ at horizon $h$ can be written as

$$H_0 : J_1' \hat{A}^h J_2^* = 0,$$

where $J_1 = [I_{k_1}, 0, \ldots, 0]'$, $J_2^* = e_p \otimes [0, I_{k_2}, 0]'$ and $e_p$ is a $p \times 1$ vector of ones. To test this null hypothesis, Lütkepohl and Burda (1997) suggested a nonlinear Wald test given by

$$W_T = \hat{\pi}' \hat{\Sigma}_\hat{\pi}^{-1} \hat{\pi},$$

where $\hat{\pi} = \text{vec}(J_1' \hat{A}^h J_2^*)$ and $\hat{A}$ is the estimated analog of $A$, where $A_1, \ldots, A_p$ are replaced by least-squares estimators. The covariance matrix

$$\Sigma_{\hat{\pi}} = E[(\hat{\pi} - \pi)(\hat{\pi} - \pi)']$$

with $\pi = \text{vec}(J_1' A^h J_2^*)$ is estimated by employing the Delta method yielding

$$\hat{\Sigma}_{\hat{\pi}} = D(\hat{a}) \hat{\Sigma}_{\hat{a}} D(\hat{a})'$$

where $a = \text{vec}(A_1, \ldots, A_p)$, $\hat{a} = \text{vec}(\hat{A}_1, \ldots, \hat{A}_p)$, $D(\hat{a}) = \partial \pi / \partial a'|_{a=\hat{a}}$ and $\hat{\Sigma}_{\hat{a}}$ is the estimated covariance matrix of $\hat{a}$.

As noted by Lütkepohl and Burda (1997), the estimator of the asymptotic covariance matrix $T \hat{\Sigma}_{\hat{\pi}}$ may converge in probability to a singular matrix as $T \to \infty$. This can happen for particular values of $a$ that imply that the rank of $D(a)$ is smaller than $pk_1, k_3$. In this case, the Wald statistic does no longer have a limiting $\chi^2$ distribution.
Dufour and Renault (2006) propose a simple test procedure based on the
$h$-step ahead autoregression

\[ y_{t+h} = \Pi_1^{(h)} y_t + \cdots + \Pi_p^{(h)} y_{t-p+1} + u_{t+h}, \quad (3) \]

where \( u_{t+h} = \varepsilon_{t+h} + B_1 \varepsilon_{t+h-1} + \cdots + B_{h-1} \varepsilon_{t+1} \) and \( B_j \) are obtained from the MA representation \( y_t = \varepsilon_t + \sum_{j=1}^{\infty} B_j \varepsilon_{t-j} \). The null hypothesis that \( y_{2,t} \) is not a cause of \( y_{1,t} \) at horizon \( h \) implies

\[ H'_0 : \quad J_1' \Pi^{(h)} J_2^* = 0, \quad (4) \]

where

\[ \Pi^{(h)} = \begin{bmatrix}
\Pi_1^{(h)} & \Pi_2^{(h)} & \cdots & \Pi_p^{(h)} \\
I_K & 0 & \cdots & 0 \\
0 & I_K & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & I_K & 0
\end{bmatrix}. \quad (5) \]

\( H'_0 \) is equivalent to \( H_0 \) in the sense that \( H_0 \) is fulfilled if and only if \( H_0 \) holds. To test \( H'_0 \), Dufour et al. (2006) suggest a (linear) Wald test given by

\[ D_T = \tilde{\pi}' \tilde{\Sigma}^{-1} \tilde{\pi}, \quad (6) \]

where \( \tilde{\pi} = \text{vec}(J_1' \hat{\Pi}^{(h)} J_2^*) \) and \( \hat{\Pi}^{(h)} \) results from replacing \( \Pi_1^{(h)}, \ldots, \Pi_p^{(h)} \) in (5) by least-squares estimators obtained from (3). Note that \( \tilde{\pi} - \hat{\pi} \xrightarrow{p} 0 \). Since the errors of the regression (3) have an MA(\( h-1 \)) representation, the usual covariance matrix estimator for the least-squares estimators has to be modified by using the approach suggested by Newey and West (1987), for example.

### 3 Asymptotic properties of OLS based test procedures

In this section the local power of the nonlinear Wald test advocated by LB (1997) and the linear test suggested by DPR (2006) are compared. To facilitate the analysis we focus on the simplest possible case. This basic framework already allows to draw some interesting conclusions.
Consider the three-dimensional VAR(1) model given by

\[
\begin{bmatrix}
  y_{1,t} \\
  y_{2,t} \\
  y_{3,t}
\end{bmatrix}
= \begin{bmatrix}
  0 & 0 & \alpha \\
  0 & 0 & 0 \\
  0 & \beta & 0
\end{bmatrix}
\begin{bmatrix}
  y_{1,t-1} \\
  y_{2,t-1} \\
  y_{3,t-1}
\end{bmatrix}
+ \begin{bmatrix}
  \varepsilon_{1,t} \\
  \varepsilon_{2,t} \\
  \varepsilon_{3,t}
\end{bmatrix}
\tag{7}
\]

\[ y_t = Ay_{t-1} + \varepsilon_t, \tag{8} \]

where \( \varepsilon_t \sim i.i.N(0, \Sigma) \) and \( E(\varepsilon_t \varepsilon'_t) = \Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2) \). The null hypothesis is

\[ H_0: \pi = \alpha \beta = 0. \]

Under this null hypothesis there is no causality at horizons \( h \geq 1 \), whereas for \( \pi \neq 0 \) \( y_{2,t} \) causes \( y_{1,t} \) at \( h = 2 \) but not at \( h \neq 2 \). Following Lütkepohl and Burda (1997) the null hypothesis can be tested by using the nonlinear Wald statistic:

\[ W_T = \frac{T(\hat{\alpha} \hat{\beta})^2}{\beta^2 \hat{\sigma}_\alpha^2 + \hat{\alpha}^2 \hat{\sigma}_\beta^2}. \tag{9} \]

