Persistence-Robust Granger Causality Testing

Dietmar Bauer
Arsenal Research
Vienna, Austria

Alex Maynard*
Department of Economics
University of Guelph, ON, Canada

October 16, 2008

Abstract

The observed persistence common in economic time series may arise from a variety of models that are not always distinguished with confidence in practice, yet play an important role in model specification and second stage inference procedures. Previous literature has introduced causality tests with conventional limiting distributions in I(0)/I(1)VAR models with unknown integration orders, based on an additional surplus lag in the specification of the estimated equation, which is not included in the tests. Building on this approach, but using an infinite order VARX framework, we provide a highly persistence-robust Granger causality test that accommodates i.a. stationary, nonstationary, local-to-unity, long-memory, and certain (unmodelled) structural break processes in the forcing variables within the context of a single χ^2 null limiting distribution. No first-stage testing or estimation is required and known lag orders are not assumed.
JEL Classification: C12, C32 *Keywords*: Granger causality, surplus lag, nonstationary VAR, local-to-unity, long-memory

---

*Corresponding author. Previously entitled "Robust Granger Causality Tests in the VARX Framework." We thank Zahid Asghar, Richard Baillie, Lynda Khalaf, J.M. Dufour, David Hendry, Peter Phillips, and participants at the Singapore Management University Conference in Honour of Peter C. B. Phillips, the 2006 NBER-NSF Time Series Conference, the 2nd International Workshop on Computational and Financial Econometrics, and the Canadian Econometric Study Group for useful discussion. This work was started while the authors were visiting the Cowles Foundation and we gratefully acknowledge their hospitality. The postdoc position of Bauer at the Cowles Foundation was financed by the Max Kade Foundation which is gratefully acknowledged. Maynard thanks the SSHRC for research funding.

1 Introduction

Since its introduction in Granger (1969), tests of Granger noncausality have become ubiquitous in economics, with applications ranging from the causal relation between money and output (Friedman and Kuttner, 1992) to the export led growth hypothesis (Marin, 1992). This paper develops a simple but flexible augmented VARX approach to Granger causality testing that is highly robust to the degree and nature of the persistence in the causing variables. The proposed causality test[1] may be employed regardless of whether the data generating process for the causal variable is characterized by stationarity, long-memory/fractional integration, a local-to-unity process, I(1) behavior, or breaks in the mean of the process.[2] Consequently no prior knowledge, pre-estimation, or pre-test is required. Likewise, known lag orders are not assumed.

These are desirable characteristics for several reasons. Frequently it is difficult to determine the degree and nature of the persistence of the forcing variables with full confidence. This can often matter in both theory and practice for second stage model specification and inference. Likewise, recent developments in the cointegration literature have also stressed the importance of allowing for fractional integration,[3] and near unit roots (Jansson and Moreira, 2006). In fact, two of the most recent studies (Phillips, 2005; Muller and Watson, 2007), emphasize agnostic approaches to the form of this persistence, motivated by robustness concerns similar in spirit to ours.

The practical difficulties associated with distinguishing $I(1)$ and $I(0)$ processes are well known. Moreover, processes with near unit roots may often be better modelled as local-to-unity (Phillips, 1987; Chan, 1988), against which unit root tests are inconsistent by design. This choice is further complicated by the possibility of structural breaks, which can in turn be difficult to distinguish from long-memory processes (Diebold and Inoue, 2001; Gourieroux and Jasiak, 2001; Granger and Hyung, 2004). Thus it can often be difficult to determine with confidence the correct model for persistent data. As Phillips (2003, p. C35) puts it "no one really understands trends, even though most of us see trends when we look at economic data."

These distinctions are important to model specification. A typical VAR may be specified in levels if unit roots are rejected, in first-differences if the variables are

[1]We address only Granger's version of causality, despite the importance of several other definitions.

[2]Formal results for the structural break model, omitted here to save space, can be found in our working paper version, available at www.amaynard.ca/papers/bm.pdf.

[3]There is now a very rich literature on fractional cointegration. See, for example, Hualde and Robinson (2007) and the many references therein.

individually integrated but not cointegrated, and in error-correction format if the variables are cointegrated. Likewise, structural breaks require explicit modelling and long-memory processes are not easily accommodated in a VAR setting. Such choices can have important practical implications. A recent example involves the role of technology shocks in macroeconomic models, for which VARs in productivity and differenced hours worked support models with frictions, whereas including hours in levels supports traditional real business cycle models (Christiano *et al.*, 2003).

The persistence of the causing variable also matters for second-stage inference. Even in simple regression models, different critical values may apply depending on whether or not the regressor contains a unit root. Moreover, both stationary and unit root asymptotics can be misleading in the presence of near unit roots (Elliott, 1998). Such inference problems may be further complicated once one allows for structural breaks or long-memory. They have also been found to matter in practical applications, such as tests of stock return predictability (Stambaugh, 1999).

Our approach builds on a rich literature, originating in the work of Park and Phillips (1989) and Sims *et al.* (1990). Their results imply that, despite the nonstandard asymptotics in I(1) systems, parameters that may be expressed as coefficients on stationary regressors retain a standard root-T normal asymptotic distribution. Similar results also hold in cointegrating systems involving nonstationary fractional integration (Dolado and Marmol, 2004). The surplus lag approach uses this result to simplify inference. In the context of unit root testing, Choi (1993) recognized that, with the addition of an extra, unnecessary lag, the autoregressive model could be rewritten so that all the parameters of interest are expressed as coefficients on stationary transformations of the data. Thus, at some cost, in terms of efficiency, inference procedures could be simplified, via the avoidance of nonstandard distributions. Toda and Yamamoto (1995), and Dolado and Lütkepohl (1996) showed how the same surplus lag approach could be applied to provide inference in finite order vector autoregression, without knowing which components are stationary and which have unit roots. Saikkonen and Lütkepohl (1996) extended these results to infinite order VARs.

This approach is very flexible with respect to inference in general I(0)/I(1) and cointegrated models. On the other hand, the pure VAR framework adopted in these surplus lag methods cannot accommodate long-memory, nor can it accommodate breaks unless they are explicitly modelled. This detracts somewhat from the advantageous features of the surplus lag approach. By incorporating an exogenously modelled component, we may accommodate a richer class of persistent processes for the forcing variable in the VARX framework, including those with long-memory/fractionally integration or unmodelled structural breaks. Moreover, we find that with the incorporation of the surplus lag, the null limit distribution continues to be unaffected by the particular form of this persistence. Likewise, these results are not dependent on knowledge of the correct lag orders. In all cases, we allow for infinite lag orders under the null hypothesis, approximated by finite order models whose lag lengths increase with sample size. Thus our results also build on the literature on reasonable approximability (Berk, 1974; Lewis and Reinsel, 1985; Lütkepohl and Saikkonen, 1997) and provide some extensions to allow for exogenous regressors, including those with long-memory. Some related extensions are provided by Poskitt (2007), who establishes autoregressive approximations to (univariate) non-invertible and stationary long-memory processes.

The simplicity and generality of the surplus lag approach does not come without cost. Naturally, the addition of an extra unnecessary lag reduces efficiency and power relative to a correctly specified model. However, as previous literature reports, the magnitude of this effects varies considerably. Power losses are greatest in cointegration tests, in which consistency against $O(T^{-1})$ alternatives is lost. Generally, the surplus lag is not recommended in this case.[4] However, efficiency losses are often far more moderate in the Granger causality tests considered here, particularly when the baseline model already includes a number of lags, as in common macroeconomic applications.

In the I(0)/I(1) context there arguably exist alternative methods that are as general, but more efficient, than existing results for the VAR-based surplus lag method. When the number of cointegrating vectors is known, error correction models provide a natural context for efficient causality testing, although in practice pre-tests are required (Toda and Phillips, 1993). The fully modified (FM) VAR estimation (Phillips, 1995; Kitamura and Phillips, 1997) is also efficient, without requiring a priori knowledge on the number of $I(0)$ and unit root components.[5] Nevertheless, most efficient tests designed for the $I(0)/I(1)$ case require adjustment in the presence of either near unit roots (Elliott, 1998) or fractional integration, whereas we show that, in the VARX context, the same surplus lag test continues to work without adjustment in both cases.

A second limitation of our approach is that we allow for long-memory in the forcing processes but not the error process for the dependent variables. The difficulty of

[4]The surplus lag approach may also be unsuited to applications, such as forecasts for the persistent variable itself, in which explicit modelling of the low-frequency behavior is unavoidable.

[5]Kim and Phillips (2004) extend FM regression, but not causality tests, to fractional cointegration.

weakening this assumption for time domain estimators is discussed in (Hidalgo, 2000; Hidalgo, 2005), who provides frequency based non-parametric causality tests, which allow for long-memory in both. On the other hand, these tests require covariance stationarity, ruling out many of the interesting cases considered here.

The remainder of the paper is organized as follows. Section 2 presents the model, Section 3 presents the large sample results, and Section 4 provides some brief simulations. Appendix A collects some technical results and the proofs of the main theorems are provided in Appendix B. The tables are included at the back of the paper. Formal results for the structural break case and additional simulations, omitted to save space, are included in the working paper version available at www.amaynard.ca/papers/bm.pdf.

2 The model

We consider tests of the null hypothesis that z_{1t} $(k_{z1} \times 1)$ does not Granger cause y_t $(k_y \times 1)$ after controlling for z_{2t}[6] $(k_{z2} \times 1)$. Let $\mathcal{F}_{t,y,z1,z2}$ and $\mathcal{F}_{t,y,z2}$ define the information sets generated by $\left\{(y'_{t-j}, z'_{1t-j}, z'_{2t-j})', j \geq 0\right\}$ and $\left\{(y'_{t-j}, z'_{2t-j})', j \geq 0\right\}$, respectively. Then we test the Granger noncausality condition

$$\mathbb{E}\left[y_t | \mathcal{F}_{t-1,y,z1,z2}\right] = \mathbb{E}\left[y_t | \mathcal{F}_{t-1,y,z2}\right]. \tag{1}$$

In practice this hypothesis is often tested by means of parameter restrictions on a joint VAR involving all three variables. However, our interest includes cases in which the forcing variable z_{1t} exhibits long-memory or structural breaks. In this context, the pure VAR may be too restrictive. The forcing variable z_{1t} is instead exogenously modelled, allowing for a number of alternative models for z_{1t} (see Section3).[7]

Under the null hypothesis the true joint DGP for $w_t := [y'_t, z'_{2t}]'$ will be assumed to be approximable by a VAR model, i.e. we assume that

$$w_t = \sum_{j=1}^{\infty} \pi_{wj} w_{t-j} + \varepsilon_t \tag{2}$$

where $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a martingale difference sequence (MDS; for detailed assumptions see Section 3). Our primary interest lies in the process for y_t, which is approximated by

$$y_t = \sum_{j=1}^{p} (\pi_{yj} y_{t-j} + \pi_{z2j} z_{2t-j}) + \varepsilon_{yt,p}. \tag{3}$$

[6]While z_{2t} is optional, the results of Dufour and Renault (1998) underline its potential importance.
[7]Note, that although exogenously modelled, z_{t1} is not strictly exogenous in a statistical sense.

In order to consider linear alternatives to Granger noncausality, we must also include lags of z_{1t} in the empirical specification. Thus, we estimate the VARX model[8]

$$y_t = \sum_{j=1}^{p} (\psi_{yj} y_{t-j} + \psi_{z2j} z_{2t-j}) + \sum_{j=1}^{p_{z1}+1} \psi_{z1j} z_{1t-j} + \varepsilon_{yt,p} \tag{4}$$

and test the joint restriction $\psi_{z1j} = 0$ for $1 \leq j \leq p_{z1}$ using a standard Wald test.

The estimated model includes a surplus lag of the forcing variable, $z_{1t-p_{z1}-1}$, which is not tested. Its role becomes apparent after reparameterizing (4) as

$$y_t = \sum_{j=1}^{p} (\psi_{yj} y_{t-j} + \psi_{z2j} z_{2t-j}) + \sum_{j=1}^{p_{z1}} \psi_{z1j} \left(z_{1t-j} - z_{1t-p_{z1}-1}\right) + \left(\sum_{j=1}^{p_{z1}+1} \psi_{z1j} \right) z_{1t-p_{z1}-1} + \varepsilon_{yt,p}. \tag{5}$$

When z_{1t} is integrated of order less than 1.5 the parameters restricted under the null hypothesis (i.e. ψ_{z1j} for $1 \leq j \leq p_{z1}$) are expressed as the coefficients on the covariance stationary variables $z_{1t-j} - z_{1t-p_{z1}-1}$ (recall that p_{z1} is fixed) and may be shown to follow a joint normal limiting distribution under suitable conditions.

The choice of p_{z1}, the lag order of z_{1t}, will generally influence test power but not large sample size. Larger choices of p_{z1} allow more general alternatives, but may reduce power against simpler alternatives. Also, p_{z1} need not be set equal to p, the lag order of w_t. This is another way in which the VARX provides additional flexibility. Even if modelling z_{1t} requires many lags, e.g. if z_{1t} has long-memory, it may still be possible to model w_t parsimoniously. Likewise, in the pure VAR framework we require a surplus lag of all variables, whereas in the VARX we require only an extra lag of z_{1t}, improving efficiency, particularly when the lag order is small, but the dimension of w_t is large.

In order to rewrite (4) in compact form define $y_t^- := [y'_{t-1}, \ldots, y'_{t-p}]'$, $z_{2t}^- := [z'_{2t-1}, \ldots, z'_{2t-p}]'$, $\psi_y := [\psi_{y1}, \ldots, \psi_{yp}]$, $\psi_{z2} := [\psi_{z21}, \ldots, \psi_{z2p}]$, and $z_{1t}^- := [z'_{1t-1}, \ldots, z'_{1t-p_{z1}}]'$, so that $\varepsilon_{yt,p} = y_t - \psi_y y_t^- - \psi_{z2} z_{2t}^-$. We define by $x_{1t}^- = z_{1t}^-$ the regressors whose coefficients ψ_{x1} are to be tested. The remaining regressors, including the surplus lag, are then grouped together as $x_{2t}^- := [(y_t^-)', (z_{2t}^-)', (z_{1t-p_{z1}-1})']'$. Thus, the estimated equation in (4) may be rewritten in single equation form as

$$y_t = \psi_{x1} x_{1t}^- + \psi_{x2} x_{2t}^- + \varepsilon_{yt,p} \tag{6}$$

where $\psi_{x1} \in \mathbb{R}^{k_y \times p_{z1} k_{z1}}$ and $\psi_{x2} \in \mathbb{R}^{k_y \times (k_y p + k_{z2} p + k_{z1})}$ or in stacked form as

$$Y = X_1 \psi'_{x1} + X_2 \psi'_{x2} + \mathcal{E}_p, \tag{7}$$

[8]When the null hypothesis holds $\psi_{yj} = \pi_{yj}$ and $\psi_{z2j} = \pi_{z2j}$.

where $Y = \left[\begin{array}{ccc} y^-_{p_{max}+1}, & \ldots, & y^-_T \end{array} \right]'$, for $p_{max} = max\{p, p_{z1}+1\}$, and X_1, X_2, and $\mathcal{E}_p$ stack x^-_{1t}, x^-_{2t} and $\varepsilon^-_{yt,p}$ in identical fashion.

