Instrumental Variable and Variable Addition Based Inference in Predictive Regressions*

Jörg Breitung† and Matei Demetrescu‡

Companion working paper: April 1, 2015

Abstract

Valid inference in predictive regressions depends in a crucial manner on the degree of persistence of the predictor variables. The paper studies test procedures that are robust in the sense that their asymptotic null distributions are invariant to the persistence of the predictor, that is, the limiting distribution is the same irrespective of whether the regressors are stationary or (nearly) integrated. Existing procedures are often conservative (e.g. tests based on Bonferroni bounds), are based on highly restrictive assumptions (such as homoskedasticity or assuming an AR(1) process for the regressor) or fail to have power against alternatives in a $1/T$ neighborhood of the null hypothesis. We first propose a refinement of the variable addition method with improved asymptotic power approaching the optimal rate. Second, inference based on instrumental variables may further improve the (local) power of the test and even achieve local power under the optimal $^1/_T$ rate. We give high-level conditions under which the suggested variable addition and instrumental variable procedures are valid no matter whether the predictor is stationary, near-integrated or integrated, or exhibits time-varying volatility. All test statistics possess a standard limiting distribution. Monte Carlo experiments suggest that tests based on simple combinations of instruments perform most promising relative to existing tests. An application to quarterly U.S. stock returns illustrates the need for robust inference.

Key words

Causality test, persistence, integration, long memory, IV estimation

JEL classification

C12 (Hypothesis Testing), C32 (Time-Series Models)

*The authors would like to thank two anonymous referees, as well as Christoph Hanck, Uwe Hassler, Søren Johansen, Lynda Khalaf, Helmut Lütkepohl, Tassos Magdalinos, Massimiliano Marcellino, Dominik Poß, Paulo M.M. Rodrigues, Michalis Stamatogiannis, Rob Taylor and Martin Wagner for very helpful comments and suggestions.

†University of Cologne, Institute of Econometrics and Statistics, Albertus-Magnus-Platz D-50923 Köln

‡*Corresponding author:* Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany. E-mail address: `mdeme@stat-econ.uni-kiel.de`

1 Introduction

Predictive regressions play an important role in empirical economics. For instance, the concept of Granger causality implies that some variable is not considered as a cause for another variable if the former variable is not able to predict the latter. In financial economics it is of interest whether variables like dividend yields or interest spreads contain information about future stock price returns (e.g. Campbell and Shiller, 1988). Other examples include tests of the rational expectation hypothesis. For example, the uncovered interest rate parity [UIP] hypothesis implies that $\mathrm{E}_t(s_{t+1}) = i_t - i_t^*$, where $\mathrm{E}_t(s_{t+1})$ denotes the expected change of future exchange rates and i_t (i_t^*) represent conformable domestic (foreign) interest rates. This hypothesis can be cast in the predictive regression framework by running a regression of the dependent variable $y_t = s_{t+1} - i_t + i_t^*$ on the interest rate differential $x_t = i_t - i_t^*$ (and possibly some other variables). If the UIP hypothesis holds, interest rate differentials should not be able to predict y_t, but in practice it is often found that the coefficient on interest rate differentials is significantly negative (e.g. Froot and Thaler, 1990).

An important practical problem with performing such predictive regressions is that in many cases the regressor is highly persistent, whereas the dependent variable is close to white noise. For example, stock price returns or exchange rate changes are approximately white noise, whereas predictors like dividend yields or interest rate differentials behave approximately like a random walk. As shown by Elliott and Stock (1994) the t-statistic may suffer from severe size distortions in such cases.