The test is based on the asymptotic distributions of the least-squares estimators

\[ \sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, \sigma_\alpha^2) \tag{10} \]
\[ \sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma_\beta^2), \tag{11} \]

where \( \sigma_\alpha^2 = \sigma_1^2 / (\beta^2 \alpha_2^2 + \sigma_3^2) \) and \( \sigma_\beta^2 = \sigma_3^2 / \sigma_3^2 \). The estimators \( \hat{\sigma}_\alpha^2 \) and \( \hat{\sigma}_\beta^2 \) are consistent estimators of \( \sigma_\alpha^2 \) and \( \sigma_\beta^2 \).

The test procedure proposed by DPR (2006) is based on the least-squares estimator of \( \pi \) in the regression \( y_{1,t} = \pi y_{2,t-2} + u_{1,t} \), where \( u_{1,t} = \varepsilon_{1,t} + \alpha \varepsilon_{3,t-1} \). Employing a robust estimator of the variance of \( \hat{\pi} \) the test statistic results as

\[ D_T = \frac{T(\hat{\pi})^2}{\hat{\sigma}_u^2 / \hat{\sigma}_2^2}, \tag{12} \]

where \( \hat{\sigma}_u^2 \) is a consistent estimator of \( \sigma_u^2 = \sigma_1^2 + \alpha^2 \sigma_3^2 \). Note that in our case \( E(y_{2,t-2}u_t)(y_{2,s-2}u_s) = 0 \) for \( t \neq s \).

To study the asymptotic power of the test procedures we consider three sequences of local alternatives:

- case A: \( \pi_T = \bar{\alpha} \bar{\beta}_T \tag{13} \)
- case B: \( \pi_T = \alpha_T \bar{\beta} \tag{14} \)
- case C: \( \pi_T = \alpha_T \beta_T \tag{15} \)
where $\alpha_T = c_\alpha/\sqrt{T}$, $\beta_T = c_\beta/\sqrt{T}$ and $c_\alpha$, $c_\beta$, $\bar{\alpha}$, $\bar{\beta}$ are some fixed numbers.

The following theorem presents the asymptotic distribution of the two alternative tests under these local alternatives.

**Theorem 1:** Let $y_t = [y_{1,t}, y_{2,t}, y_{3,t}]'$ be generated as in (7), where the parameters $\alpha$ and $\beta$ are replaced by the sequences in (13) – (15). As $T \to \infty$ the limiting distributions of the test statistics $W_T$ and $D_T$ defined in (9) and (12) are given by

**case A:** $W_T \xrightarrow{d} (\lambda_1 + z_1)^2$, $D_T \xrightarrow{d} (\lambda_1^* + z_1^*)^2$

**case B:** $W_T \xrightarrow{d} (\lambda_2 + z_2)^2$, $D_T \xrightarrow{d} (\lambda_2^* + z_2^*)^2$

**case C:** $W_T \xrightarrow{d} \frac{(\lambda_1 + z_1)(\lambda_2 + z_2)}{(\lambda_1 + z_1)^2 + (\lambda_2 + z_2)^2}$, $D_T \xrightarrow{d} z_2^2$,

where $z_j$ and $z_j^*$, $j \in \{1, 2\}$, are standard normally distributed random variables with $E(z_1z_2) = 0$, $E(z_1z_1^*) = 0$ and

$\lambda_1 = \frac{c_\beta \sigma_2}{\hat{\sigma}_3}$, \quad $\lambda_1^* = \frac{c_\beta \sigma_2}{\sqrt{\hat{\sigma}_3^2 + (\sigma_1^2/\bar{\alpha}^2)}}$

$\lambda_2 = \frac{c_\alpha \sqrt{\beta^2 \sigma_2^2 + \sigma_3^2}}{\sigma_1}$, \quad $\lambda_2^* = \frac{c_\alpha \beta \sigma_2}{\sigma_1}$.

**Proof:** Case A: For the nonlinear Wald statistic it follows from (10) and (11) as $T \to \infty$

$W_T = \frac{[\hat{\alpha}c_\beta + \sqrt{T}(\hat{\beta} - \beta_T)]^2}{\hat{\sigma}_2^2 \sigma_3^2 / \sigma_2^2} + o_p(1)$

$= (\lambda_1 + z_1)^2 + o_p(1),$

where $z_1 = \varepsilon_{3t}/\sigma_3 \sim N(0, 1)$. Similarly, we obtain

$D_T = \frac{[\hat{\alpha}c_\beta + \sqrt{T}(\hat{\pi} - \pi_T)]^2}{\sigma_1^2 + \hat{\alpha}^2 \sigma_3^2 / \sigma_2^2} + o_p(1)$

$= (\lambda_1^* + z_1^*)^2$ (16)

where

$\lambda_1^* = \hat{\alpha} \sigma_2^2 \hat{\beta} / (\sigma_1^2 + \hat{\alpha}^2 \sigma_3^2)$

$= c_\beta \sigma_2^2 / (\sigma_3^2 + \sigma_1^2/\hat{\alpha}^2)$

$= c_\beta \sigma_2^2 / (\sigma_3^2 + \sigma_1^2/\bar{\alpha}^2)$

7
and $z_1^* \sim N(0,1)$. Note that $z_1$ and $z_1^*$ are independent as we assume that the errors are uncorrelated Gaussian random variables.