The null hypothesis of no-Granger causality is then $H_0 : \psi_{x1} = 0$ and the alternative hypothesis is $H_A : \psi_{x1} \neq 0$. Defining $X_{1.2} = X_1 - X_2(X_2'X_2)^{-1}X_2'X_1$, with rows denoted by $\left(x^-_{1.2t}\right)'$, as the residual from the projection of X_1 on X_2, we estimate the parameter of interest ψ_{x1} by $\hat{\psi}_{x1} = Y'X_{1.2}\left(X_{1.2}'X_{1.2}\right)^{-1}$ and the variance of $vec\left(\hat{\psi}_{x1}\right)$ by $\hat{\Sigma}_{x1} := \left((X_{1.2}'X_{1.2})^{-1} \otimes \hat{\Sigma}_\varepsilon\right)$, for $\hat{\Sigma}_\varepsilon := \frac{1}{T}\hat{\mathcal{E}}_p'\hat{\mathcal{E}}_p$, with the rows of $\hat{\mathcal{E}}_p$ given by $\hat{\varepsilon}_{yt,p}'$ for $\hat{\varepsilon}_{yt,p} := y_t - \hat{\psi}_{x1}x^-_{1t} - \hat{\psi}_{x2}x^-_{2t}$.[9] The standard Wald test for $\psi_{x1} = 0$ then takes the form:

$$\hat{W} := \mathrm{vec}(\hat{\psi}_{x1})'\hat{\Sigma}_{x1}^{-1}\mathrm{vec}(\hat{\psi}_{x1}) = \mathrm{vec}(Y'X_{1.2})'\left((X_{1.2}'X_{1.2})^{-1} \otimes \hat{\Sigma}_\varepsilon^{-1}\right)\mathrm{vec}(Y'X_{1.2}). \quad (8)$$

3 Large sample robustness results

In this section we show that the Wald statistic $\hat{W}$ for a test of Granger noncausality in the surplus lag VARX obeys a standard Chi-squared null limiting distribution under a variety of assumptions regarding the nature of the persistence in z_{1t}. We first state the assumptions on the innovation process for the endogenous variables w_t in (2):[10]

Assumption N: *The noise $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a strictly stationary ergodic martingale difference sequence adapted to the increasing sequence of sigma algebras $\mathcal{F}_t$ generated by $\varepsilon_t, \varepsilon_{t-1}, \ldots$. Further assume that $\mathbb{E}\{\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\} = \mathbb{E}\varepsilon_t\varepsilon_t' = \Sigma > 0$, $\mathbb{E}\{\varepsilon_{t,a}\varepsilon_{t,b}\varepsilon_{t,c}|\mathcal{F}_{t-1}\} = \omega_{a,b,c}$ (constant) where $\varepsilon_{t,a}$ denotes the a-th coordinate of ε_t, and $\mathbb{E}\{\varepsilon_{t,i}^4\} < \infty$.*

Many of the results presented below may be proved under more general assumptions on the innovations. In particular, finite fourth moments are often unnecessary. However, the above assumptions are standard in VAR models (Saikkonen and Lütkepohl, 1996, use similar but stronger assumptions) and provide a single set of assumptions that are sufficient for most of our results. A second restriction is the assumed conditional homoskedasticity of the innovations. If this restriction is dropped the asymptotic distributions change and robust standard errors would be needed. This is not pursued.

Under the null hypothesis we have $y_t = \varepsilon_{yt,p} + \psi_{x2}x^-_{2t}$ and hence $Y'X_{1.2} = \mathcal{E}_p'X_{1.2}$. This motivates the following high level assumptions where $\hat{\Gamma}_{1.2} := T^{-1}X_{1.2}'X_{1.2}$ is used:

Assumption HL: *Let $p = p(T)$, let p_{z1} be a fixed integer, and assume that*
(i) $\hat{\Sigma}_\varepsilon \to \Sigma$ in probability.

[9] Here $\otimes$ stands for the Kronecker product corresponding to columnwise vectorization.
[10] Note that $\mathcal{F}_{t-1,y,z2} = \mathcal{F}_{t-1}$ under the null hypothesis.

(ii) $\hat{\Gamma}_{1.2} \to \Gamma_{1.2}$ *in probability for some matrix* $\Gamma_{1.2} \in \mathbb{R}^{k_{z1}p_{z1} \times k_{z1}p_{z1}}, \Gamma_{1.2} > 0$.
(iii) $p(T)$ *is such that* $T^{-1/2} vec(\sum_{t=p+1}^{T} \varepsilon_{t,p} (x_{1.2t}^{-})') \xrightarrow{d} N(0, \Gamma_{1.2} \otimes \Sigma)$.

From these high level assumptions the standard asymptotics for the Wald test are immediate from (8).

Theorem 1 *Let Assumption HL hold for* $\varepsilon_{t,p} = y_t - \psi_{x2} x_{2t}^{-} - \psi_{x1} x_{1t}^{-}$. *Then, under* $H_0 : \psi_{x1} = 0$, $\hat{W} \xrightarrow{d} \chi^2(k_y p_{z1} k_{z1})$.

We show below that in a multitude of circumstances Assumption HL is fulfilled.

3.1 Infinite Order Stationary $VARX$

Before proceeding to the near unit-root and long-memory models, we first extend the approximation results of Lewis and Reinsel (1985) from the VAR to the VARX model in order to establish results for the stationary case. We employ the following assumptions:[11]

Assumption P1:
(i) The noise $(\varepsilon_t)_{t \in \mathbb{Z}}$ *fulfills Assumption N.*
(ii) $\sum_{j=1}^{\infty} \|\pi_{w,j}\|_2 < \infty$ *and* $\det \pi_w(z) \neq 0, for |z| \leq 1$, *where* $\pi_w(z) := I - \sum_{j=1}^{\infty} \pi_{w,j} z^j$.
(iii) The integer p *increases with* T *such that* $T^{1/2} \sum_{j=p+1}^{\infty} \|\pi_{w,j}\|_2 \to 0$ *and* $p^3/T \to 0$.
(iv) The process $(z_{1t})_{t \in \mathbb{Z}}$ *is generated according to the equation*

$$z_{1t} = \nu_t + \sum_{j=1}^{\infty} \theta_j \nu_{t-j} + \sum_{j=1}^{\infty} \phi_j \varepsilon_{t-j} \tag{9}$$

where $(\nu_t)_{t \in \mathbb{Z}}$ *fulfills Assumption N with* $\mathbb{E}\nu_t \nu_t' > 0$ *and is independent of the process* $(\varepsilon_t)_{t \in \mathbb{Z}}$. *Here* $\sum_{j=1}^{\infty} \|[\theta_j, \phi_j]\|_2 < \infty$ *is assumed.*

Assumptions (ii) and (iii) match those of Lewis and Reinsel (1985, Theorem 2, p. 398). However, the process $(z_{1t})_{t \in \mathbb{Z}}$ is not modelled endogenously, with the advantage of allowing the lag order p_{z1} for z_{1t} to vary freely, i.e. it is not tied to the approximation properties. Also, over-differenced processes are allowed for z_{1t}, as it does not require a $VAR(\infty)$ representation. The following result extends Theorem 3 of Lewis and Reinsel (1985) to the VARX framework:

[11] We define $|| \cdot ||_2$ as the Euclidean norm $\|x\|_2 = \sqrt{x'x}$, when applied to the vector x and as the induced matrix norm $max \{\|Ax\|_2 : x(n \times 1), \|x\|_2 = 1\}$ when applied to the $m \times n$ matrix A.

Theorem 2 *Let* $x_{2t}^{-} := [y_{t-1}', \ldots, y_{t-p}', z_{2t-1}', \ldots, z_{2t-p}', z_{1t-p_{z1}-1}']'$ *and* $x_{1t}^{-} := [z_{1t-1}', \ldots, z_{1t-p_{z1}}']'$. *Then Assumption P1 implies Assumption HL.*

The theorem shows that when the true process follows a VARX(∞,p_{z1}) the Wald test statistic can be used as if the true process was a $VARX(p, p_{z1})$. From the proof it is clear that in this special case the result also holds without the surplus lag $z_{1t-p_{z1}-1}$.

3.2 Infinite order I(1) and near-I(1) models

One of the main motivations behind the surplus lag approach was to obtain results without unit root and cointegration pre-testing in cases where components of y_t and/or z_t might be (co)integrated. Perhaps equally important to empirical applications is the near unit root model (Phillips, 1987; Chan, 1988), which approximates well the case in which the largest roots are indistinguishable from, but still less than, one. This often poses a challenge for inference since, in a local-to-unity model, the critical values of econometric tests designed for the I(0)/I(1) framework, including those that involve pre-testing, typically depend on the value of the localization parameter, which cannot be consistently estimated using a single observation of a time series (see e.g. Elliott (1998)). We will use the following assumptions:

Assumption P2 :

(i) Define $A_{T,w} := 1 + C_w/T, C_w = diag(c_1, c_2, \ldots, c_{k_y+k_{z2}-n})$ *and* $c_i \leq 0$ *for* $i = 1, \ldots c_{k_y+k_{z2}-n}$. *There exists a nonsingular matrix* $\Gamma = [\gamma_\perp, \gamma], \gamma \in \mathbb{R}^{(k_y+k_{z2})\times n}, 0 \leq n \leq k_y + k_{z2}$ *such that the process* $(v_t)_{t\in\mathbb{Z}}$ *obtained as (for suitable value* w_0*)*

$$v_t := \Big(\; (\gamma_\perp' w_t - A_{T,w}\gamma_\perp' w_{t-1})', \;\; (\gamma' w_t)' \; \Big)' \tag{10}$$

has an VAR(∞*) representation* $\sum_{j=0}^{\infty} \pi_{v,j} v_{t-j} = \varepsilon_t$ *where* $(\varepsilon_t)_{t\in\mathbb{Z}}$ *fulfills Assumption N.*

(ii) For $\pi_v(z) := \sum_{j=0}^{\infty} \pi_{v,j} z^j$ *we assume* $\det \pi_v(z) \neq 0, |z| \leq 1$.

(iii) Summability of the power series: $\sum_{j=1}^{\infty} j\|\pi_{v,j}\|_2 < \infty$.

(iv) The integer p *increases with* T *such that* $p^3/T \to 0$ *and* $T^{1/2}\sum_{j=p+1}^{\infty} \|\pi_{v,j}\|_2 \to 0$.

(v) Let $A_{T,z} := I + C_z/T$ *where* $C_z := S\,diag(c_{z,1}, \ldots c_{z,k_{z1}})S^{-1}$, $c_{z,i} \leq 0$ *for* $i = 1, \ldots k_{z1}$, *and* $S \in \mathbb{R}^{k_{z1}\times k_{z1}}$ *is nonsingular. The process* $(z_{1t} - A_{T,z}z_{1t-1})_{t\in\mathbb{Z}}$ *for some value* z_{10} *fulfills Assumption P1(iv) where additionally* $\sum_{j=1}^{\infty} j\|[\theta_j, \phi_j]\|_2 < \infty$ *holds.*

Under Assumption P2 y_t, z_{1t} and z_{2t} are all defined as triangular arrays[12] that can be either stationary, integrated, or near-integrated. Cointegrating relations may exist.

[12] For notational simplicity we follow common practice in suppressing the dependence on T.

The matrices of largest roots $A_{T,w}$ and $A_{T,z}$ depend on the matrices of local-to-unity parameters C_w and C_z respectively, allowing for a different local-to-unity parameter (c_i and $c_{z,i}$) in each element of $\gamma_\perp' w_t$ and z_{1t}. The component $\gamma' w_t$ is stationary, allowing for cointegration in w_t with cointegration rank n. The no cointegration case ($n = 0$) is also included. Cointegration between w_t and z_{1t} is allowed for, but not explicitly modeled. Results for exact unit roots hold when $c_i = c_{z,i} = 0$.

The theorem below shows that $\hat{W}$ has an asymptotic normal null distribution that is invariant to both the local-to-unity parameters and the cointegrating rank.

Theorem 3 *Let* $x_{2t}^- := [y_{t-1}', \ldots, y_{t-p}', z_{2t-1}', \ldots, z_{2t-p}', z_{1t-p_{z1}-1}']'$ *and* $x_{1t}^- := [z_{1t-1}', \ldots, z_{1t-p_{z1}}']'$. *Then Assumption P2 implies Assumption HL.*

In the special case of exact unit roots ($C_w = 0$, $C_z = 0$) the theorem extends the robustness results of Saikkonen and Lütkepohl (1996) to the VARX model. The asymptotic normality result in the more general local-to-unity framework is a rare property that underlines the practical value of the surplus lag method as a robust test.

3.3 Long-memory forcing variables

Models of fractional integration originating from (Granger and Joyeux, 1980; Hosking, 1981) provide another useful method of spanning the I(0)/I(1) divide. A variable z_{1t} is said to be integrated of order d if its fractional difference $(1-L)^d z_{1t}$ is I(0). Thus values of $0 < d < 1$ provide an intermediate between I(0) and I(1) models, in which shocks do decay, but only at a hyperbolic rate. These slow decay rates have been found useful for modelling a number of phenomena in economics and finance, such as volatilities (Baillie, 1996). For $d < 0.5$, the process fits into a larger class of stationary long-memory models. $d > 0.5$ corresponds to nonstationary fractional integration.

3.3.1 Stationary long-memory

Assumption P1 imposed short-memory via the summability assumptions on the MA(∞) representation of $(z_{1t})_{t\in\mathbb{N}}$. We now relax this condition.

Assumption P4 :
(i) Assumption P1, (i) - (iii) hold. Additionally $(\varepsilon_t)_{t\in\mathbb{Z}}$ *is assumed to be i.i.d.*
(ii) The process $(z_{1t})_{t\in\mathbb{Z}}$ *is generated according to the equation (9), where* $(\nu_t)_{t\in\mathbb{Z}}$ *fulfills Assumption N and is independent of the process* $(\varepsilon_t)_{t\in\mathbb{Z}}$. *Here* $\|[\theta_j, \phi_j]\|_2 \leq c j^{d-1}$ *for*

some constant $0 < c < \infty$ *and* $-0.5 < d < 0.5$ *is assumed.*
(iii) p *is chosen such that* $p = o(T^{1-2d})$ *and Assumption P1(iii) is fulfilled.*

Since the squared coefficients for $d \approx 0.5, d \leq 0.5$ are just summable, the conditions on the impulse response sequences are close to minimal. Thus, the assumptions on the exogenous inputs include many long-memory processes, including fractionally integrated processes and sums of fractionally integrated processes. On the other hand, we now require an additional condition on p, the number of lags included in the approximation for $1/3 < d < 1/2$ since in this case the estimates of the covariance sequence, including the cross covariance with lags of y_t and z_{2t}, are extremely unreliable. In fact, their covariances are of order $O(T^{4d-2})$ and hence arbitrarily small fractions of the sample size are obtained as convergence orders for values close to $d = 0.5$. This in turn limits the range of admitted processes via the assumption that $T^{1/2} \sum_{j=p+1}^{\infty} \|\pi_{w,j}\|_2 \to 0$. In some situations this is not a severe limitation. If the joint process w_t is a VARMA process then any rate of the form $p = T^c$ will fulfill the approximation restriction and choosing $c < 1 - 2d$ the condition on p is met.