We start within the framework of Elliott and Stock and consider as a baseline model the dynamic system given by the triangular representation

$$y_t = \beta x_{t-1} + u_t \quad (1)$$

$$x_t = \varrho x_{t-1} + v_t, \quad (2)$$

$t = 2, \ldots, T$, with

$$\Sigma = \mathrm{E}\left(\begin{pmatrix} u_t \\ v_t \end{pmatrix}\begin{pmatrix} u_t & v_t \end{pmatrix}\right) = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}.$$

Note that the regressor x_t is assumed to be weakly exogenous since $\mathrm{E}(x_{t-1}u_t) = 0$ but $\mathrm{E}(x_{t-1}u_{t-1}) \neq 0$ whenever $\sigma_{uv} \neq 0$. If $\sigma_{uv} = 0$, then the regressor is strictly exogenous. We first abstract from any deterministic component such as an intercept or linear trend to focus on the main issues without the extra notational burden. In Section 3.2 we expand our model accordingly and show that deterministic terms can be easily dealt with in the usual manner.

To model persistent regressors, the variable x_t is often assumed to be nearly integrated such that

$$\varrho = 1 - \frac{c}{T} \tag{3}$$

for $c \geq 0$. We are interested in testing the null hypothesis $\beta = 0$ whatever the value of c may be. Under suitable regularity conditions (e.g. Elliott and Stock, 1994) the ordinary least squares [OLS] t-statistic for the hypothesis $\beta = 0$ in (1) is asymptotically distributed as

$$t_{ls} \xrightarrow{d} \omega \frac{\int_0^1 J_c(r)\,\mathrm{d}W_v(r)}{\sqrt{\int_0^1 J_c^2(r)\,\mathrm{d}r}} + \sqrt{1-\omega^2}\mathcal{Z}\,, \tag{4}$$

where $\omega = \sigma_{uv}/(\sigma_u\sigma_v)$, $J_c(r)$ represents an Ornstein-Uhlenbeck process such that $T^{-1/2}x_{[rT]} \Rightarrow \sigma_v J_c(r)$ with $J_c(r) = W_v(r) - c\int_0^r e^{-c(r-s)}W_v(s)\,\mathrm{d}s$ and $W_v(r)$ a standard Brownian motion obtained as $T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \sigma_v W_v(r)$. The standard normal random variable $\mathcal{Z}$ is independent of $W_v(r)$ (and thus of $J_c(r)$ as well). Hence, the actual distribution of the ordinary t-statistic is nonstandard and depends on the parameter c whenever $\sigma_{uv} \neq 0$.

Should the driving process x_t be stationary, i.e. $-1 < \varrho < 1$ fixed, standard asymptotic theory applies. The problem in applied research is that the nature of x_t is typically unknown, and pre-testing to check whether $|\varrho| = 1$ has been shown to induce serious size-distortions when ϱ is close to unity (Elliott and Stock, 1994).

For the baseline model given by (1) with x_t generated as in (2) there already exist a number of test procedures that are robust to the value of of the autoregressive coefficient ϱ. Elliott and Stock (1994) proposed a Bayesian mixture procedure, Cavanagh et al. (1995) consider various tests based on conservative bounds (as refined by Campbell and Yogo, 2006). The work of Jansson and Moreira (2006) and Chen and Deo (2009) can be casted in a restricted likelihood framework. The asymptotic analysis of these procedures is however confined to the near-integrated case.

Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) proposed testing strategies that enable applied researchers to conduct valid inference irrespective of the nature of the autoregressive roots (local to unity or stationary) of the examined time series. The idea is to augment the testing equation with additional (redundant) variables such that the coefficients to be tested are attached to stationary variables. Bauer and Maynard (2012) show that variable addition [VA] also works in the context of VAR(∞) processes with unknown persistence. Although such a robust approach is very appealing, we argue that this class of tests may suffer from a dramatic loss of power. Specifically, they only have power in $1/\sqrt{T}$ neighborhoods of the null hypothesis compared with the typical rate of $1/T$ for tests involving nearly integrated regressors.[1] The shortcoming is shared

[1] The variable addition approach of Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) is much more general and may perform more favorable in other applications such as testing for causality in cointegrated systems.

to some extent by the nonparametric approach of Maynard and Shimotsu (2009) with a local power characterized by the rate $1/T^{0.75}$ (see their Lemma 9). Gorodnichenko et al. (2012) propose a quasi-differencing procedure applicable, like the VA method of Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996), in general dynamic models; but, like VA, the procedure only has power in $1/\sqrt{T}$ neighborhoods of the null.[2] Finally, Phillips and Magdalinos (2009) propose an instrumental variable [IV] procedure with local power arbitrarily close to the optimal $1/T$ of the size-distorted OLS estimator.