**Case B:** Using (14) as the sequence of local alternatives we have

$$W_T = \frac{[c_\alpha \hat{\beta} + \sqrt{T}(\hat{\alpha} - \alpha_T) \hat{\beta}]}{\beta_2 \sigma_1^2 / (\beta_2 \sigma_2^2 + \sigma_3^2)} + o_p(1)$$

$$= (\lambda_2 + z_2)^2,$$

where $\lambda_2^2 = c_\alpha^2 (\beta_2 \sigma_2^2 + \sigma_3^2) / \sigma_1^2$ and $z_2 \sim N(0,1)$. Furthermore,

$$D_T = \frac{[\hat{\beta} c_\alpha + \sqrt{T}(\pi - \pi_T)]^2}{\sigma_1^2 / \sigma_2^2} + o_p(1)$$

and, therefore, $D_T = (\lambda_2^* + z_2^*)^2$ with $\lambda_2^* = \hat{\beta} c_\alpha \sigma_2 / \sigma_1$.

**Case C:** Under the sequence (15) we obtain

$$W_T = \frac{(\hat{\alpha} \hat{\beta})^2 / \sigma_\alpha^2 \sigma_\beta^2}{(\beta_2 \sigma_2^2 / \sigma_3) + (\hat{\alpha}^2 / \sigma_\alpha^2)}$$

$$= \frac{z_1^2 z_2^2}{z_1^2 + z_2^2} + o_p(1).$$

For the linear test statistic we have

$$D_T = \frac{[\sqrt{T}(\pi - \pi_T)]^2}{\sigma_1^2 / \sigma_2^2} + o_p(1)$$

$$= z_2^2 + o_p(1).$$

**Remark 1:** Under the null hypothesis (i.e. $c_\alpha = 0$ or $c_\beta = 0$) the tests have a $\chi^2(1)$ limiting distribution in cases A and B. In case C only the $D_T$ statistic has a $\chi^2$ limiting distribution, whereas $W_T$ has a nonstandard limiting distribution. Since $z_1^2 / (1 + z_2^2 / z_1^2) < z_1^2$ it follows that using critical values from a $\chi^2(1)$ distribution yields a conservative test procedure.

**Remark 2:** To cope with cases having a non-standard limiting distribution, LB (1997) suggest to set eigenvalues of the estimated analog of the asymptotic covariance matrix $T \hat{\Sigma}_\pi$ equal to zero if they fall below some critical value $c$. In their simulation LB found that a value of $c_T = \sqrt{T}$ performs best. In our case
$T \tilde{\Sigma}$ is a scalar. In the nonstandard case C the estimator of the asymptotic variance is $O_p(T^{-1})$, whereas for the cases A and B the estimator is $O_p(1)$. It follows that a threshold value generated as $c_T = kT^{-\delta}$ with $0 < \delta < 1$ is able to identify case C. According to the simulation results of LB (1997) the “middle value” $\delta = 1/2$ performs best.\footnote{Note that the the modified test statistic based on the generalized inverse suggested by LB (1997) is not applicable in this case.}

**Remark 3:** In cases A and B the asymptotic properties of $W_T$ are identical to the asymptotic properties of the minimum statistic $M_T = \min(t_\alpha^2, t_\beta^2)$, where $t_\alpha$ and $t_\beta$ are the usual t-statistics for a test of $\alpha = 0$ and $\beta = 0$ respectively. Since in case A (or B) $t_\alpha^2$ ($t_\beta^2$) tends to infinity, it follows that the test is asymptotically equivalent to a Wald test of $\alpha = 0$ ($\beta = 0$).

**Remark 4:** It is easy to see that the score statistic and the likelihood ratio statistic are asymptotically equivalent to the Wald statistic. This is due to the fact that $\ell(\hat{\alpha}, \hat{\beta}) = \max \{\ell(0, \hat{\beta}), \ell(\hat{\alpha}, 0)\}$, where $\ell(\cdot)$ denotes the log-likelihood function and $(\hat{\alpha}, \hat{\beta})$ are the restricted estimators subject to the restriction $\alpha \beta = 0$.

**Remark 5:** Since $\lambda_1 \geq \lambda_1^*$ and $\lambda_2 \geq \lambda_2^*$, it follows that the $W_T$ is asymptotically more powerful than $D_T$. However the efficiency loss of $D_T$ depends on the ratio $\hat{\alpha}^2/\sigma_1^2$ (in case A) and $\sigma_3^2/\sigma_1^2$ (in case B) and, therefore, the efficiency loss may be very large or negligible, depending on the parametrization of the process (see also Remark 8).

## 4 Tests based on generalized least-squares

In section 3 it is shown that for the process (7) the DPR test is generally less powerful than the LB test. This may be due to fact that the OLS estimator of the two-step ahead autoregression is inefficient and can be improved by using a generalized least-squares (GLS) approach.

First, it should be noted that the GLS estimator is more efficient only in case A. In cases B and C the error of the 2-step ahead autoregression $u_{1,t} = \varepsilon_{1,t} + \alpha \varepsilon_{3,t}$
is asymptotically uncorrelated if $\alpha$ tends to zero.

There are two possibilities to construct a GLS type estimator. First, the system can be transformed using the lags of the variables. Since

$$y_t = A^2 y_{t-2} + u_t,$$

where $u_t = (1 + AL)\varepsilon_t$, the variable transformation results as $\tilde{y}_t = (1 + AL)^{-1} y_t = y_t - Ay_{t-1} + A^2 y_{t-2}$, where we make use of the fact that in model (7) $A^j = 0$ for $j \geq 3$. The GLS test procedure is obtained as the usual Wald statistic for the hypothesis $\pi = 0$ in the OLS regression $\tilde{y}_{1,t} = \pi \tilde{y}_{2,t-2} + \tilde{\varepsilon}_{1,t}$.