In this setting the advantage of the VARX framework is clearly visible. If instead one modelled the process $[y_t', z_{1t}', z_{2t}']'$ using a VAR(p) then a large p would be required for a small approximation error $\varepsilon_{yt,p} - \varepsilon_{yt}$ due to the slow decay of the coefficients in the true VAR(∞) representation. Again it can be shown that Assumption HL holds. The following result also holds if the surplus lag $z_{1t-p_{z1}-1}$ is omitted.

Theorem 4 *Let* $x_{2t}^- := [y_{t-1}', \ldots, y_{t-p}', z_{2t-1}', \ldots, z_{2t-p}', z_{1t-p_{z1}-1}']'$ *and* $x_{1t}^- := [z_{1t-1}', \ldots, z_{1t-p_{z1}}']'$. *Then Assumption P4 implies Assumption HL.*

3.3.2 Nonstationary long-memory

We next establish that the surplus lag test also retains robustness under the following set of assumptions, which allow for forcing variables with nonstationary long-memory.

Assumption P5 :

(i) Assumption P1, (i) - (iii) hold. Additionally $(\varepsilon_t)_{t \in \mathbb{Z}}$ *is assumed to be i.i.d. and* $\sum_{j=1}^{\infty} j^{1+\delta} \|\pi_{w,j}\| < \infty$ *for some* $\delta > 0$.
(ii) There exists full column rank matrices $\beta \in \mathbb{R}^{k_{z1} \times (k_{z1} - c_{z1})}$ *and* $\beta_\perp \in \mathbb{R}^{k_{z1} \times c_{z1}}, \beta'\beta_\perp = 0$ *such that for* $\beta_\perp' z_{10} = 0$

$$\begin{bmatrix} \beta_\perp'(z_{1t} - z_{1t-1}) \\ \beta' z_{1t} \end{bmatrix} = v_t, t \in \mathbb{N}, \ \textit{where } v_{i,t} = \sum_{j=0}^{\infty} L_i(j) \frac{\Gamma(j + d_i)}{\Gamma(d_i)\Gamma(j+1)} \alpha_i' \begin{pmatrix} \nu_{t-j} \\ \varepsilon_{t-j} \end{pmatrix}, \quad (11)$$

for $-0.5 < d_i < 0.5$*,* $\|\alpha_i\|_2 = 1$*,* $\lim_{j\to\infty} L_i(j) = 1$*, and* $(\nu_t)_{t\in\mathbb{Z}}$ *i.i.d. and independent of* ε_t*, with* $\mathbb{E}\nu_t = 0$*,* $\mathbb{E}\nu_t\nu_t' > 0$ *and finite fourth moments.*
(iii) Defining $d_{\max} := \max(d_1, \ldots, d_{k_{z1}})$*, and* $d_{\min} := \min(d_1, \ldots, d_{c_{z1}})$*, p is chosen such that* $p = o_p\left(T^{\min\{1/3, 1-2d_{\max}, 1/3(1+2d_{\min})\}}\right)$ *and* $T^{1/2}\sum_{j=p+1}^{\infty}\|\pi_{w,j}\|_2 \to 0$*.*

Type I Nonstationary fractional integration (see Marinucci and Robinson (1999)) in the forcing variable is allowed for through the hyperbolic rates of decay on $\beta_\perp'(z_{1t} - z_{1t-1})$, through (11), which allows for different values of d in each element of $\beta_\perp' z_{1t}$. The cointegrating residuals, given by $\beta' z_{1t}$, may be fractionally integrated of order $-0.5 < d_i < 0.5$. The inclusion of the slowly varying coefficients, $L_i(j)$, lends flexibility to the short-memory dynamics, allowing for models such as the ARFIMA(p,d,q), as discussed in Davidson and Hashimzade (2007). The required restrictions on the increase of p as a function of the sample size are striking. Assumption P4 showed problems for d_i close to 0.5 due to the bad estimates of the covariance sequence. Assumption P5 indicates difficulties for d_i near -0.5, which results from the slow divergence rate of the nonstationary component, with integration $1 + d_i$ only slightly above 0.5. The borderline case $d_i = 0.5$ has not been analyzed.

Theorem 5 *Let* $x_{2t}^- := [y_{t-1}', \ldots, y_{t-p}', z_{2t-1}', \ldots, z_{2t-p}', z_{1t-p_{z1}-1}']'$ *and* $x_{1t}^- := [z_{1t-1}', \ldots, z_{1t-p_{z1}}']'$*. Then Assumption P4 implies Assumption HL.*

The theorem shows that when the forcing variables are fractionally integrated of order $0.5 < d < 1.5$ the null asymptotics remain standard. In the special case when the lag length p is known and finite, the validity of the excess lag test may be partially anticipated by the results of Dolado and Marmol (2004) who generalize the findings of Sims *et al.* (1990) to allow for nonstationary fractional integration. However, the above result appears to be the first to directly establish the validity of the surplus lag method with nonstationary fractionally integrated regressors. The allowance for unknown and possibly infinite order models complicates the analysis non-trivially.

3.4 Deterministic components and breaks

Because standard normal null asymptotics apply, it is expected that the above results can be extended to allow for de-meaning and de-trending, although this would require the generalization of a few technical results, such as Lemmas 3 and 4. Naturally, any structural breaks in the process for the dependent variable would require explicit modelling, including knowledge or estimation of the break date. On the other hand, the VARX structure does not require specific modelling of the forcing variable z_{1t}, thus allowing for unmodelled breaks in z_{1t} under suitable regularity conditions. For example, Theorem 7 of our working paper version showed that the same asymptotic normality result holds using demeaned data, in a model allowing for a finite number of level shifts in z_{1t}. The result is excluded in this version to save space.

4 Simulation results

Extensive finite sample size and power results in cointegrated I(1) models can be found in Dolado and Lütkepohl (1996) and Swanson *et al.* (2003) for the surplus-lag VAR and in our working paper version for the surplus-lag VARX. Below we study test size in Granger causality tests in near unit root and fractionally integrated models.

4.1 Simulation Models

In the models below, the null hypothesis that z_{1t} does not Granger cause y_t is imposed through the restriction that $\delta = 0$. The innovation process is specified as[13]

$$\varepsilon_t' = (\varepsilon_{1t}, \varepsilon_{2t}) \sim \text{i.i.d. } N(0, \Sigma) \text{ for } \Sigma = \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{bmatrix}, \text{ and } \sigma_{12} = -0.8. \tag{12}$$

4.1.1 Models with near-unit-root/local-to-unity

We define $c \leq 0$ as the local-to-unity coefficient and $a_T = 1 + c/T$. We include one model with non-cointegrated near unit roots (no-cointegration):

$$\begin{bmatrix} \Delta y_t \\ \Delta z_{1t} \end{bmatrix} = \begin{bmatrix} a_T - 1 & 0 \\ 0 & a_T - 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{1t-1} \end{bmatrix} + \begin{bmatrix} 0.5 & \delta \\ 0.3 & 0.5 \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta z_{1t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \tag{13}$$

and two that allow for cointegration between near unit roots. In the first (z-adjusts):

$$\begin{bmatrix} \Delta y_t \\ \Delta z_{1t} \end{bmatrix} = \begin{bmatrix} a_T - 1 & 0 \\ a_T & -1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{1t-1} \end{bmatrix} + \begin{bmatrix} 0.5 & \delta \\ 0.3 & 0.5 \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta z_{1t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{14}$$

(y_t, z_{1t}) have cointegrating vector $(1, -1)$ and z_{1t} adjusts to restore long-run equilibrium. In the second model (y-adjusts), specified by,

$$\begin{bmatrix} \Delta y_t \\ \Delta z_{1t} \end{bmatrix} = \begin{bmatrix} -1 & a_T\delta_1 \\ 0 & c/T \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{1t-1} \end{bmatrix} + \begin{bmatrix} 0.5 & \delta_2 \\ 0.3 & 0.5 \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta z_{1t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \tag{15}$$

[13] Granger noncausality (condition 1) has no implication for the residual cross-correlation, σ_{12}.

it is y_t that performs this adjustment and therefore the cointegrating vector $(-1, \delta_1)$ specifies an alternative to the null model of Granger noncausality. Both (14) and (15) are specializations of (Elliott, 1998, eq. 2).

4.1.2 Models involving fractional integration in z_{1t}

In this case, we assume that z_{1t} is fractionally integrated of order d and model it as:

$$z_{1t} = (1-L)^{-d}\varepsilon_{2t} \tag{16}$$

for $0 < d < 1$. We consider two models. In the first model:

$$\Delta y_t = 0.5\Delta y_{t-1} + \delta\Delta z_{1t-1} + \varepsilon_{1t}, \tag{17}$$

y_t is I(1) under both the null and alternative. Because z_{1t} is I(d) for $d < 1$, y_t and z_{1t} cannot cointegrate even under the alternative hypothesis. Finally, we consider

$$\Delta y_t = -0.5(y_{t-1} - \delta_1 z_{1t-1}) + \delta_2\Delta z_{t-1} + \varepsilon_{1t}, \tag{18}$$

in which y_t is I(0) under H_0: $\delta_1 = \delta_2 = 0$ and either I(0) ($\delta_1 = 0, \delta_2 \neq 0$) or I(d) and cointegrated with z_{1t} (with cointegrating vector $(1, -\delta_1)$) under the alternative.

4.2 Test procedures

As an example of the surplus lag VARX causality test (Surplus-VARX) analyzed in theoretical sections above, we estimate the autoregressive distributive lag or ARX model $\hat{y}_{1t} = \hat{a}(0) + \sum_{i=1}^{2}\hat{a}(i)y_{1t-i} + \sum_{i=1}^{3}\hat{b}(i)z_{1t-i}$, in which z_{1t} is an unmodelled forcing process, and test $H_0 : b(1) = b(2) = 0$. The implication of Granger noncausality for the surplus lag $b(3)$ is not tested and the forcing variable z_{1t} is not explicitly modelled. Because we do not require any extra surplus lags of y_{1t} we base the test on a surplus lag ARX(2,3) rather than a surplus-lag ARX(3,3).[14] As a basis of comparison, we also include results for Toda and Phillips (1993), based on a vector error correction model, with pre-tests for unit roots and cointegration rank. The procedure is both efficient and robust to unknown degrees of (integer) integration and cointegrating relations.

[14]Similar results were obtained when selecting the lag order p by the Akaike information criterion.

4.3 Rejection rates under the null

Both surplus-lag and the Toda-Phillips test, to which we compare, are known to provide good size in I(0), I(1), and cointegrated models. Here we study the robustness of finite sample size in the local-to-unity and fractionally integrated models of Section 4.1.

The results are divided into two tables. Table 1 shows simulation results under the local-to-unity DGPs of Section 4.1.1, using $c = -5.0$. The results in Columns 3-5, correspond to simulation models (13), (14), and model (15), respectively, with $\delta = 0$. In Table 2, we report the empirical size of the causality tests under the fractionally integrated models of Section 4.1.2, in which z_{1t} is generated by (16) using values of $d = 0.4$ (stationary long-memory) in Columns 3 and 5 and $d = 0.8$ (nonstationary fractional integration) in Columns 4 and 6. In Columns 3-4, the dependent variable, y_t, is generated by (17), with $\delta = 0$. In Columns 5-6 y_t is given by (18), with $\delta_1 = \delta_2 = 0$. Results for the Toda-Phillips and surplus-lag VARX tests are shown in the top and bottom panels, respectively. Within each panel we present results for sample sizes of $T = 50$, 100, 200, and 500. The table entries show finite sample rejection rates under the null hypotheses for a five-percent nominal test based on one thousand simulations.

Despite the fact that the Toda-Phillips procedure is not designed for the local-to-unity or fractionally integrated models, its size was still quite accurate across a number of these specifications.[15] Nevertheless, some size distortions were still detected under these models. On the other hand, the surplus-VARX based test provided reasonably accurate rejection rates in moderate sample sizes across all specifications considered. This underlines its value as a causality test that is particularly robust to misspecification of integration orders. Of course this robustness does not come without cost. As we discussed earlier, the addition of an unnecessary lag may be expected to reduce test power. Power comparisons included in our working paper, show that power losses associated with the surplus-lag are large when the alternative is described by a cointegrating vector, but can be quite modest against other forms of causality.

5 Conclusion

Employing a surplus lag in VAR based tests has been known to provide for inference which is invariant to possible $I(1)$ nonstationarity without necessitating unit root or cointegration pre-tests (Toda and Yamamoto, 1995; Dolado and Lütkepohl, 1996;

[15]Its accuracy was also found to improve for smaller values of $|\sigma_{12}|$ in (12).

Saikkonen and Lütkepohl, 1996). This provides for robust inference at some cost in terms of efficiency. On the other hand, there are arguably more efficient competing methods, which make fuller use of the I(0)/I(1) framework, without requiring knowledge on cointegration orders (Toda and Phillips, 1993; Kitamura and Phillips, 1997).

As our results demonstrate, the full advantage of the surplus approach becomes more apparent once one departs from both the pure VAR model and the I(0)/I(1) framework, of which it makes little explicit use, in order to allow for more general models of persistence. In particular, by applying the surplus lag to a VARX, in which the causing variables are exogenously modelled, we have shown that the same Chi-squared test statistic and critical values can be used to test Granger causality under a variety of possible data generating processes that may characterize the persistence in the forcing variable. These include the $I(0)$, $I(1)$ and cointegrated models considered earlier in the VAR context, as well as stationary and nonstationary long-memory, local-to-unity, and certain structural break models. Our simulation results suggest good finite sample size in both local-to-unity and fractionally integrated models.

References

Baillie, R. T (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics* **73**, 5–59.

Berk, K. N (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2**, 489–502.

Chan, N. H (1988). The parameter inference for nearly nonstationary time series. *Journal of the American Statistical Association* **83**(403), 857–862.

Chan, N. H and W Palma (1998). State space modeling of long-memory processes. *The Annals of Statistics* **26**, 719–740.

Choi, I (1993). Asymptotic normality of the least-squares estimates for higher order autorgressive integrated processes with some applications. *Econometric Theory* **9**, 263–282.

Christiano, L, M Eichenbaum and R Vigfusson (2003). What happens after a technology shock. Mimeo, Northwestern University.

Davidson, J (1994). *Stochastic Limit Theory*. Oxford University Press.

Davidson, J and N Hashimzade (2007). Convergence to stochastic integrals with fractionally integrated processes: Theory, and applications to cointegrating regression. Technical report. University of Exeter.

Diebold, F. X and A Inoue (2001). Long memory and regime switching. *Journal of Econometrics* **105**, 131–159.

Dolado, J and F Marmol (2004). Asymptotic inference results for multivariate long-memory processes. *Econometrics Journal* **7**, 168–190.