We therefore study strategies to conduct inference in the presence of regressors with unknown persistence such that the power of the resulting tests remains close to that of an optimal test, while the limiting null distributions do not change with the degree of persistence of the regressors. Specifically, we generalize VA and IV procedures and provide classes of tests which exhibit similarities with the IVX approach of Phillips and Magdalinos (2009).

This paper's contributions are as follows. We consider a model with conditional and unconditional heteroskedasticity as well as short-run dynamics of the predictor and show in Section 2 that the original variable addition test may suffer from severe loss of (asymptotic) power. Although we demonstrate that the power of the VA procedure can be substantially improved by employing certain transformations of the involved variables, some loss of power remains. We then develop alternative test procedures based on instrumental variables that share with the VA tests the invariance to the persistence of the predictor. At the same time, an appropriate choice of instruments yields tests with power against a sequence of alternatives converging to the null hypothesis at the optimal rate. Moreover, the instruments we propose do not require additional data. In Section 3, we study the possibility of improving inference in the IV setup by combining different instruments. Our methods can easily be extended to deal with deterministic components and an arbitrary number of regressors. Section 4 compares the proposed methods with existing alternatives by means of Monte Carlo experiments. Section 5 illustrates the proposed methods with U.S. stock returns, and the final section concludes.

Before proceeding to the main part of the paper, let us introduce some notation. Boldface characters denote vectors. The lag operator is denoted by L, the fractional difference operator $(1-L)^d = \Delta^d = \sum_{j\geq 0} \delta_j L^j$ is given by the usual series expansion, and its truncated version, $\mathbb{I}(t>0)\,\Delta^d$ with $\mathbb{I}$ the indicator function, is denoted by Δ^d_+ such that $\Delta^d_+ x_t = \sum_{j=0}^{t-1} \delta_j x_{t-j}$. The coefficients δ_j are known to decay at a hyperbolic rate, $\delta_j \sim j^{-d-1}$ as $j \to \infty$. Finally, "$\Rightarrow$" denotes weak convergence in a space of càdlàg

[2]It should be noted, however, that these methods are designed to work against a wider range of alternatives than we consider in (1). As argued by Lettau and Van Nieuwerburgh (2008) and Maynard and Shimotsu (2009), it is the stationary component of the predictor that matters for forecasting series like stock price returns. Accordingly, the rate of the sequence of local alternatives may be a misleading guide for assessing the power against economically relevant alternatives.

functions on $[0,1]$ endowed with a suitable norm.

2 Variable addition and instrument variable tests

We first extend the baseline model to allow for more plausible characteristics.

Assumption 1 *Let*

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} \bar{u}_t \\ \sum\limits_{j\geq 0} b_j \bar{v}_{t-j} \end{pmatrix}$$

where $\sum_{j\geq 0} j\,|b_j| < \infty$ *and* $\sum_{j\geq 0} b_j \neq 0$*, the innovations* $\bar{u}_t$ *and* $\bar{v}_t$ *are a bivariate white noise sequence with a component structure,*

$$\begin{pmatrix} \bar{u}_t \\ \bar{v}_t \end{pmatrix} = H\left(\frac{t}{T}\right) \begin{pmatrix} \widetilde{u}_t \\ \widetilde{v}_t \end{pmatrix}$$