Alternatively, we may apply the transformation

$$y_{1,t}^* = y_{1,t} - \alpha \left( y_{3,t-1} - \beta y_{2,t-2} \right)$$

and compute the usual Wald statistic of $\pi = 0$ from the OLS regression $y_{1,t}^* = \pi y_{2,t-2} + \varepsilon_{1,t}$.

The asymptotic properties of the GLS procedure are presented in Theorem 2: Assume that $y_t = [y_{1,t}, y_{2,t}, y_{3,t}]'$ is generated as in (7), where the parameter $\beta$ is replaced by the sequences (13). Furthermore, let $\tilde{D}_T$ denote the Wald test of $\pi = 0$ based on $\tilde{y}_{1,t} = \pi \tilde{y}_{2,t-2} + \tilde{\varepsilon}_{1,t}$ and $D_T^*$ denotes the Wald test of $\pi = 0$ in the model $y_{1,t}^* = \pi y_{2,t-2} + \varepsilon_{1,t}$. As $T \rightarrow \infty$ the limiting distributions of the test statistics are given by

$$\tilde{D}_T \overset{d}{\rightarrow} (\lambda_3 + z_3)^2,$$

$$D_T^* \overset{d}{\rightarrow} (\lambda_3^* + z_3)^2,$$

where

$$\lambda_3 = \frac{\bar{\alpha} c \beta \sqrt{\sigma_2^2 + \tilde{\alpha} \sigma_3^2}}{\sigma_1}, \quad \lambda_3^* = \frac{\bar{\alpha} c \beta \sigma_2}{\sigma_1^2},$$

and $z_3 \sim N(0,1)$.

**Proof:** For the process (7) we have

$$y_{1,t} = \varepsilon_{1,t} + \tilde{\alpha} \varepsilon_{3,t-1} + \pi T \varepsilon_{2,t-2}$$

$$\tilde{y}_{1,t} = \varepsilon_{1,t} + \pi T \varepsilon_{2,t-2}$$

$$\tilde{y}_{2,t} = \varepsilon_{2,t} - \tilde{\alpha} \varepsilon_{3,t-1} ,$$

10
where \( \pi_T = \bar{\alpha}c_\beta / \sqrt{T} \). It follows that

\[
\tilde{D}_T = \left[ \bar{\alpha}c_\beta + \sqrt{T} (\bar{\pi}_T - \pi_T) \right]^2 + o_p(1)
\]

where \( \bar{\pi}_T \) denotes the least-square estimator of \( \pi_T \) obtained from an OLS regression of \( \bar{y}_{1,t} \) on \( \bar{y}_{2,t-2} \). Similarly, we obtain

\[
y_{1,t}^* = \varepsilon_{1,t} + \pi_T \varepsilon_{2,t-2} \quad \text{and} \quad D_T^* = \left[ \bar{\alpha}c_\beta + \sqrt{T} (\bar{\pi}_T^* - \pi_T) \right]^2 + o_p(1)
\]

where \( \bar{\pi}_T^* \) denotes the least-square estimator of \( \pi_T \) obtained from an OLS regression of \( y_{1,t}^* \) on \( y_{2,t-2} \).

\[\blacksquare\]

**Remark 6:** It turns out that the test \( \tilde{D}_T \) is more powerful than \( D_T^* \). This is not surprising as the \( \tilde{D}_T \) is asymptotically equivalent to the Wald test based on the ML estimator of \( \pi_T \).

**Remark 7:** Comparing the noncentrality parameters \( \lambda_3 \) and \( \lambda_3^* \) to the noncentrality parameter of the OLS based test statistic \( \lambda_1^* \) it is obvious that the GLS versions have a larger local power relative to the OLS version of the test.

**Remark 8:** It is surprising to learn that the GLS procedure may be even more powerful than the nonlinear Wald (LB) test. Comparing \( \lambda_3 \) and \( \lambda_3^* \) with \( \lambda_1^* \) shows that both GLS tests are asymptotically more powerful if \(|\bar{\alpha}| > \sigma_1/\sigma_2\). The reason is that according to Remark 3, the LB test is asymptotically equivalent to a (squared) \( t \)-statistic for the hypothesis \( \beta = 0 \). Thus, this test does not involve the parameter \( \alpha \). On the other hand the test \( \tilde{D}_T \) is used to test the hypothesis \( \pi_T = \bar{\alpha}c_\beta / \sqrt{T} = 0 \) and, therefore, the power of the test increases with \(|\bar{\alpha}|\).

In order to investigate the small sample properties of alternative test procedures we generate data according to the three-variate process (7), where \( \varepsilon_t \overset{i.i.d.}{\sim} N(0, I_3) \). If \( \pi = \alpha \beta = 0 \), then \( y_{2,t} \) is not a cause of \( y_{1,t} \) at horizon \( h = 2 \). The nonlinear Wald statistic is denoted by “LB” and the OLS based linear test suggested by Dufour et al. (2006) is labelled as “DPR”. The (infeasible) GLS test
statistic is based on the regression
\[ \tilde{y}_{1,t} = (y_{1,t} - \alpha y_{3,t-1} + \alpha \beta y_{2,t-2}) = \pi \tilde{y}_{2,t-2} + \tilde{\varepsilon}_{1,t}, \]
where \( \tilde{y}_{2,t} = y_{2,t} \) and the true values of \( \alpha \) and \( \beta \) are used. The feasible version of the test is obtained by replacing \( \alpha \) and \( \beta \) by the respective least-squares estimators. By noting that
\[ \hat{\alpha} y_{3,t-1} - \hat{\alpha} \hat{\beta} y_{2,t-2} = \hat{\alpha} (\beta y_{2,t-2} + \varepsilon_{3,t} - \hat{\beta} y_{2,t-2}) = \hat{\alpha} \varepsilon_{3,t} - \hat{\alpha} (\beta - \beta) y_{2,t-2}, \quad (18) \]
it can be shown that if \( \alpha = 0 \)
\[ T^{-1/2} \sum_{t=3}^{T} (y_{1,t} - \hat{\alpha} y_{3,t-1} + \hat{\alpha} \hat{\beta} y_{2,t-2}) y_{2,t-2} \]
\[ = T^{-1/2} \sum_{t=3}^{T} y_{1,t} y_{2,t-2} + o_p(1) \]
and, therefore, the feasible GLS statistic (FGLS) is asymptotically equivalent to the infeasible GLS statistic. However, if \( \alpha \neq 1 \), then (18) is \( O_p(1) \) and the difference between the GLS and FGLS statistics does not tend to zero. Thus the FGLS statistic is no longer \( \chi^2 \) distributed.