Dolado, J and H Lütkepohl (1996). Making Wald tests work for cointegrated VAR systems. *Econometric Reviews* **15**, 369–386.

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

Elliott, G (1998). On the robustness of cointegration methods when regressors have almost unit roots. *Econometrica* **66**, 149–158.

Friedman, B and K Kuttner (1992). Money, income, prices, and interest rates. *American Economic Reivew* **82**, 472–92.

Gourieroux, C and J Jasiak (2001). Memory and infrequent breaks. *Economics Letters* **70**, 29–41.

Granger, C. W and N Hyung (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance* **11**, 399–421.

Granger, C. W. J (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–459.

Granger, C. W. J and R Joyeux (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* **1**, 15–39.

Hall, P and C. C Heyde (1980). *Martingale Limit Theory and its Application*. Academic Press.

Hannan, E. J (1976). The asymptotic dibribution of serial covariances. *Annals of Statistics* **4**(2), 396–399.

Hannan, E. J and M Deistler (1988). *The Statistical Theory of Linear Systems*. John Wiley. New York.

Hidalgo, F. J (2000). Nonparametric test for causality with long-range dependence. *Econometrica* **68**, 1465–1490.

Hidalgo, F. J (2005). A bootstrap causality test for covariance stationary processes. *Journal of Econometrics* **126**, 115–143.

Hosking, J. R. M (1981). Fractional differencing. *Biometrika* **68**, 165–176.

Hosking, J. R. M (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *Journal of Econometrics* **73**, 261–284.

Hualde, J and P. M Robinson (2007). Root-n-consistent estimation of weak fractional cointegration. *Journal of Econometrics* **140**, 450–484.

Jansson, M and M. J Moreira (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica* **74**, 681–714.

Kim, C. S and P. C Phillips (2004). Fully modified estimation of fractional cointegration models. Mimeo.

Kitamura, Y and P. C. B Phillips (1997). Fully modified IV, GIVE and GMM estimation with possibly nonstationary regressors and instruments. *Journal of Econometrics* **80**, 85–123.

Lewis, R and G. C Reinsel (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis* **16**, 393–411.

Lütkepohl, H and P Saikkonen (1997). Impulse response analysis in infinite order cointegrated vector autoregressive processes. *Journal of Econometrics* **81**, 127–157.

Marin, D (1992). Is the export-led growth hypothesis valid for industrialized countries?. *The Review of Economics and Statistics* **74**, 678–688.

Marinucci, D and P. M Robinson (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference* **80**, 111–122.

Muller, U. K and M. W Watson (2007). Low-frequency robust cointegration testing. Mimeo, Princeton University.

Palma, W and M Zevallos (2004). Analysis of the correlation structure of square time series. *Journal of Time Series Analysis* **25**, 529 – 550.

Park, J. Y and P. C Phillips (1989). Statistical inference in regressions with integrated processes: Part II. *Econometric Theory* **5**, 95–131.

Phillips, P. C. B (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* **74**, 535–547.

Phillips, P. C. B (1995). Fully modified least squares and vector autoregression. *Econometrica* **63**, 1023–1078.

Phillips, P. C. B (2003). Laws and limits of econometrics. *Economic Journal* **113**, pp. C26–C52.

Phillips, P. C. B (2005). Challenges of trending time series econometrics. *Mathematics and Computers in Simulation* **86**, 401–416.

Phillips, P. C. B and V Solo (1992). Asymptotics for linear processes. *Annals of Statistics* **20**, 971–1001.

Poskitt, D. S (2007). Autoregressive approximation in nonstandard situations: The fractionally integrated and non-invertible cases. *Annals of the Institute of Statistical Mathematics* **59**, 697–725.

Saikkonen, P and H Lütkepohl (1996). Infinite-order cointegrated vector autoregressive processes. *Econometric Theory* **12**, 814–844.

Sims, C. A, J. H Stock and M. W Watson (1990). Inference in linear time series models with some unit roots. *Econometrica* **58**, 113–144.

Stambaugh, R. F (1999). Predictive regressions. *Journal of Financial Economics* **54**, 375–421.

Swanson, N. R, A Ozyildirim and M Pisu (2003). A comparison of alternative causality and predictive ability tests in the presence of integrated and cointegrated economic variables. In: *Computer Aided Econometrics* (David Giles, Ed.). pp. 91–148. Springer Verlag. New York.

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

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

A Technical lemmas

Lemma 1 *Let* $w_t = \sum_{j=0}^{\infty} \phi_{w,j}\varepsilon_{t-j}$ *where* $(\varepsilon_t)_{t\in\mathbb{Z}}$ *is an i.i.d. sequence of random variables having zero mean and finite fourth moments. Let* $\hat{\gamma}_j := T^{-1}\sum_{t=1+p}^{T} w_t w_{t-j}'$ *and* $\gamma_j := \mathbb{E}w_t w_{t-j}'$. *Assume that* $\phi_{w,j} = O(j^{d-1})$ *where* $-0.5 < d < 0.5$. *Then:*

$$\mathbb{E}vec(\hat{\gamma}_j - \mathbb{E}\hat{\gamma}_j)vec(\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k)' = \begin{cases} O(T^{4d-2}) & , \text{ for } \quad 0.25 < d < 0.5 \\ O(T^{-1}\log T) & , \quad d = 0.25, \\ O(T^{-1}) & , \quad -0.5 < d < 0.25 \end{cases}$$

All $O(.)$ *terms hold uniformly in* $1 \le j, k \le p$ *and* $1 \le p \le T$.

Proof: We employ Theorem 1, 3 and 5 of Hosking (1996). However, these results apply to fixed lags, whereas we require expressions uniformly in the lag. First note that using $\Omega := \mathbb{E}\varepsilon_t\varepsilon_t'$ we have for some constant $0 < K < \infty$ not depending on $j \in \mathbb{Z}$

$$\|\gamma_j\|_2 = \|\sum_{i=j}^{\infty} \phi_{w,i}\Omega\phi_{w,i-j}'\|_2 \le C\sum_{i=j}^{\infty}\|\phi_{w,i}\|_2\|\phi_{w,i-j}\|_2 \le Kj^{2d-1}$$

since $\|\Omega\|_2 < C, \|\phi_{w,i}\|_2 \le C_k i^{d-1}$ for some $K < \infty$ (see Lemma 2, (Palma and Zevallos, 2004)). The vector case is only notationally more complex and hence we only show the result for the case of scalar w_t. Then we obtain

$$\mathbb{E}\hat{\gamma}_j\hat{\gamma}_k = T^{-2}\sum_{t,s=1+p}^{T}\mathbb{E}w_{t+j}w_t w_s w_{s+k}.$$

Note that $\mathbb{E}w_t w_s w_r w_0 = \gamma_{t-s}\gamma_r + \gamma_{t-r}\gamma_s + \gamma_t\gamma_{s-r} + \kappa_4(t,s,r)$ for

$$\kappa_4(t,s,r) := \sum_{a=-\infty}^{\infty}\phi_{w,a+t}\phi_{w,a+s}\phi_{w,a+r}\phi_{w,a}(\mathbb{E}\varepsilon_t^4 - 3(\mathbb{E}\varepsilon_t^2)^2)$$

where for notational simplicity $\phi_{w,a} = 0, a < 0$ is used. It follows that $\mathbb{E}w_0^4 \le M_4 < \infty$ since $\|\phi_{w,a}^4\|_2 = O(a^{4d-4}) = o(a^{-2})$. Next

$$T^{-2} \sum_{t,s=1+p}^{T} \mathbb{E}w_{t+j}w_t w_s w_{s+k} = T^{-2} \sum_{t,s=1+p}^{T} \gamma_j\gamma_k + \gamma_{t-s+j}\gamma_{t-s-k} + \gamma_{t+j-s-k}\gamma_{t-s} + \kappa_4(t-s, t-s+j, k). \tag{19}$$

The first term is equal to $(T-p)^2 T^{-2}\gamma_j\gamma_k = \mathbb{E}\hat{\gamma}_j\mathbb{E}\hat{\gamma}_k$ independent of the value of d. The derivation of the bounds for the remaining terms in (19) will be done separately for the different cases for d. First consider $0.25 < d < 0.5$. The last term in (19) is majorized by the first term in (A.2) of Hosking (1996) and hence can be bounded by $M_{4\epsilon}T^{-1}\gamma_j\gamma_k$ where $M_{4\epsilon}$ is the fourth cumulant of ε_t. In fact this holds for any $d < 0.5$. The two middle terms can be dealt with using $\|\gamma_l\|_2 \le Kl^{2d-1}$ as shown above:

$$\begin{aligned}\left| T^{-2} \sum_{t,s=1+p}^{T} \gamma_{t-s+j}\gamma_{t-s-k} \right| &\le T^{-1} \sum_{l=1-T+p}^{T-1-p} |\gamma_{l+j}\gamma_{l-k}| \frac{T-|l|-p}{T} \\ &\le T^{-1} \left(\sum_{l=1-T+p}^{T-1-p} \gamma_{l+j}^2 \right)^{1/2} \left(\sum_{l=1-T+p}^{T-1-p} \gamma_{l-k}^2 \right)^{1/2}\end{aligned}$$

and for $j \ge 0$, using Lemma 3.2. (i) of Chan and Palma (1998), we have

$$\sum_{l=1-T+p}^{T-1-p} \gamma_{l+j}^2 \le \sum_{l=1-T+p}^{T-1+2j-p} \gamma_{l+j}^2 = \sum_{l=1-T-j+p}^{T-1+j-p} \gamma_l^2 = O((T-p+j)^{4d-1}) = O(T^{4d-1}).$$

This holds for $d \ne 0.25$. For $d = 0.25$ the same argument shows the bound $O(\log T)$ (cf. Hosking, 1996, top of p. 278). For $j \le 0$ the analogous argument can be used extending the sum to the negative integers. Combining these expressions we obtain $\mathbb{E}\hat{\gamma}_j\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_j\mathbb{E}\hat{\gamma}_k = \Delta_{j,k}$ where $\mathbb{E}|\Delta_{j,k}| \le MT^{4d-2}$ for $0.25 < d < 0.5$.
For $d = 0.25$ the same bound on the last term in (19) applies as for $0.25 < d < 0.5$. Further $\mathbb{E}|\Delta_{j,k}| \le M(\log T)/T$ for $d = 0.25$ by standard summability arguments showing that $\sum_{j=1}^{T} j^{-1} = O(T \log T)$ (see e.g. Hosking (1996), top of p. 278). This shows the claim for $d = 0.25$.
For $d < 0.25$ it follows that the middle two terms are of order $O(T^{-1})$ independent of j, k, p. Hence $\mathbb{E}|\Delta_{j,k}| \le M/T$ for $d < 0.25$. All bounds hold uniformly in $1 \le j, k \le p$ and $1 \le p \le T$. $\square$
Inspecting the proof it follows that it also applies (with $d = 0$) to linear processes $v_t = \sum_{j=0}^{\infty} \theta_{v,j}\varepsilon_{t-j}$ where $(\varepsilon_t)_{t\in\mathbb{Z}}$ fulfills Assumption N if $\sum_{j=0}^{\infty} \|\theta_{v,j}\|_2 < \infty$.

Lemma 2 *Let* $(\varepsilon_t)_{t\in\mathbb{Z}}$ *fulfill Assumption N. Let* $v_{t,p} = \sum_{j=0}^{\infty} \phi_{p,j}\varepsilon_{t-j}, t \in \mathbb{Z}, p \in \mathbb{N}$. *Then if* $\sup_{p\in\mathbb{N}} \sum_{j=0}^{\infty} \|\phi_{p,j}\|_2^2 < \infty$ *it follows that* $\sup_{p\in\mathbb{N}} \mathbb{E}\|v_{t,p}\|_2^4 < \infty$.

The proof is omitted to save space. It is included in the working paper version.

Lemma 3 *Let* Γ *denote the Gamma function and let* $L_i(j)$ *satisfy* $lim_{j\to\infty} L_i(j) = 1$ *for* $i = 1, \ldots, k_u$. *Then define* v_t *by* $\Delta v_t = u_t$, $t > 0$ *and* $v_t = 0$, $t \leq 0$, *where* $u_{i,t} = \sum_{j=0}^{\infty} \theta_{u,j,i}(\alpha_i'\varepsilon_{t-j})$, $\|\alpha_i\|_2 = 1$, $(\varepsilon_t)_{t\in\mathbb{Z}}$ *is i.i.d. with mean zero and finite fourth moments and* $\theta_{u,j,i} := \Gamma(d_i)^{-1}(j+1)^{(d_i-1)}L_i(j)$, *for* $0 < d_i < 1/2$ *and* $\theta_{u,j,i} := a_{j,i} - a_{j-1,i}$ *for* $j > 0$ *and* $\theta_{u,0,i} := a_{0,i}$ *for* $a_{j,i} := \Gamma(1+d_i)^{-1}(j+1)^{d_i}L_i(j)$ *for* $-1/2 < d_i < 0$. *Further let* $w_t = \sum_{j=0}^{\infty} \theta_{w,j}\varepsilon_{t-j}$ *for* $0 < \|\sum_{j=0}^{\infty} \theta_{w,j}\|_2 < \infty$ *and* $\theta_{w,j} := O(j^{-1-\delta})$ *for* $\delta > 0$. *Then using* $D_T := diag\left(T^{-(d_1+1)}, \ldots, T^{-(d_{k_u}+1)}\right)$ *and* $D_{T,0} := diag\left(T^{-(d_{1,0}+1)}, \ldots, T^{-(d_{k_u,0}+1)}\right)$, *for* $d_{i,0} := \max(d_i, 0)$, *we have (uniformly in* $p = o(T^{1/3})$*)*

$$(i) \qquad D_T \sum_{t=p+1}^{T} v_t v_t' D_T \xrightarrow{d} \Xi_d, \quad \text{where } \det \Xi_d \neq 0 \text{ a.s.}$$

$$(ii) \qquad \max_{0\leq j\leq H_T} \|D_{T,0} \sum_{t=p+1}^{T} v_t w_{t-j}'\|_2 = O_P(1), \quad \text{where } H_T = o(T^{1/3})$$

$$(iii) \qquad T^{-(1+\max(d_i+d_j,0))} \sum_{t=p+1}^{T} v_{i,t}u_{j,t}' = O_P(1),$$

$$(iv) \qquad D_T \sum_{t=p+1}^{T} v_{t-1}\varepsilon_t' = O_P(1).$$

Proof: (i), (iii), and (iv) follow from Proposition 4.1 and Theorem 4.1 of Davidson and Hashimzade (2007). For (ii), the convergence in distribution of $T^{-(d_{i,0}+1)} \sum_{t=p+1}^{T} v_{i,t}w_{j,t+1}'$ follows from Theorem 4.1. of Davidson and Hashimzade (2007). The uniform result in j can be derived from the following argument:

$$\begin{aligned}
T^{-(d_{i,0}+1)} \sum_{t=p+1}^{T} v_{i,t}w_{t-j}' &= T^{-(d_{i,0}+1)} \sum_{t=p+1}^{T} (v_{t,i} - v_{t-j-1,i})w_{t-j}' + T^{-(d_{i,0}+1)} \sum_{t=p+1}^{T} v_{t-j-1,i}w_{t-j}' \\
&= T^{-(d_{i,0}+1)} \sum_{r=0}^{j} \sum_{t=p+1}^{T} \Delta v_{t-r,i}w_{t-j}' + T^{-(d_{i,0}+1)} \sum_{t=p+1}^{T} v_{t-j-1,i}w_{t-j}' \\
&= T^{-d_{i,0}} \sum_{r=0}^{j} \left(T^{-1} \sum_{t=p+1}^{T} u_{t-r,i}w_{t-j}'\right) + T^{-(d_{i,0}+1)} \sum_{t=p+1}^{T} v_{t-j-1,i}w_{t-j}'.
\end{aligned}$$