The upper panel of Table 1 presents the empirical rejection rates for \( T = 200 \) and 10,000 Monte Carlo replications. According to Remark A, the LB test is conservative if \( \alpha \beta = 0 \). This is confirmed by the first row of Table 1, showing that the empirical rejection rate for this test is equal to zero. If one of the two parameters \( \alpha \) and \( \beta \) is different from zero, the LB test still tends to underreject but the size bias is fairly small and tends to zero if the parameter is large. The empirical sizes for the other tests are close to the nominal value of 0.05. However, if \( \alpha \) is large, the FGLS statistic possesses a substantial size bias which was predicted by the theoretical considerations presented above.

The lower panel of Table 1 reports the empirical power of the test procedures. If all variances of the innovations are equal to one, the LB test outperforms all other tests. The GLS and FGLS test procedures yield a slight improvement compared to the OLS based DPR test procedure. However, if the variance \( \sigma^2_1 \) becomes small and \( \alpha \) is large, then the GLS procedure can be more powerful than
Table 1: Size and power of alternative test procedures

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>LB</th>
<th>DPR</th>
<th>GLS</th>
<th>FGLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.000</td>
<td>0.050</td>
<td>0.053</td>
<td>0.053</td>
</tr>
<tr>
<td>0.0</td>
<td>0.5</td>
<td>0.044</td>
<td>0.049</td>
<td>0.050</td>
<td>0.049</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>0.049</td>
<td>0.049</td>
<td>0.048</td>
<td>0.048</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0</td>
<td>0.045</td>
<td>0.054</td>
<td>0.049</td>
<td>0.082</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0</td>
<td>0.049</td>
<td>0.054</td>
<td>0.051</td>
<td>0.168</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5</td>
<td>0.861</td>
<td>0.289</td>
<td>0.292</td>
<td>0.297</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2</td>
<td>0.774</td>
<td>0.252</td>
<td>0.294</td>
<td>0.320</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
<td>0.728</td>
<td>0.139</td>
<td>0.149</td>
<td>0.158</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2</td>
<td>0.795</td>
<td>0.494</td>
<td>0.751</td>
<td>0.692</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2</td>
<td>0.799</td>
<td>0.664</td>
<td>0.992</td>
<td>0.904</td>
</tr>
</tbody>
</table>

Note: The nominal size is 0.05 and $T = 200$. If not indicated otherwise, the error variances are set equal to unity.

the LB test. Indeed this is what we expect from Theorem 2 as the Pitman drift $\lambda_3$ is positively related to $\alpha$ and inversely related to $\sigma_1$.

5 Testing for long-run causality

In this section we consider a tests for the infinite horizon $h \to \infty$. Since for a stationary variable $y_{t,1}$ the conditional variance is equal to

$$\lim_{h \to \infty} Var(y_{t,1+h}|y_t, y_{t-1}, \ldots) = \lim_{h \to \infty} Var(y_{t,1+h}|y_t^0, y_{t-1}, \ldots) = Var(y_t)$$

it is not possible to test for long-run causality within a stationary framework. Therefore, in this section we will allow the autoregressive lag-polynomial to have $k-r$ ($0 \leq r < k$) unit roots. If $r < k$ we say that $y_t$ is cointegrated, which may include 'trivial' cointegration vectors of the form $[0,0,1]$, i.e., the variable $y_{3,t}$ may be stationary. To simplify the exposition we will (again) focus on the VAR(1) model.

To derive the restrictions from the AR representation of the process we start from the VAR representation of the system that is obtained the error correc-
tion representation of a cointegrated system by adding $y_{t-1}$ to both sides of the equation:

$$y_t = (\alpha \beta' + I_k) y_{t-1} + \varepsilon_t$$

$$= Ay_{t-1} + \varepsilon_t,$$

where $\alpha$ and $\beta$ are $k \times r$ matrices with full column rank. The columns of the matrix $\beta$ represent the $r$ linear combinations defining the cointegration relationships of the system. Consider the Jordan decomposition $A = V \Lambda V^{-1}$, where $\Lambda$ is a diagonal matrix of the (ordered) eigenvalues and $V$ is the matrix of orthonormal eigenvectors. The infinite horizon forecast results as

$$\lim_{h \to \infty} E(y_{t+h}|y_t) = V^{-1} \Lambda V y_t = \left( \sum_{j=1}^{r} w_j v_j' \right) y_t,$$

where $\Lambda = \lim_{h \to \infty} \Lambda^h$ is a diagonal matrix with $k - r$ ones and $r$ zeros on the leading diagonal and $v_j$ ($w_j$) is the $j$'th column of $V$ ($V^{-1}$). It follows that the variable $y_{2,t}$ is a long-run cause for $y_{1,t}$ if $w_{1,j}v_{2,j} = 0$, where $v_{i,j}$ ($w_{i,j}$) is the $i$'th element of $v_j$ ($w_j$). Note that for $j = k - r + 1, \ldots, k$ the eigenvectors $v_j$ lie inside the cointegration space spanned by the cointegration vectors $\beta_1, \ldots, \beta_r$, or

$$v_j \in \text{sp}(\beta) \text{ for } j = k - r + 1, \ldots, k,$$  \hspace{1cm} (19)

whereas the remaining eigenvectors are related to the common trends, i.e., $v_j y_t \sim I(1)$ for $j = 1, \ldots, k - r$. Pre-multiplying (19) by the orthogonal complement $\alpha'_\perp$ with $\alpha'_\perp \alpha = 0$ yields $\alpha'_\perp y_t = \alpha'_\perp y_{t-1} + \alpha'_\perp \varepsilon_t$ and, therefore,

$$v_j \in \text{sp}(\alpha'_\perp) \text{ for } j = 1, \ldots, k - r.$$  \hspace{1cm} (20)

Hence, we can write $V^{-1} = \Psi[\alpha'_\perp, \beta]'$, where $\Psi$ is a block diagonal matrix that yields an appropriate normalization of $\alpha'_\perp$ and $\beta'_\perp$. It is easy to verify that

$$V = [\beta'_\perp, \alpha'] \text{ and } \Psi = \text{diag}\{(\alpha'_\perp \beta'_\perp)^{-1}, (\beta' \alpha)^{-1}\}$$

and, therefore,

$$y_t = V \Lambda V^{-1} y_{t-1} + \varepsilon_t$$

$$= \beta'_\perp (\alpha'_\perp \beta'_\perp)^{-1} \alpha'_\perp y_{t-1} + \alpha \Lambda_2 (\beta' \alpha)^{-1} \beta' y_{t-1} + \varepsilon_t.$$
where $\Lambda_2$ is a $r \times r$ diagonal matrix with the stable eigenvalues on the leading diagonal. It follows that if $y_2$ is not a long-run cause of $y_1$, then the (2,1)-element of the matrix $C = \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp$ is equal to zero. Bruneau and Jondeaux (1999) and Yamamoto and Kurozumi (2006) suggested a Wald test statistic of the hypothesis $c_{12} = 0$. Since the matrix $C$ is a complicated nonlinear function of the VAR parameters, the Wald statistic is difficult to compute. We therefore suggest a much simpler test procedure.

From the Beveridge-Nelson decomposition we know that

$y_t = C \left( \sum_{j=1}^{t} \varepsilon_j \right) + O_p(1)$.  

Let $v_{2,t} = \varepsilon_{2,t} - E(\varepsilon_{2,t}|\varepsilon_{1,t}, \varepsilon_{3,t})$, $\sigma^2_{v_2} = \text{var}(v_{2,t})$ and define the respective partial sum process as $S_t = v_{2,1} + \cdots + v_{2,t}$. It follows that

$$\omega_{1v} = \lim_{T \to \infty} E \left( \frac{1}{T} y_{1,T} S_T \right) = c_{12} \sigma^2_{v_2}.$$  \hspace{1cm} (21)

Accordingly, $c_{12}$ is proportional to the long-run covariance between $\Delta y_{1,t}$ and $v_{2,t}$. It is interesting to note that the impulse response of $\Delta y_{1,t+h}$ with respect to the orthogonalized shock $v_{2,t}$ is given by $\phi_h = E(\Delta y_{1,t+h} v_{2,t}) = \text{and, therefore}$

$$\lim_{h \to \infty} E(y_{1,t+h} v_{2,t}) = \lim_{h \to \infty} \left[ \sum_{j=1}^{h} E(\Delta y_{1,t+j} v_{2,t}) \right] = \sum_{j=0}^{\infty} \phi_j.$$

It follows that $\omega_{1v}$ measures the limit of the cumulated impulse response with respect to the shock $v_{2,t}$.

To derive an estimator for the long-run covariance let us first assume that $v_{2,t}$ is observed. In this case, a natural (Newey-West type) estimator is

$$\hat{\omega}_{1v}(\ell) = \frac{1}{(\ell + 1)T} \sum_{t=1}^{T+\ell} (\Delta y_{1,t} + \cdots + \Delta y_{1,t-\ell})(v_{2,t} + \cdots + v_{2,t-\ell})$$,  \hspace{1cm} (22)

where $\Delta y_{1,s} = 0$ and $v_{2,s} = 0$ for $s < 1$ and $s > T$. In the following lemma it is shown that $\hat{\omega}_{1v}(\ell)$ is identical to the usual Newey-West (1987) estimator for the long-run covariance.
Lemma 1: Let \( x_t \) and \( y_t \) be sequences observed for \( t = 1, \ldots, T \) and set \( x_s = 0 \) and \( y_s = 0 \) for \( s < 1 \) and \( s > T \). Then,

\[
\hat{\omega}_{xy} = \frac{1}{(\ell + 1)T} \sum_{t=1}^{T+\ell} (x_t + \cdots + x_{t-\ell})(y_t + \cdots + y_{t-\ell}) = \sum_{j=-\ell}^{\ell} w_{|j|} \hat{\gamma}_j
\]

where \( w_j = (\ell + 1 - j)/(\ell + 1) \), \( \hat{\gamma}_j = T^{-1} \sum_{t=j+1}^{T} x_{t-j} y_t \) for \( j \geq 0 \) and \( \hat{\gamma}_j = T^{-1} \sum_{t=|j|+1}^{T} x_t y_{t-|j|} \) for \( j < 0 \).