The first term is the sum of $j+1$ estimated covariances to which we apply Lemma 1:

$$\sum_{r=0}^{j}\left(T^{-1}\sum_{t=p+1}^{T}u_{t-r,i}w'_{t-j}\right)=\sum_{r=0}^{j}\mathbb{E}u_{t-r,i}w'_{t-j}+\sum_{r=0}^{j}\left(T^{-1}\sum_{t=p+1}^{T}\left[u_{t-r,i}w'_{t-j}-\mathbb{E}u_{t-r,i}w'_{t-j}\right]\right)+O(pT^{-1})$$

which is of order $O(p^{d_{0,i}})+O_P((j+1)f_T)$ where $f_T=T^{2d_{0,i}-1}$ for $0.25<d_{0,i}<0.5$, $f_T=T^{-1/2}\sqrt{\log T}$ for $d_{0,i}=0.25$ and $f_T=T^{-1/2}$ for $d_{0,i}<0.25$. Here $\sum_{r=1}^{j}\mathbb{E}u_{t-r-1,i}w'_{t-j}=O(p^{d_{0,i}})$ is used which is straightforward to derive. Hence the first term above is of order $o(1)+O_P(jf_TT^{-d_{0,i}})=o_P(1)$ for $d_i>0$ and of order $O(1)+O_P(jT^{-1/2})=O_P(1)$ for $d_{0,i}<0$ uniformly in $0\le j\le T^{1/3}$.$\square$

Lemma 4 *Let $v_{t,T}-A_Tv_{t-1,T}=u_t, t\in\mathbb{N}, A_T=I-diag(c_1,\dots,c_k)/T, c_i\ge 0$ for $i=1,\dots k$, where u_t is stationary and ergodic with finite second moments generated according to $\sum_{j=0}^{\infty}\pi_{u,j}u_{t-j}=\varepsilon_t$ where $(\varepsilon_t)_{t\in\mathbb{Z}}$ fulfills Assumption N, and where, for $\pi_u(z):=\sum_{j=0}^{\infty}\pi_{u,j}z^j$, we have $\det\pi_u(z)\ne 0, |z|\le 1$ and $\sum_{j=0}^{\infty}\|\pi_{u,j}\|_2<\infty$. The recursions are started at $v_{0,T}=v_0, T\in\mathbb{N}$ which is assumed to be deterministic. Further let $w_t=\sum_{j=0}^{\infty}\phi^{\varepsilon}_{w,j}\varepsilon_{t-j}+\phi^{\eta}_{w,j}\eta_{t-j}$ where $\sum_{j=0}^{\infty}j\|\phi^{\varepsilon}_{w,j}\|_2<\infty, \sum_{j=0}^{\infty}\|\phi^{\eta}_{w,j}\|_2<\infty$ and $(\eta_t)_{t\in\mathbb{Z}}$ fulfills Assumption N and is independent of $(\varepsilon_t)_{t\in\mathbb{Z}}$. Then:*

(i) $\mathbb{E}\|v_{t,T}\|_2^2=O(t)$ *uniformly in T.*

(ii) $\mathbb{E}\|T^{-3/2}\sum_{t=p+1}^{T}v_{t,T}w'_t\|_2^2=O(T^{-1})$.

(iii) $T^{-2}\sum_{t=p+1}^{T}v_{t,T}v'_{t,T}\xrightarrow{d}\int_0^1J_c(w)J_c(w)'dw$ *where $J_c(w)$ denotes an Ornstein-Uhlenbeck process.*

(iv) $T^{-1}\sum_{t=p+1}^{T}v_{t,T}u'_t\xrightarrow{d}\int_0^1J_c(w)dB(w)'+\sigma_u$ *for some matrix σ_u. Here $B(w)$ denotes the Brownian motion associated with $T^{-1/2}u_t$.*

Proof: (i) According to the assumptions it follows that $u_t=\sum_{j=0}^{\infty}\phi_{u,j}\varepsilon_t$ (Lewis and Reinsel, 1985, p. 395, l.3). Further $\sum_{j=-\infty}^{\infty}\|\mathbb{E}u_0u'_j\|_2<\infty$ follows. The recursive definition of $v_{t,T}$ implies that $v_{t,T}=A_T^tv_0+\sum_{i=0}^{t-1}A_T^iu_{t-i}$. Consequently

$$\mathbb{E}\|v_{t,T}\|_2^2=\mathbb{E}(A_T^tv_0+\sum_{i=0}^{t-1}A_T^iu_{t-i})'(A_T^tv_0+\sum_{i=0}^{t-1}A_T^iu_{t-i})=\mathbb{E}v'_0(A_T^t)'A_T^tv_0+\sum_{i,j=0}^{t-1}\mathbb{E}u'_{t-i}(A_T^i)'A_T^ju_{t-j}.$$

Since $c_i\ge 0$ for $i=1,\dots k$, it follows that the elements of the diagonal matrix A_T are all less than one and hence $v_0(A_T^t)'A_T^tv_0=O(1)$. For the second term note that

$$|\sum_{i,j=0}^{t-1}\mathbb{E}u'_{t-i}(A_T^i)'A_T^ju_{t-j}|\le\sum_{i,j=0}^{t-1}\|\mathbb{E}u_{t-i}u'_{t-j}\|_2\le t\sum_{j=-\infty}^{\infty}\|\mathbb{E}u_0u'_j\|_2=O(t).$$

(ii) We will only deal with the univariate case, the multivariate case is only notationally more difficult. The process $(w_t)_{t\in\mathbb{N}}$ can be decomposed as $w_t := w_t^\varepsilon + w_t^\eta = (\sum_{j=0}^\infty \phi_{w,j}^\varepsilon \varepsilon_{t-j}) + (\sum_{j=0}^\infty \phi_{w,j}^\eta \eta_{t-j})$. Since ε_s and η_t are independent it follows that

$$\mathbb{E}v_{t,T}v_{s,T}w_t w_s = \mathbb{E}v_{t,T}v_{s,T}w_t^\varepsilon w_s^\varepsilon + \mathbb{E}v_{t,T}v_{s,T}\mathbb{E}w_t^\eta w_s^\eta \tag{20}$$

because $\mathbb{E}v_{t,T}v_{s,T}w_t^\varepsilon w_s^\eta = \mathbb{E}v_{t,T}v_{s,T}w_t^\varepsilon \mathbb{E}w_s^\eta = 0$ and expectations exist by Assumption N. We bound the contribution to $\mathbb{E}\|T^{-3/2}\sum_{t=p+1}^T v_{t,T}w_t\|_2^2$ of the second term in (20) by

$$T^{-3}\sum_{t=1+p}^T\sum_{s=1+p}^T |\mathbb{E}v_{t,T}v_{s,T}\mathbb{E}w_t^\eta w_s^\eta| \le T^{-3}\sum_{t=1+p}^T\sum_{s=1+p}^T t^{1/2}s^{1/2}|\mathbb{E}w_t^\eta w_s^\eta| = O(T^{-1})$$

due to $\sum_{j=-\infty}^\infty \|\mathbb{E}w_t^\eta w_{t-j}^\eta\|_2 < \infty$.

For the first term in (20), we use the Beveridge-Nelson decomposition (Phillips and Solo, 1992) $w_t^\varepsilon = \phi_w(1)\varepsilon_t + w_t^* - w_{t-1}^*$. We then rewrite $\sum_{j=p+1}^T v_{t,T}w_t^\varepsilon$ as a sum of several terms and show that the expectation of the square of each summand is of the required order. Of course, the cross terms are then of the same order. It follows that

$$\begin{aligned} T^{-3/2}\sum_{t=1+p}^T v_{t,T}w_t^\varepsilon &= T^{-3/2}\sum_{t=1+p}^T v_{t,T}\varepsilon_t\phi_w(1) + T^{-3/2}\sum_{t=1+p}^T v_{t,T}(w_t^* - w_{t-1}^*) \\ &= T^{-3/2}\sum_{t=1+p}^T v_{t,T}\varepsilon_t\phi_w(1) - T^{-3/2}\sum_{t=p}^{T-1}(v_{t+1,T} - v_{t,T})w_t^* \\ &\quad + T^{-3/2}v_{T,T}w_T^* - T^{-3/2}v_{p,T}w_p^* . \end{aligned} \tag{21}$$

Since $v_{T,T} = A_T^T v_0 + \sum_{i=0}^{T-1} A_T^i u_{T-i}$ it follows from finite fourth moments of u_t that $\mathbb{E}v_{T,T}^4 = O(T^4)$ and finite fourth moments of w_T^* (see the proof of Lemma 1) then imply via the Cauchy-Schwartz inequality that $\mathbb{E}v_{T,T}^2(w_T^*)^2 = O(T^2)$. Therefore the two last terms in the expression above contribute terms of the order $O(T^{-1})$ to $\mathbb{E}\|T^{-3/2}\sum_{t=p+1}^T v_{t,T}w_t\|_2^2$ as required. Further $v_{t,T} = A_T v_{t-1,T} + u_t$ and

$$\mathbb{E}\left(T^{-3/2}\sum_{t=1+p}^T v_{t-1,T}\varepsilon_t\right)^2 = T^{-3}\sum_{t,s=1+p}^T \mathbb{E}v_{t-1,T}\varepsilon_t v_{s-1,T}\varepsilon_s = T^{-3}\sum_{t=1+p}^T \mathbb{E}v_{t-1,T}^2\mathbb{E}\varepsilon_t^2 = O(T^{-1})$$

due to $\mathbb{E}\{\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\} = \mathbb{E}\varepsilon_t\varepsilon_t'$ and $\mathbb{E}v_{t,T}^2 = O(t)$. Obviously $\mathbb{E}(T^{-3/2}\sum_{t=p+1}^T u_t\varepsilon_t)^2 = O(T^{-1})$. Finally $v_{t,T} - v_{t-1,T} = v_{t,T} - A_T v_{t-1,T} + (A_T - 1)v_{t-1,T} = u_t - c/T v_{t-1,T}$ and therefore the square of the second term in (21) equals

$$T^{-3}\sum_{t,s=1+p}^T u_{t+1}u_{s+1}w_t^* w_s^* - \frac{c}{T}(v_{t,T}u_{s+1}w_t^* w_s^* + v_{s,T}u_{t+1}w_t^* w_s^*) + \frac{c^2}{T^2}v_{t,T}v_{s,T}w_s^* w_t^*.$$

Now $\mathbb{E}v_{t,T}^4 = O(t^4)$ and hence $\mathbb{E}v_{t,T}u_{s+1}w_t^* w_s^* \le (\mathbb{E}v_{t,T}^4)^{1/4}(\mathbb{E}u_{s+1}^4)^{1/4}(\mathbb{E}(w_t^*)^4)^{1/2} = O(t)$. Therefore (ii) follows.

The proofs for (iii) and (iv) are omitted since they closely follow previously established results. (iii) and (iv) are proved in Lemma 1 (c) and (d) of Phillips (1987) for the univariate case ($k = 1$) and in Lemma 1 (iii) and (iv) of (Elliott, 1998) for the multivariate case, in both cases under different assumptions on the process u_t. The main fact used in both cases, however, is that the process $X_T(t) = T^{-1/2}\sigma^{-1}\sum_{s=1}^{\lfloor tT \rfloor} u_s, 0 \leq t \leq T$ converges weakly to a Brownian motion. It is a standard result that this holds under our assumptions (see e.g. Hall and Heyde, 1980, Theorem 4.1.). □

Lemma 5 *Let the process* $(w_t)_{t\in\mathbb{Z}}$ *be generated according to Assumption P2 (i)-(ii) and be partitioned as* $w_t' = [y_t', z_{2t}']'$. *Accordingly let* ε_{yt} *denote the first block of* ε_t. *Define* $\pi_{w,0,T} := I, \Gamma' := \begin{pmatrix} \gamma_\perp' \\ \gamma' \end{pmatrix}, \pi_{w,j,T} := (\Gamma')^{-1}[\pi_{v,j}\Gamma' - \pi_{v,j-1}\begin{pmatrix} A_{T,w}\gamma_\perp' \\ 0 \end{pmatrix}], j \geq 1$. *Let* $\varepsilon_{yt,p} := \sum_{j=0}^{p-1}[I_s, 0]\pi_{w,j,T}w_{t-j} - [I_s, 0](\Gamma')^{-1}\pi_{v,p-1}\begin{pmatrix} A_{T,w}\gamma_\perp' \\ 0 \end{pmatrix} w_{t-p} = \varepsilon_{yt} - \sum_{j=p}^{\infty}[I_s, 0](\Gamma')^{-1}\pi_{v,j}v_{t-j}$. *Then, for a suitable constant* $c < \infty$ *not depending on* p,

$$\mathbb{E}(\|\varepsilon_{yt,p} - \varepsilon_{yt}\|_2^2)^{1/2} \leq c\sum_{j=p}^{\infty}\|\pi_{v,j}\|_2 \tag{22}$$

Proof: Using (10) and the definition of $\pi_{w,j,T}$ to substitute for w_t and $\pi_{w,j,T}$ respectively in the equation for $\varepsilon_{yt,p}$ we obtain $\varepsilon_{t,p} = \sum_{j=0}^{p-1}\pi_{v,j}v_{t-j}$ where $\varepsilon_t = \sum_{j=0}^{\infty}\pi_{v,j}v_{t-j}$. Then (22) follows by Lewis and Reinsel (1985), p. 397, (2.9) and $\varepsilon_{yt,p} = [I_s, 0](\Gamma')^{-1}\varepsilon_{t,p}$. □

Remark 1 *The Lemma holds for both the stationary (see Assumption P1) and (co)-integrated I(1) processes as special cases when* $\gamma_\perp = 0$ *and* $c = 0$, *respectively.*

Lemma 6 *Let* $R_T \in \mathbb{R}^{g_T \times g_T}$ *denote a sequence of (possibly random) matrices whose dimension* g_T *depends on the sample size* T. *Let* $\hat{R}_T$ *denote a sequence of random matrices such that* $\|\hat{R}_T - R_T\|_2 = O_P(f_T)$ *where* $f_T \to 0$. *Then if* $\sup_{T\in\mathbb{N}}\|R_T^{-1}\|_2 < \infty$ *a.s. it follows that* $\|\hat{R}_T^{-1} - R_T^{-1}\|_2 = O_P(f_T)$.