Proof: Using the lag operator \( L \) we write

\[
\hat{\omega}_{xy} = \frac{1}{(\ell + 1)T} \sum_{t=1}^{T+\ell} [(1 + L + \cdots + L^\ell)x_t][(1 + L + \cdots + L^\ell)y_t]
\]

\[
= \frac{1}{(\ell + 1)T} \sum_{t=1}^{T+\ell} x_t[L^{-\ell}(1 + L + \cdots + L^\ell)^2]y_t
\]

\[
= \frac{1}{(\ell + 1)T} \sum_{t=1}^{T+\ell} x_t(y_{t-\ell} + 2y_{t-\ell+1} + \cdots + (\ell + 1)y_t + \ell y_{t+1} + \cdots + y_{t+\ell})
\]

\[
= \sum_{j=-\ell}^{\ell} w_{|j|} \hat{\gamma}_j .
\]

To obtain a simple test statistic for the hypothesis \( \omega_{1v} = 0 \) in the case that \( v_{2,t} \) is estimated by the residual of a regression \( \hat{\varepsilon}_{2,t} \) on \( \hat{\varepsilon}_{1,t} \) and \( \hat{\varepsilon}_{3,t} \), where \( \hat{\varepsilon}_{i,t} \) is the residual of the \( i \)th VAR equation, we consider the following regression equation:

\[
y_{2,t} = a_1 y_{1,t} + a_3 y_{3,t} + \hat{\beta}' y_{t-1} + \theta_{\ell} z_{t+1} + \tilde{v}_{2,t} \tag{23}
\]

where \( \hat{\beta} \) is a \( T \)-consistent estimator of \( \beta \) and

\[
z_{t+1}^\ell = (\ell y_{1,t+1} + (\ell - 1)y_{1,t+2} + \cdots + 2y_{1,t+\ell-1} + y_{1,t+\ell}, \quad t = 1, \ldots, T-1 \tag{24}
\]

with \( y_{1,s} = 0 \) for \( s > T \). Note that we leave out the lags in the construction of the variable \( z_{t+1}^\ell \) since \( E(y_{1,s} v_{2,t}) = 0 \) for \( s = t, t-1, \ldots \) by construction. In the
following theorem we show that if $y_{2,t}$ is not a long-run cause of $y_{1,t}$, then the $t$-statistic of the hypothesis $\theta_\ell = 0$ has a standard normal limiting distribution.

**Theorem 3:** Let $y_t = [y_{1,t}, y_{2,t}, y_{3,t}]'$ be generated by a VAR(1), where $\varepsilon_t$ is i.i.d. with $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon'_t) = \Sigma$ (pos. def.) and $E(\varepsilon_{1,t}^4) < \infty$ for all $i, t$. 1 $\leq r \leq k$ eigenvalues of the matrix $A$ are equal to one, whereas all other eigenvalues are inside the complex unit circle. If $y_{2,t}$ is not a long-run cause for $y_{1,t}$, then as $T \to \infty$, $\ell \to \infty$ and $T/\ell^3 \to 0$ we have

$$t_{\theta_\ell} \overset{d}{\to} N(0, 1)$$

where $t_{\theta_\ell}$ denotes the $t$-statistic for the hypothesis $\theta_\ell = 0$ in the regression (??).

**Proof:** Since $\hat{\beta}$ is a super-consistent estimator for $\beta$ we assume that $\beta$ is given. Let $\xi_t = [y_{1,t}, y_{3,t}, y'_{t-1}\beta]'$ such that

$$\hat{\theta}_\ell = A_T / B_T$$

and

$$A_T = \sum_{t=2}^{T-1} v_{2,t} z_{t+1} - \sum_{t=2}^{T-1} v_{2,t} \xi_t \left( \sum_{t=2}^{T-1} \xi_t \xi_t' \right)^{-1} \left( \sum_{t=2}^{T-1} \xi_t z_{t+1} \right)$$

$$B_T = \sum_{t=2}^{T-1} (z_{t+1} \ell) - \sum_{t=2}^{T-1} z_{t+1} \xi_t' \left( \sum_{t=2}^{T-1} \xi_t \xi_t' \right)^{-1} \left( \sum_{t=2}^{T-1} \xi_t z_{t+1} \right).$$

As shown by Priestley (1981, p. 469) we have as $T/\ell^3 \to 0$

$$\frac{1}{\sqrt{T}} \sum_{t=2}^{T-1} v_{2,t} z^\ell_{t+1} \overset{d}{\to} N(0, \sigma^2_{v_2} s^2_\varepsilon),$$

where $\sigma^2_{v_2} = \text{var}(v_{2,t})$ and $s^2_\varepsilon = \text{plim} T^{-1} \sum_{t=1}^{T} (z^\ell_{t+1})^2$. Furthermore,

$$\frac{1}{\sqrt{T}} \sum_{t=2}^{T-1} v_{2,t} \xi_t \overset{d}{\to} N(0, \sigma^2_{v_2} S_{\xi \xi})$$

$$\frac{1}{T} \sum_{t=2}^{T-1} z^\ell_{t+1} \xi_t \overset{p}{\to} s_{z\xi}$$

17
where $S_{ξξ} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} ξ_t ξ_t'$ and, thus,

$$T^{-1/2} A_T \xrightarrow{d} N(0, \sigma^2 v_2 (s_z^2 - s_z' S_{ξξ}^{-1} s_z')) .$$

Since $\lim_{T \to \infty} T^{-1} B_T = s_z^2 - s_z' S_{ξξ}^{-1} s_z$ it follows that

$$t_{θ_ℓ} = \frac{A_T}{\hat{σ}_v \sqrt{B_T}}$$

has a standard normal limiting distribution. ■

**Remark 10:** For the ease of exposition we focused on the three-variate VAR(1) model. In the more general case $p \geq 1$ and $k \geq 3$ variables the test regression is

$$y_{2,t} = \sum_{i \neq 2} a_i y_{i,t} + \sum_{i=1}^{p} \sum_{j=1}^{3} b_{ij} y_{i,t-j} + θ_ℓ z_{t+1}^θ + v_{2,t}$$

for $t = p + 1, \ldots, T - ℓ$. The variable $z_{t+1}^θ$ is constructed as in (24). It is straightforward to show that the result in Theorem 3 continues to hold in the more general case.