Proof: See Lewis and Reinsel (1985), p. 397, l. 11. □

Lemma 7

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A^{-1}B \\ I \end{bmatrix}\left[D - CA^{-1}B\right]^{-1}\left[\begin{array}{cc} -CA^{-1} & I \end{array}\right] \tag{23}$$

Proof: This can be verified by simple algebraic manipulations. □

Lemma 8 *Under Assumption P1(i), (ii) and (iv) let* $\Gamma_p := \mathbb{E}(x_t^-)(x_t^-)'$ *where* $x_t^- = [(x_{2t}^-)', (x_{1t}^-)']'$ *as defined in Theorem 2. Then* $\sup_{p\in\mathbb{N}} \|\Gamma_p^{-1}\|_2 < \infty$.

Proof: Since $z_{1t} = z_{1t}^\nu + z_{1t}^\varepsilon$ where $z_{1t}^\nu = \nu_t + \sum_{j=1}^\infty \theta_j \nu_{t-j}$ and $z_{1t}^\varepsilon = \sum_{j=1}^\infty \phi_j \varepsilon_{t-j}$ are mutually independent, we have $\mathbb{E} z_{1t-i} z_{1t-j}' = \mathbb{E} z_{1t-i}^\nu (z_{1t-j}^\nu)' + \mathbb{E} z_{1t-i}^\varepsilon (z_{1t-j}^\varepsilon)'$. Let x_{1t}^ε and x_{1t}^ν denote the components of x_{1t}^- generated from ε_t and ν_t respectively. Then

$$\begin{aligned}\Gamma_p &= \mathbb{E}\begin{bmatrix} y_t^-(y_t^-)' & y_t^-(z_{2t}^-)' & y_t^-(z_{1t-p_{z1}-1}^\varepsilon)' & y_t^-(x_{1t}^\varepsilon)' \\ z_{2t}^-(y_t^-)' & z_{2t}^-(z_{2t}^-)' & z_{2t}^-(z_{1t-p_{z1}-1}^\varepsilon)' & z_{2t}^-(x_{1t}^\varepsilon)' \\ z_{1t-p_{z1}-1}^\varepsilon(y_t^-)' & z_{1t-p_{z1}-1}^\varepsilon(z_{2t}^-)' & z_{1t-p_{z1}-1}^\varepsilon(z_{1t-p_{z1}-1}^\varepsilon)' & z_{1t-p_{z1}-1}^\varepsilon(x_{1t}^\varepsilon)' \\ x_{1t}^\varepsilon(y_t^-)' & x_{1t}^\varepsilon(z_{2t}^-)' & x_{1t}^\varepsilon(z_{1t-p_{z1}-1}^\varepsilon)' & x_{1t}^\varepsilon(x_{1t}^\varepsilon)' \end{bmatrix} \\ &+\mathbb{E}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & z_{1t-p_{z1}-1}^\nu(z_{1t-p_{z1}-1}^\nu)' & z_{1t-p_{z1}-1}^\nu(x_{1t}^\nu)' \\ 0 & 0 & x_{1t}^\nu(z_{1t-p_{z1}-1}^\nu)' & x_{1t}^\nu(x_{1t}^\nu)' \end{bmatrix} \stackrel{def}{=} \Gamma_p^\varepsilon + \Gamma_p^\nu\end{aligned}$$

Clearly $0 \le \Gamma_p^\varepsilon, 0 \le \Gamma_p^\nu$. Also the largest eigenvalue of both matrices are bounded uniformly in p (see Theorem 6.6.10. of Hannan and Deistler (1988) for Γ_p^ε; the nonzero eigenvalues of Γ_p^ν do not depend on p). Furthermore the matrix in the third and fourth block row and block column of Γ_p^ν is positive definite, since z_{1t} contains the term ν_t. For the heading subblock built from the first and second block row and columns of Γ_p^ε the smallest eigenvalue is bounded uniformly in p by Theorem 6.6.10. on p. 265 of Hannan and Deistler (1988). Suppose then that the uniform bound on the eigenvalues of Γ_p does not hold. Then there exists a sequence $p_T \to \infty$ and a sequence of unit norm vectors x_p such that $x_p'\Gamma_p x_p \to 0$. Then $x_p'\Gamma_p^\varepsilon x_p + x_p'\Gamma_p^\nu x_p \to 0$ and hence partitioning $x_p = [x_{p,1}', x_{p,2}', x_{p,3}', x_{p,4}']'$ where $x_{p,i}$ corresponds to the partitioning used previously it follows that $\mathbb{E}(x_{p,3}' z_{1t-p_{z1}-1}^\nu + x_{p,4}' x_{1t}^\nu)(x_{p,3}' z_{1t-p_{z1}-1}^\nu + x_{p,4}' x_{1t}^\nu)' \to 0$. It follows that $\|x_{p,3}\|_2 + \|x_{p,4}\|_2 \to 0$. From Theorem 6.6.10 of Hannan and Deistler (1988) it also follows that $\mathbb{E}(x_{p,1}' y_t^- + x_{p,2}' z_{2t}^-)(x_{p,1}' y_t^- + x_{p,2}' z_{2t}^-)' \to 0$ implies $\|x_{p,1}\|_2 + \|x_{p,2}\|_2 \to 0$. But this produces a contradiction to $\|x\|_2 = 1$. This shows the claim. □

Lemma 9 *Let* $(w_t)_{t\in\mathbb{Z}}$, $(\varepsilon_{yt,p})_{t\in\mathbb{Z}}$, *and* $\pi_{w,j,T}$, $j \ge 0$ *be defined as in Lemma 5. Then, under* $H_0 : \gamma_{z1j} = 0$ *for all* j, *and for* $T > max(c_i)$, *(4) can be reformulated as*

$$\Delta y_t = \Psi_{0,p,T}(\gamma_\perp' w_{t-1}) + \sum_{j=1}^{p} \Xi_{j,p,T} v_{t-j} + \left(\sum_{j=1}^{p_{z1}+1} \psi_{z1j}\right) z_{1t-p_{z1}-1} + \sum_{j=1}^{p_{z1}} \psi_{z1j}(z_{1t-j} - z_{1t-p_{z1}-1}) + \varepsilon_{yt,p}, \tag{24}$$

where $\sup_{p,T}(\sum_{j=1}^{\infty}\|\Xi_{j,p,T}\|_2)<\infty$, $\Psi_{0,p,T}:=-[I:0](\Gamma')^{-1}[I:0]'-\sum_{j=1}^{p-1}\pi_{\perp,j}A_{T,w}^{-(j-1)}-[I:0](\Gamma')^{-1}\pi_{v,p-1}[I:0]'A_T^{2-p}$, *and* $\Xi_{j,p,T}:=[\Xi_{1,j,p,T},\Xi_{2,j,T}]$ *for* $\Xi_{1,j,p,T}:=\sum_{h=j+1}^{p-1}\pi_{\perp,h}A_{T,w}^{-(h-j)}+(\Gamma')^{-1}\pi_{v,p-1}[I:0]'A_T^{j-p+1}$ *for* $j=1,\ldots,p-1$, *and* $\Xi_{1,p,p,T}:=0$, $\Xi_{2,1,T}:=-[I:0](I+\pi_{w,1,T})(\Gamma')^{-1}[0:I]'$, $\Xi_{2,j,T}:=-[I:0]\pi_{w,j,T}(\Gamma')^{-1}[0:I]'$ *for* $j=2,\ldots p-1$, $\Xi_{2,p,T}=0$, *and* $\pi_{\perp,j}:=[I:0]\pi_{w,j,T}(\Gamma')^{-1}[I:0]'$.

Remark 2 *A similar reformulation is employed in (A.2) of Saikkonen and Lütkepohl (1996) for the VAR case with* $A_{T,w}=I$. *However, the derivations and notation differ.*

Proof: Using $[\psi_{yj},\psi_{z2j}]=-[I,0]\pi_{w,j,T}, j=1,\ldots,p-1, [\psi_{yp},\psi_{z2p}]=[I_s,0](\Gamma')^{-1}\pi_{v,p-1}\left(\gamma_\perp A'_{T,w},\right)'$ (since $\gamma_{z1j}=0$ under H_0) and subtracting $y_{t-1}=[I:0]w_{t-1}$ from both sides of (4) and using $w_t=(\Gamma')^{-1}\Gamma'w_t=(\Gamma')^{-1}((\gamma_\perp'w_t)',v_{2,t}')'$, for $v_{2,t}=[0:I]v_t$, we obtain

$$\Delta y_t=[I:0]\left[-(\Gamma')^{-1}\begin{bmatrix}\gamma_\perp'w_{t-1}\\ v_{2,t-1}\end{bmatrix}-\sum_{j=1}^{p}\pi_{w,j,T}(\Gamma')^{-1}\begin{bmatrix}\gamma_\perp'w_{t-j}\\ v_{2,t-j}\end{bmatrix}\right]+\sum_{j=1}^{p_{z1}+1}\psi_{z1j}z_{1t-j}+\varepsilon_{yt,p}. \tag{25}$$

Defining $v_{1,t}:=[I:0]v_t=\gamma_\perp'w_t-A_{T,w}\gamma_\perp'w_{t-1}$ and noting that $A_{T,w}$ is invertible for $T>max(c_i)$, the terms involving $\gamma_\perp'w_{t-j}$ in (25) can be re-expressed as:

$$\left[-[I:0](\Gamma')^{-1}[I:0]'-\sum_{j=1}^{p}\pi_{\perp,j}A_{T,w}^{-(j-1)}\right]\gamma_\perp'w_{t-1}-\sum_{j=1}^{p-1}\sum_{h=j+1}^{p}\pi_{\perp,h}A_{T,w}^{-(h-j)}v_{1,t-j}.$$

Likewise, the terms involving z_{1t-j} may be re-expressed as in (5),yielding (24).
Since, by using (10) to substitute for v_j $j=0,1,2\ldots$ in $\sum_{j=0}^{\infty}\pi_{v,j}v_{t-j}=\varepsilon_t$, $\pi_{w,j,T}$ may be expressed as a linear finite lag function of $\pi_{v,j}$, $\sum_{j=1}^{\infty}j\|\pi_{w,j,T}\|<\infty$ follows by Assumption P2 (iii). $\sup_{p,T}(\sum_{j=1}^{\infty}\|\Xi_{1,j,p,T}\|_2)\le[I:0]\sum_{j=1}^{\infty}\sum_{h=j+1}^{\infty}\|\pi_{w,h,T}\|_2(\Gamma')^{-1}[I:0]'<\infty$ and absolute summability of $\Xi_{2,j}$ both follow. □

B Proof of Theorems

The proof of the theorems will be given based on the following lemma, which introduces a new set of high level conditions sufficient for Assumptions HL to hold:

Lemma 10 *Let* $(w_t)_{t\in\mathbb{Z}}$, $(\varepsilon_{yt,p})_{t\in\mathbb{Z}}$, *and* $\pi_{w,j,T}$, $j\ge 0$ *be defined as in Lemma 5. Assume that* $z_t^-\in\mathbb{R}^{k_{zp}}$ *is a vector, which is* $\mathcal{F}_{t-1}$ *measurable such that* $y_t=A(p)z_t^-+\varepsilon_{yt,p}=[A_1(p),A_2(p),A_3(p)][\left(z_{t,1}^-\right)',\left(z_{t,2,p}^-\right)',z_{3,t}']'+\varepsilon_{yt,p}$ *where* $z_t^-\in\mathbb{R}^{k_{zp}}$ *is partitioned as* $z_t^-=[\left(z_{1,t}^-\right)',\left(z_{2,t,p}^-\right)',z_{3,t}']'$ *such that* $z_{t,1}^-=\left[z_{t-1,1}',\ldots,z_{t-p_1,1}'\right]'\in\mathbb{R}^{k_{z1}}$ *(where* p_1 *is*

fixed) and $z_{3,t} \in \mathbb{R}^{k_{z3}}$ *do not depend on* p *and* $z_{2,t,p} = \left[z'_{2t-1}, \ldots, z'_{2t-p}\right]'$ *depends on* p. *Further let* p *tend to infinity as a function of the sample size such that* $p^3/T \to 0$ *and* $T^{1/2}\sum_{j=p+1}^{\infty} \|\pi_{v,j}\|_2 \to 0$ *such that* $\mathbb{E}(\|\varepsilon_{yt,p} - \varepsilon_{yt}\|_2^2)^{1/2} = o(T^{-1/2})$.
Then the following conditions are sufficient for Assumption HL to hold: There exists a matrix R_T *and a scaling matrix* $D_T = diag(I_{k_{z1}}T^{-1/2}, IT^{-1/2}, I_{k_{z3}}f_T)$ *such that (*λ_{max} *denotes the maximal eigenvalue)*

$$\max_{T\in\mathbb{N}} \lambda_{max}(\mathbb{E}R_T) = O(1) \quad , \quad \lambda_{max}(R_T) = O_P(1), \lambda_{max}(R_T^{-1}) = O_P(1), \tag{26}$$

$$R_T = \begin{bmatrix} R_{1,1} & R_{T,1,2} & 0 \\ R_{T,2,1} & R_{T,2,2} & 0 \\ 0 & 0 & R_{T,3,3} \end{bmatrix}, \tag{27}$$

$$\hat{R}_T := D_T \sum_{t=p+1}^{T} z_t^-(z_t^-)'D_T, \text{ such that } \|\hat{R}_T - R_T\|_2 = o_P(p^{-1/2}), \text{ and } \mathbb{E}\hat{R}_T = O(1) \text{ elementwise} \tag{28}$$

$$\sup_{l\in\mathbb{R}^{kzp}, \|l\|_2=1} T^{-1/2} \sum_{t=p+1}^{T} (\mathbb{E}\|l'D_T z_t^-\|_2^2)^{1/2} = O(1), \tag{29}$$

$$vec\left[\sum_{t=p+1}^{T} \varepsilon_{yt}(z_t^-)'D_T R_T^{-1} \begin{pmatrix} I & 0 & 0 \end{pmatrix}'\right] \xrightarrow{d} Z, \tag{30}$$

where $Z \sim N(0, \Gamma_{1.2}^{-1} \otimes \Sigma)$*, where* $\Gamma_{1.2} := \lim_{T\to\infty} R_{1,1} - R_{T,1,2}R_{T,2,2}^{-1}R_{T,2,1} > 0$.

Proof: Consider[16]

$$\begin{aligned}\hat{A}(p) &:= \sum_{t=p+1}^{T} y_t(z_t^-)'(\sum_{t=p+1}^{T} z_t^-(z_t^-)')^{-1} = A(p) + \sum_{t=p+1}^{T} \varepsilon_{yt,p}(z_t^-)'D_T(D_T \sum_{t=p+1}^{T} z_t^-(z_t^-)'D_T)^{-1}D_T \\ &+ O(T^{-1}) = A(p) + \left(\sum_{t=p+1}^{T} \varepsilon_{yt,p}(z_t^-)'D_T\right)\hat{R}_T^{-1}D_T + O(T^{-1}),\end{aligned}$$

where $\sum_{t=p+1}^{T} \varepsilon_{yt,p}(z_t^-)'D_T = \sum_{t=p+1}^{T} \varepsilon_{yt}(z_t^-)'D_T + \sum_{t=p+1}^{T}(\varepsilon_{yt,p} - \varepsilon_{yt})(z_t^-)'D_T$ and

$$\begin{aligned}\mathbb{E}\|\sum_{t=p+1}^{T}(\varepsilon_{yt,p} - \varepsilon_{yt})(z_t^-)'D_T\|_2 &\leq \sum_{t=p+1}^{T}(\mathbb{E}\|\varepsilon_{yt,p} - \varepsilon_{yt}\|_2^2)^{1/2}(\mathbb{E}\|D_T(z_t^-)\|_2^2)^{1/2} \\ &= (T^{1/2}\left(\mathbb{E}\|\varepsilon_{y1,p} - \varepsilon_{y1}\|_2^2\right)^{1/2})\left(T^{-1/2}\sum_{t=p+1}^{T}(\mathbb{E}\|D_T(z_t^-)\|_2^2)^{1/2}\right) = o(p^{1/2}). \end{aligned} \tag{31}$$

[16]The $O(T^{-1})$ term is due to the dependence of $A(p)$ on $A_{T,z}/T, A_{T,w}/T$ in the local-to-unity case, see Lemma 5.

Here (29) and Lemma 5 are used. Moreover letting $\varepsilon_{yt(i)}, i = 1, \ldots, k_y$, denote a coordinate of ε_{yt} we have

$$\mathbb{E}(\sum_{t=p+1}^{T} \varepsilon_{yt(i)}(z_t^-)' D_T)'(\sum_{t=p+1}^{T} \varepsilon_{yt(i)}(z_t^-)' D_T) = \sum_{t=p+1}^{T} \mathbb{E}\varepsilon_{yt(i)}^2 \mathbb{E} D_T z_t^-(z_t^-)' D_T = \mathbb{E}\varepsilon_{y1(i)}^2 \mathbb{E}\hat{R}_T$$

using the martingale difference property. Therefore $\| \sum_{t=p+1}^{T} \varepsilon_{yt,p}(z_t^-)' D_T \|_2 = O_P(p^{1/2})$. Consequently, $\|(\hat{A}(p) - A(p)) D_T^{-1}\|_2 = O_P(p^{1/2})$ using (29), (26, 28) and Lemma 6. Then consider $\hat{\Sigma}_\varepsilon := T^{-1} \sum_{t=p+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$: We obtain

$$\begin{aligned}
\hat{\Sigma}_\varepsilon &= \frac{1}{T} \sum_{t=p+1}^{T} (y_t - \hat{A}(p) z_t^-)(y_t - \hat{A}(p) z_t^-)' \\
&= \frac{1}{T} \sum_{t=p+1}^{T} (\varepsilon_{yt,p} - (\hat{A}(p) - A(p)) z_t^-)(\varepsilon_{yt,p} - (\hat{A}(p) - A(p)) z_t^-)' \\
&= \frac{1}{T} \sum_{t=p+1}^{T} \varepsilon_{yt,p}\varepsilon_{yt,p}' - \frac{1}{T} \sum_{t=p+1}^{T} \varepsilon_{yt,p}(z_t^-)'(\hat{A}(p) - A(p))' - \frac{1}{T} \sum_{t=p+1}^{T} (\hat{A}(p) - A(p)) z_t^- \varepsilon_{yt,p}' \\
&\quad + (\hat{A}(p) - A(p)) \left(\frac{1}{T} \sum_{t=p+1}^{T} z_t^-(z_t^-)' \right) (\hat{A}(p) - A(p))' \\
&= \Sigma + o_P(1) + O_P(p/T) = \Sigma + o_P(1).
\end{aligned}$$

Here the bound follows from $T^{-1} \sum_{t=p+1}^{T} \varepsilon_{yt,p}\varepsilon_{yt,p}' \to \Sigma$, which can be shown using Lemma 5 and the ergodicity of $(\varepsilon_t)_{t \in \mathbb{Z}}$, implying that $T^{-1} \sum_{t=p+1}^{T} \varepsilon_t \varepsilon_t' \to \Sigma$ almost surely. Further $\|(\hat{A}(p) - A(p)) D_T^{-1}\|_2 = O_P(p^{1/2})$, $\|\hat{R}_T\|_2 = O_P(1)$ and $\| \sum_{t=p+1}^{T} D_T z_t^- \varepsilon_{yt,p}' \|_2 = O_P(p^{1/2})$ are used. This shows HL (i).
Next, note that $\hat{\Gamma}_{1.2}^{-1}$ equals the (1,1) block of $\hat{R}_T^{-1}$. Then (28) and (26) imply HL (ii). With respect to HL (iii) note that $x_{1.2t}^- = [\hat{\Gamma}_{1.2}, 0] D_T^{-1} \hat{R}_T^{-1} D_T z_t^-$. Therefore

$$T^{-1/2} \sum_{t=p+1}^{T} \varepsilon_{yt,p}(x_{1.2t}^-)' = \sum_{t=p+1}^{T} \varepsilon_{yt,p}(z_t^-)' D_T \hat{R}_T^{-1} \begin{bmatrix} \hat{\Gamma}_{1.2} \\ 0 \\ 0 \end{bmatrix} = \sum_{t=p+1}^{T} \varepsilon_{yt}(z_t^-)' D_T R_T^{-1} \begin{bmatrix} \hat{\Gamma}_{1.2} \\ 0 \\ 0 \end{bmatrix} + o_P(1)$$

since $\| \sum_{t=p+1}^{T} (\varepsilon_{yt} - \varepsilon_{yt,p})(z_t^-)' D_T l \|_2 = o_P(1)$ similar to (31) and $\sum_{t=p+1}^{T} \varepsilon_{yt}(z_t^-)' D_T = O_P(p^{1/2})$ as used above. Then (30) and HL (ii) imply HL (iii). □

B.1 Proof of Theorem 2

Proof: The proof uses a number of results of Lewis and Reinsel (1985), henceforth LR. We verify the conditions of Lemma 10 where $z_{1,t}^- := x_{1t}^-, z_{2,t,p}^- := x_{2t}^-$ and $z_{3,t}$

does not occur. Thus $k_{zp} = k_{z1}p_{z1} + p(k_y + k_{z2}) + k_{z1}$ and $D_T = T^{-1/2}I$. Also, by assumption, all variables are stationary with bounded variance. Then $\mathbb{E}\hat{R}_T = (T-p)/TR_T$. The maximum eigenvalue of R_T is bounded uniformly in $T \in \mathbb{N}$ since z_t^- is a vector containing only lags of the vector process $[w_t', z_{1t}']'$, which has bounded spectrum due to the summability assumptions on the autoregression coefficients (see e.g. Hannan and Deistler, 1988, p. 265). The bound on the minimum eigenvalue of R_T is derived in Lemma 8. This verifies (26), (27) and (29).

Each entry in $\hat{R}_T - R_T$ is equivalent to an estimated covariance at some lag up to an approximation error due to the different limits of summation. Lemma 1 shows that the variance of the estimators of the covariances are of order $O(T^{-1})$, see also Hannan (1976), Chapter 4. The change in the summation introduces an error of order $O_P(pT^{-1})$ since the difference is a sum of a maximum of p terms each of variance $O(T^{-2})$. Thus all entries in $\hat{R}_T - R_T$ are of order $O_P(T^{-1/2})$ and therefore $\|\hat{R}_T - R_T\|_2 = O_P(pT^{-1/2})$. Then $p/T^3 \to 0$ implies that $pT^{-1/2} = o(p^{-1/2})$ showing (28).

Finally (30) follows as in Theorem 3 of LR (see also Theorem 7.4.9. of Hannan and Deistler, 1988). The only change in the arguments lies in the different definition of the regressors and correspondingly the replacement of Γ_p of LR by R_T. In the proof the uniform bound on $\lambda_{max}(R_T^{-1})$ derived above is crucial. Details are omitted. □

B.2 Proof of Theorem 3

Proof: The proof builds on Saikkonen and Lütkepohl (1996), henceforth SP96. We re-parameterize the auxiliary model (4) using (24), which is permissible for our purpose since we test only $\psi_{z1j}, j = 0, \ldots, p_{z1}$, whose estimates coincide in (4) and (24). Note that in (24) there are two variables containing nonstationary regressors: $(\gamma_\perp' w_{t-1})$ and $z_{1t-p_{z1}-1}$. Assumption P2 allows for full column rank matrices $\beta \in \mathbb{R}^{(n+k_{z1})\times n_z}$ with $0 \le n_z \le n + k_{z1}$ and $\beta_\perp \in \mathbb{R}^{(n+k_{z1})\times(n+k_{z1}-n_z)}$ such that $\beta'\beta_\perp = 0$[17] where $(\tilde{n}_{t,\perp})_{t\in\mathbb{N}}, \tilde{n}_{t,\perp} := \beta'[(\gamma_\perp' w_{t-1})', z_{1t-p_{z1}-1}']'$ is stationary and $(\tilde{n}_t)_{t\in\mathbb{N}}, \tilde{n}_t := \beta_\perp'[(\gamma_\perp' w_{t-1})', z_{1t-p_{z1}-1}']'$ is integrated (but not cointegrated). Thus instead of (24), we consider

$$\Delta y_t = [\psi_{x1}, \tilde{\psi}_{x2}, \tilde{\Psi}_0][(z_{1,t}^-)', (z_{2,t,p}^-)', z_{3,t}']' + \varepsilon_{yt,p} = A(p,T)z_t^- + \varepsilon_{yt,p}, \tag{32}$$

where $z_{1,t}^- := \tilde{x}_{1t}^- = [(z_{1t} - z_{1t-p_{z1}-1})', \ldots, (z_{1t-p_{z1}} - z_{1t-p_{z1}-1})']'$, $z_{2,t,p}^- := [\tilde{n}_{t,\perp}', v_{t-1}', \ldots, v_{t-p+1}', (\gamma' w_{t-p})']'$ and $z_{3,t} := \tilde{n}_t$ analogously to the definition in Lemma A.3. of SP96. Here $(z_{2,t,p}^-)_{t\in\mathbb{Z}}$

[17] Cointegration between $\gamma_\perp' w_{t-1}$ and $z_{1t-p_{z1}-1}$ is allowed for, but not imposed. The no cointegration case is accommodated by taking $n_z = 0$.

is stationary for given value of p. $z^-_{1,t} := [(z_{1t} - z_{1t-p_{z1}-1})', \ldots, (z_{1t-p_{z1}} - z_{1t-p_{z1}-1})']'$ behaves essentially as a stationary process since $z_{1t-j} - A_{T,z}^{p_{z1}+1-j} z_{1t-p_{z1}-1}$ is stationary (as a finite sum of stationary terms) and therefore

$$z_{1t-j} - z_{1t-p_{z1}-1} = z_{1t-j} - A_{T,z}^{p_{z1}-j+1} z_{1t-p_{z1}-1} + (A_{T,z}^{p_{z1}-j+1} - 1) z_{1t-p_{z1}-1},$$

where $A_{T,z}^{p_{z1}-j+1} - 1 = O(T^{-1})$. Thus it follows from Lemma 4 that the second term is negligible and it is sufficient to verify the conditions of Lemma 10.

Define $\hat{R}_T := D_T(\sum_{t=p+1}^T z_t^-(z_t^-)')D_T$ for $D_T := \text{diag}(T^{-1/2}I, T^{-1/2}I, T^{-1}I)$, with partitioning corresponding to that of z_t^- in (32). The last k_{z3} coordinates of z_t^- are integrated. The rest are stationary, apart from lower order remainders. Further let

$$R_T := \begin{bmatrix} \mathbb{E}z^-_{1,t}(z^-_{1,t})' & \mathbb{E}z^-_{1,t}(z^-_{2,t})' & 0 \\ \mathbb{E}z^-_{2,t}(z^-_{1,t})' & \mathbb{E}z^-_{2,t}(z^-_{2,t})' & 0 \\ 0 & 0 & T^{-2}\sum_{t=p+1}^T \tilde{n}_t \tilde{n}_t' \end{bmatrix},$$

such that obviously (27) holds. Here the submatrix built of the first two block rows and columns of R_T has uniformly bounded eigenvalues (both from below and from above) due to Lemma 8 as in the proof of Theorem 2. The nonsingularity (in probability) of the (3,3) block of R_T follows from the convergence in distribution (cf. Lemma 4 (iii)) to an almost sure positive definite random matrix. Therefore $\lambda_{max}(R_T) = O_P(1)$ and $\lambda_{max}(R_T^{-1}) = O_P(1)$ establishing (26). $\mathbb{E}\hat{R}_T = O(1)$ is easy to verify from the results of the proof of Theorem 2 and $\mathbb{E}\tilde{n}_t\tilde{n}_t' = O(t)$ from standard theory.

Next, Lemmas 1 (for $d = 0$) and 4 (ii) imply that each entry in $\hat{R}_T - R_T$ has variance uniformly of order $O(T^{-1})$. Thus $\|\hat{R}_T - R_T\|_2 = O_P(p/T^{-1/2})$ showing (28) for $p = o(T^{1/3})$. Then consider $\mathbb{E}\|l'D_T z_t^-\|_2^2 = \mathbb{E}(T^{-1}\|l_1' z^-_{1,t}\|_2^2 + T^{-1}\|l_2' z^-_{2,t}\|_2^2 + T^{-2}\|l_3' z_{3,t}\|_2^2)$ where $l' = [l_1', l_2', l_3']$ is partitioned in accordance with z_t^-. By Lemma 4 (i), $\mathbb{E}\|z_{3,t}\|_2^2 = O(t)$. Due to stationarity of the remaining terms $\mathbb{E}\|l'D_T z_t^-\|_2^2 = O(T^{-1})$, analogously to the proof in Theorem 2, and (29) follows. Finally, in $\sum_{t=p+1}^T \varepsilon_{yt}(z_t^-)'D_T R_T^{-1}[I, 0, 0]'$ the nonstationary terms do not occur due to the block diagonal structure of R_T. Thus analogous arguments as in the proof of Theorem 2 imply that (30) holds. □

B.3 Proof of Theorem 4

Proof: The proof follows that of Theorem 2, except that the impulse response sequence corresponding to z_{1t} is not summable. (Note that w_t is short-memory.) Hence let $D_T = T^{-1/2}I$, $R_T = \mathbb{E}z_t^-(z_t^-)'$, and $\hat{R}_T := T^{-1}\sum_{t=p+1}^T z_t^-(z_t^-)'$, where z_t^- is defined as in the proof of Theorem 2. To show $\|\hat{R}_T - R_T\|_2 = o_P(p^{-1/2})$, note that every entry in this matrix converges in mean square since, by Lemma 1, the variances are of order $O(T^{\max(4d-2,-1)})$ for $d \neq 0.25$ and of order $O(T^{-1}\log T)$ for $d = 0.25$. Note that $\mathbb{E}\hat{\gamma}_j = (T-p)/T\gamma_j$. Hence $\mathbb{E}\hat{R}_T = (T-p)/TR_T$. Thus the expectation of the sum of squared entries of $\hat{R}_T - R_T$ is of order $O(T^{4d-2}p + p^2T^{-1})$, $O(pT^{-1}\log T + p^2T^{-1})$, and $O(pT^{-1} + p^2T^{-1})$ for $0.25 < d < 0.5$, $d = 0.25$, and $d < 0.25$, respectively. This follows since there are only $O(p)$ terms involving the long-memory processes, as w_t has short memory and contributes p^2 terms of order $O(T^{-1})$. Hence, for obtaining $\|\hat{R}_T - R_T\|_2 = o_P(p^{-1/2})$ it suffices that $p^2T^{4d-2} + p^3T^{-1} \to 0$ for $0.25 < d < 0.5$, $(p^2\log T + p^3)/T \to 0$ for $d = 0.25$, and $p^3T^{-1} \to 0$ otherwise. This shows (28). The bounds in (26) follow from Lemma 8 (which did not use the short memory assumption on z_{1t}) as in the proof of Theorem 2. Since $z_{3,t}$ does not occur (27) follows trivially. Stationarity and finite variances of $(z_{1t})_{t\in\mathbb{N}}$ implies (29) as in the proof of Theorem 2.