**Remark 11:** The test equation (??) is based on the error correction representation. Alternatively, the test may be based on the unrestricted VAR

$$\Delta y_{2,t} = a_1 \Delta y_{1,t} + a_2 \Delta y_{2,t} + b' y_{t-1} + θ_ℓ z_{t+1}^θ + v_{2,t}.$$  \hspace{1cm} (26)

Although, the estimation of this representation is less efficient than the error correction representation, it is not difficult to show that the asymptotic properties of the test based on (??) are identical to the test based on (23).

**Example.** To investigate the small sample properties of the test we consider a simple example. Let

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} 0 & a & b \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} ε_{1,t} \\ ε_{2,t} \\ ε_{3,t} \end{bmatrix}$$

$$y_t = Ay_{t-1} + ε_t$$
### Table 1: Size and power of the long-run causality test

<table>
<thead>
<tr>
<th>a</th>
<th>$\ell = 4$</th>
<th>$\ell = 8$</th>
<th>$\ell = 12$</th>
<th>$\ell = 16$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0520</td>
<td>0.0504</td>
<td>0.0546</td>
<td>0.0529</td>
</tr>
<tr>
<td>0.1</td>
<td>0.3538</td>
<td>0.2211</td>
<td>0.1781</td>
<td>0.1486</td>
</tr>
<tr>
<td>0.2</td>
<td>0.8655</td>
<td>0.6157</td>
<td>0.4430</td>
<td>0.3461</td>
</tr>
<tr>
<td>0.3</td>
<td>0.9952</td>
<td>0.9044</td>
<td>0.7520</td>
<td>0.6028</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9999</td>
<td>0.9897</td>
<td>0.9205</td>
<td>0.8174</td>
</tr>
<tr>
<td>0.5</td>
<td>1.0000</td>
<td>0.9994</td>
<td>0.9834</td>
<td>0.9335</td>
</tr>
</tbody>
</table>

**Note:** The entries report the rejection frequencies of 10,000 replications of model (27). The nonlinear Wald test (LB) is computed by using a numerical derivative and the DPR test employs a rectangular kernel with truncation lag $h - 1$. The nominal size is 0.05 and $T = 200$. The entries report the rejection frequencies of 10,000 replications of model (27). The nonlinear Wald test (LB) is computed by using a numerical derivative and the DPR test employs a rectangular kernel with truncation lag $h - 1$. The nominal size is 0.05 and $T = 200$.

with error correction representation

$$\Delta y_t = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & -a & -b \end{bmatrix} y_{t-1} + \varepsilon_t$$

$$\beta'_\perp = \begin{bmatrix} b & 0 & 1 \\ 0 & b & -a \end{bmatrix}$$ and $$\alpha'_\perp = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and $A^j = A$ for $j = 1, 2, \ldots$. In this model, the (1, 2) element of the matrix $\beta'_\perp (\alpha'_\perp \beta'_\perp)^{-1} \alpha'_\perp$ is equal to $a$ and, therefore, $y_{2,t}$ is not a long-run cause of $y_{1,t}$ whenever $a = 0$.

To investigate the small sample properties of the test we generate data according to (27), where we let $b = 1$. If $a = 0$, then $y_{2,t}$ is not a long-run cause of $y_{1,t}$. The sample size is $T = 200$ and 10,000 replications are used to compute the rejection rates presented in Table 3. It turns out that the power of the test depends sensitively on the truncation lag $\ell$ and, therefore, rules for the truncation lag should be preferred that tend to suggest small numbers of $\ell$. For example the rule $k = \text{int}\{4(T/100)^2/9\}$ proposed by Newey-West renders $k = 4$ which performs best in our simulations.

### 6 Conclusions

In this paper we compare the asymptotic properties of multistep causality tests suggested by LB (1997) and variants of the test suggested by DPR (2006). We
show that under a sequence of local alternatives the former test is uniformly more powerful than the computationally simpler test suggested the latter paper. Furthermore we derive the (nonstandard) limiting distribution of the test in cases where the classical Wald tests fails. Our theoretical results imply that the power gap of the original DPR test depends on the relative variances of the VAR innovations. Therefore, the original DPR test should not be used in cases where the error variance of the causing variable is small relative to the other error variances.

The main reason of the (sometimes dramatic) loss of power of the DPR test is the use of the inefficient OLS estimator for estimating the $h$-step ahead autoregression. Our theoretical analysis suggest that the GLS variants of the tests may even be more powerful than the LB test. Therefore, a substantial improvement of the test procedure can be achieved by using more efficient estimators based on a GLS regression. An important problem is, however, that in general the limiting distribution of the feasible GLS estimator is different from a $\chi^2$ distribution and depends on the parameters of the system. Therefore, bootstrap or subsampling techniques have to be employed to obtain valid critical values of the feasible GLS procedure.

Finally we show that a simple test for long-run causality (i.e. for causality at horizon $h \to \infty$) can be constructed based on a semi-parametric (Newey-West type) estimator of the long-run covariance between the first variable and the orthogonalized innovation of the second variable. This approach sidestep the difficulties arising from the fact that the null hypothesis implies a highly nonlinear restriction on the VAR parameters in the system. In contrast, our proposed test statistic can be computed from a simple least squares regression.

References


Bruneau, C. and E. Jondeau (1999), Long-run Causality, with Application to International Links between Long-term Interest Rates, Oxford Bulletin