It remains to verify (30). In the following we will only deal with the scalar output case (i.e. $k_y = 1$). The multivariate case is only notationally more difficult. It is sufficient to show that $T^{-1/2}\sum_{t=p+1}^{T}\varepsilon_{yt}(\alpha_p' z_t^-)$ is asymptotically normal with $\alpha_p' R_T \alpha_p \to \alpha_\infty' R_\infty \alpha_\infty$ for vector sequences α_p such that $0 < c < \inf_{p\in\mathbb{N}}\|\alpha_p\|_2 \leq \sup_{p\in\mathbb{N}}\|\alpha_p\|_2 \leq C$ for some constants $0 < c < C < \infty$ and $\|[\alpha_p', 0]' - \alpha_\infty\|_2 \to 0$ holds. Clearly the columns of R_T^{-1} fulfill these requirements. In this respect we use the three series criterion of Hall and Heyde (1980, Theorem 3.2, p. 58): With $X_{Tt} = \varepsilon_{yt}(\alpha_p' z_t^-)/\sqrt{T}$ we obtain that $(X_{Tt})_{1\leq t\leq T}$ is a martingale difference sequence with respect to the sigma field generated by $\varepsilon_s, \nu_s, s \leq t$. Below we deal only with the univariate case. The multivariate case follows as usual from the Cramer-Wold device (see e.g. Davidson, 1994, Theorem 25.5.). Then Theorem 3.2. states that $\sum_{t=1}^{T} X_{Tt} \xrightarrow{d} \mathcal{N}(0,\eta^2)$ if

$$(i)\max_{1\leq t\leq T}|X_{Tt}| \xrightarrow{p} 0, \quad (ii)\sum_{t=1}^{T} X_{Tt}^2 \xrightarrow{p} \eta^2(\text{a constant}), \quad (iii)\ \mathbb{E}\max_{1\leq t\leq T} X_{Tt}^2 \text{ is bounded in } T\ .$$

Assume that $\alpha_p' R_T \alpha_p \to \tilde{\eta}^2$ (for some constant $\tilde{\eta}$) as $p \to \infty$. Then it holds that $\mathbb{E}\varepsilon_{yt}^2(\alpha_p' z_t^-)^2 = \mathbb{E}\varepsilon_{yt}^2\mathbb{E}(\alpha_p z_t^-)^2 < M$ for some constant $0 < M < \infty$ uniformly in $p \in \mathbb{N}$ due to the conditional homoskedasticity and the assumption of finite second moments of z_t^-. Then $\mathbb{E}\max_{1\leq t\leq T} X_{Tt}^2 \leq \sum_{t=1}^{T}\mathbb{E}X_{Tt}^2 \leq M$ such that (iii) follows. Secondly,

$$\sum_{t=1}^{T} X_{Tt}^2 = T^{-1}\sum_{t=1}^{T}\varepsilon_{yt}^2(\alpha_p' z_t^-)^2 = T^{-1}\sum_{t=1}^{T}(\varepsilon_{yt}^2 - \mathbb{E}\varepsilon_{yt}^2)\alpha_p' z_t^-(z_t^-)'\alpha_p + \left(T^{-1}\sum_{t=1}^{T}\alpha_p' z_t^-(z_t^-)'\right)\alpha_p\mathbb{E}\varepsilon_{yt}^2$$

where $\alpha_p'(T^{-1}\sum_{t=1}^{T} z_t^-(z_t^-)')\alpha_p = \alpha_p'\hat{R}_T\alpha_p \to \tilde{\eta}^2$ since $\|\hat{R}_T - R_T\|_2 \to 0$. Therefore

it is sufficient to show that $T^{-1}\sum_{t=1}^{T}(\varepsilon_{yt}^2 - \mathbb{E}\varepsilon_{yt}^2)\alpha_p' z_t^-(z_t^-)'\alpha_p$ converges to zero. According to Davidson (1994, Theorem 19.7) this hold for our assumptions if $|(\varepsilon_{yt}^2 - \mathbb{E}\varepsilon_{yt}^2)(\alpha_p' z_t^-)^2|$ can be shown to be uniformly integrable (uniformly over t and p). Now $\mathbb{E}(\varepsilon_{yt}^2 - \mathbb{E}\varepsilon_{yt}^2)^2(\alpha_p' z_t^-)^4 = (\mathbb{E}(\varepsilon_{yt}^2 - (\mathbb{E}\varepsilon_{yt}^2))^2)(\mathbb{E}\alpha_p' z_t^-)^4$ due to the i.i.d. assumption on $(\varepsilon_t)_{t\in\mathbb{Z}}$. But $\mathbb{E}(\varepsilon_{yt}^2 - (\mathbb{E}\varepsilon_{yt}^2))^2 < \infty$ due to finite fourth moments. In order to show that $\sup_{p\in\mathbb{N}}\mathbb{E}(\alpha_p' z_t^-)^4 < \infty$ for $\sup_p \|\alpha_p\|_2 < \infty$ we use Lemma 2: Clearly $\alpha_p' z_t^- = \sum_{j=0}^{\infty}\phi_{p,j}^{\nu}\nu_{t-j} + \phi_{p,j}^{\varepsilon}\varepsilon_{t-j}$. Thus it suffices to show that $\sup_p \sum_{j=0}^{\infty}\|[\phi_{p,j}^{\nu}, \phi_{p,j}^{\varepsilon}]\|_2^2 < \infty$, which follows since $\sup_p \|\alpha_p\|_2$ is bounded by assumption and for each of y_t, z_{1t} and z_{2t} the summability assumption is easily verified. Uniform integrability then follows from Davidson (1994, Theorem 12.10.). It follows that (ii) holds.
Finally (i) holds since it is implied by ($\mathbf{I}(.)$ denoting the indicator function)

$$\sum_{t=1}^{T}\mathbb{E}\left[X_{Tt}^2\mathbf{I}(X_{Tt}^2 > \epsilon)\right] = T\mathbb{E}\left[X_{T1}^2\mathbf{I}(X_{T1}^2 > \epsilon)\right] \to 0$$

for each $\epsilon > 0$ (see Hall and Heyde, 1980, (3.6), p. 53). Here convergence is implied by $\mathbb{E}[\varepsilon_{y1}(\alpha_p' z_1)]^4 = \mathbb{E}\varepsilon_{y1}^4\mathbb{E}(\alpha_p' z_1)^4 < \infty$ as shown previously. This concludes the proof. □

B.4 Proof of Theorem 5

Proof: The proof of Theorem 5 combines the arguments from the proof of Theorems 3 and 4. Analogously to equation (24) we obtain

$$y_t = \sum_{j=1}^{p-1}\pi_j y_{t-j} + \sum_{j=1}^{p}\psi_j z_{2t-j} + \left(\sum_{j=1}^{p_{z1}+1}\psi_{z1j}\right)B^{-1}(Bz_{1t-p_{z1}-1}) + \sum_{j=1}^{p_{z1}}\psi_{z1j}(z_{1t-j} - z_{1t-p_{z1}-1}) + \varepsilon_{yt,p}$$

where $B := [\beta, \beta_\perp]$. Note that $z_{1t-j} - z_{1t-p_{z1}-1} = \sum_{i=j}^{p_{z1}}\Delta z_{1t-i} = \sum_{i=j}^{p_{z1}} x_{1t-i}$ is stationary for each $1 \le j < p_{z1}$. Define $z_{1t}^- := \left[z_{1t-1}' - (z_{1t-p_{z1}-1})', \ldots, z_{1t-p_{z1}}' - (z_{1t-p_{z1}-1})'\right]'$, $z_{2,t,p}^- := [(y_t^-)', (z_{2t}^-)', (\beta' z_{1t-p_{z1}-1})']'$ and $z_{3,t} := \beta_\perp' z_{1t-p_{z1}-1}$. Then in $z_t^- := [(z_{1,t}^-)', (z_{2,t,p}^-)', z_{3,t}']'$ the last coordinates (i.e. $z_{3,t}$) are fractionally integrated while the rest are stationary. Let $D_T := \text{diag}\left(T^{-1/2}I, T^{-(d_1+1)}, \ldots, T^{-(d_{c_{z1}}+1)}\right)$, $\hat{R}_T := D_T\sum_{t=p+1}^{T} z_t^-(z_t^-)'D_T$, and

$$R_T := \begin{bmatrix} \mathbb{E}z_{1,t}^-(z_{1,t}^-)' & \mathbb{E}z_{1,t}^-(z_{2,t}^-)' & 0 \\ \mathbb{E}z_{2,t}^-(z_{1,t}^-)' & \mathbb{E}z_{2,t}^-(z_{2,t}^-)' & 0 \\ 0 & 0 & [\hat{R}_T]_{3,3} \end{bmatrix}.$$

Obviously (27) holds with this choice. The uniform bound on the eigenvalues of R_T follows as in the proof of Theorem 4 and from

$$\operatorname{diag}\left(T^{-(d_1+1)},\ldots,T^{-(d_{c_{z1}}+1)}\right)\sum_{t=p+1}^{T} z_{3,t}z_{3,t}'\operatorname{diag}\left(T^{-(d_1+1)},\ldots,T^{-(d_{c_{z1}}+1)}\right)\xrightarrow{d}\Xi \quad (33)$$

where Ξ is a.s. positive definite by Lemma 3 (i). Consequently (26) holds.
Next we show that (28) also holds. $\hat{R}_T - R_T$ consists of six types of subblocks: The terms involving only $z_{1,t}^-$ and $z_{2,t}^-$ can be analyzed exactly as in the proof of Theorem 4, with $d_{\max} := \max(d_1,\ldots,d_{k_{z1}})$ replacing d: The upper bound on the increase of p as a function of T shows that the sum of squares of these entries is of order $O_P(p^{-1})$. The $(3,3)$ block of $\hat{R}_T - R_T$ is zero by definition. The remaining two terms include terms of the form $T^{-(d_r+3/2)}\sum_{t=p+1}^T [z_{3,t}]_r[(\beta' z_{1t-j})']_s = O_p(T^{\max(d_r+d_s,0)-dr-1/2})$ $T^{-(d_r+3/2)}\sum_{t=p+1}^T [z_{3,t}]_r[\Delta z_{1t-j}']_s = O_p(T^{\max(d_r+d_1,\ldots,d_r+d_{cz},0)-d_r-1/2})$ by Lemma 3 (iii). Both terms are $o_p(p^{-1/2})$ since $|d_s|,|d_r| < 0.5$ and, by Assumption P5 (iii), $p < T^{\min_s(1-2d_s,(1+2d_r)/3,1/3)}$ for $r = 1,\ldots,c_{z1}$ and $s = 1,\ldots,k_{z1}$. Likewise, defining $d_{r,0} := \max(0,d_r)$, it follows from Lemma 3 (ii) that[18]

$$\max_{0\le j\le H_T}\left\| T^{-d_r-3/2}\sum_{t=p+1}^{T}[z_{3,t}]_r[y_{t-j}', z_{2t-j}']\right\|_2 = O_P(T^{d_{r,0}-d_r-1/2}), \quad \text{for } H_T = o(T^{1/3}),\ r = 1,\ldots,c_{z1}.$$

Thus the sum over these terms is $O_P(pT^{d_{0,r}-d_r-1/2}) = o_p(p^{-1/2})$ since, by Assumption P5 (iii), $p < T^{1/3}$ (covers $0 \le d_r < 1/2$) and $p < T^{2/3(1/2+d_r)}$ (covers $-1/2 < d_r < 0$).

Further $\mathbb{E}[z_{3,t}]_r^2 = O(T^{2d_r+1})$ follows from Davidson and Hashimzade (2007). Thus (28) holds under the restrictions on p imposed in Assumption P5. From (33) it also follows that the contribution of this block to $\mathbb{E}\|l' D_T z_t^-\|_2^2$ is $O(1)$, showing (29). Finally the arguments to show (30) are analogous to those used in the proof of Theorem 4 since the nonstationary components are not involved. This concludes the proof. □

[18]The summability condition of Assumption P5 (i) implies the rate condition on $\theta_{w,j}$ in Lemma 3.

Table 1: Null rejection rates: local-to-unity models.

Method	sample size	Simulation DGP no-cointegration (13, $\delta = 0$)	$z_{1t}-$ adjusts; (14, $\delta = 0$)	y_t- adjusts; (15, $\delta' = (0,0)$)
	50	0.143	0.083	0.390
Toda-Phillips	100	0.130	0.072	0.326
	200	0.126	0.052	0.344
	500	0.130	0.049	0.332
	50	0.099	0.076	0.090
Surplus-VARX	100	0.058	0.057	0.052
	200	0.042	0.041	0.060
	500	0.066	0.057	0.059

Table entries show empirical rejection rates under the null hypothesis for a nominal 0.05% test. The DGP in Columns 3-5 is given by (13), (14), and (15), respectively, with $c = -5.0$ and $\delta = 0$. Panels 1-2 employ tests based on Toda and Phillips (1993) and the surplus-lag ARX(2,3), respectively.

Table 2: Null rejection rates: fractionally integrated models

Method		Simulation DGP: no cointegration (eq. 17, $\delta = 0$)		test cointegration (eq. 18, $\delta' = (0,0)$)	
		$d = 0.4$	$d = 0.8$	$d = 0.4$	$d = 0.8$
	50	0.128	0.106	0.126	0.120
Toda-	100	0.108	0.073	0.072	0.100
Phillips	200	0.082	0.082	0.042	0.100
	500	0.067	0.090	0.048	0.129
	50	0.098	0.125	0.103	0.126
Surplus-	100	0.092	0.093	0.087	0.080
VARX	200	0.078	0.064	0.061	0.069
	500	0.070	0.068	0.057	0.052

Table entries show empirical rejection rates under the null hypothesis for a nominal 5% test. The DGP is given by (16) & (17) in Columns 3-4 and by (16) & (18) in Columns 5-6. Panels 1-2 employ tests based on Toda and Phillips (1993) and the surplus-lag ARX(2,3), respectively.