such that $H(r)$ *is a* 2×2 *matrix of piecewise Lipschitz functions, invertible for all* $r \in [0,1]$*, and* $(\widetilde{u}_t, \widetilde{v}_t)'$ *is a martingale difference sequence with identity covariance matrix satisfying* $\sup_t \left| T^{-1} \sum_{j\geq 1} \sum_{k\geq 1} \mathrm{E}\left(\widetilde{v}_{t-j}\widetilde{v}_{t-k}\widetilde{v}_t^2\right)\right| < \infty$ *as well as* $\sup_t \mathrm{E}\left(\left\|(\widetilde{u}_t, \widetilde{v}_t)'\right\|^{4+\epsilon}\right) < \infty$ *for some* $\epsilon > 0$.

The assumption allows the increments of the predictor process x_t to be serially correlated. The so-called 1-summability condition of the moving average coefficients is standard in the literature on integrated series. The disturbances u_t are uncorrelated with the increments of x_t at all lags (i.e. x_t is weakly exogenous with respect to u_t). The martingale difference assumption for the innovations is natural for the empirical applications we have in mind and allows for conditional heteroskedasticity. The summability condition on the cross-product moments $\mathrm{E}\left(v_{t-j}v_{t-k}v_t^2\right)$ slightly restricts the serial dependence in the conditional variances and is fulfilled by independent sequences, for instance.

At the same time, unconditional time heteroskedasticity is captured by the matrix $H(r)$ since

$$\mathrm{E}\left(\begin{pmatrix} \bar{u}_t \\ \bar{v}_t \end{pmatrix} \begin{pmatrix} \bar{v}_t & \bar{v}_t \end{pmatrix}\right) = H\left(\frac{t}{T}\right) H\left(\frac{t}{T}\right)'.$$

Should $H(r)$ be a diagonal matrix for all $r \in (0,1]$, the innovations $\bar{u}_t$ and $\bar{v}_t$ may have time-varying variance but are uncorrelated. In general, the off-diagonal element of $H(r)H(r)'$ is not restricted to zero, allowing for contemporaneous and time varying correlation among the innovations. A leading case of such time heteroskedasticity are breaks in the variance or correlation, captured by piecewise constant elements of $H(r)$.[3]

[3]Note that a break in the covariance need not imply a break in the correlation; but constant-correlation models imply restrictions on the matrix H which we do not impose.

Such changing variances and covariances are quite plausible with macroeconomic and financial data; see e.g. Stock and Watson (2002), Sensier and van Dijk (2004) or Clark (2009).

Under this assumption, an invariance principle holds (see e.g. Cavaliere et al., 2010)

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix} u_t \\ v_t \end{pmatrix} \Rightarrow \begin{pmatrix} B_{H,u}(r) \\ B_{H,v}(r) \end{pmatrix},$$

where

$$\begin{pmatrix} B_{H,u}(r) \\ B_{H,v}(r) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \sum_{j\geq 0} b_j \end{pmatrix} \int_0^r H(s)\mathrm{d}\mathbf{W}(\mathrm{s})$$

with $\mathbf{W}(s)$ a vector of two independent standard Wiener processes. The processes $B_{H,u}(r)$ and $B_{H,v}(r)$ are individually time-transformed Brownian motions whose correlation may also vary in time. Concretely, their covariance at time r is given by $\sum_{j\geq 0} b_j \int_0^r H(s)H'(s)\mathrm{d}s$. See Cavaliere et al. (2010) for a more detailed discussion. Furthermore we have

$$\frac{1}{\sqrt{T}} x_{[rT]} \Rightarrow J_{c,H,v}(r) \quad \text{where} \quad J_{c,H,v}(r) = B_{H,v}(r) - c\int_0^r e^{-c(r-s)} B_{H,v}(s)\mathrm{d}s$$

i.e. the Ornstein-Uhlenbeck process driven by the time-transformed Brownian motion $B_{H,v}(r)$; cf. Phillips (1987) and Cavaliere (2004). In the case of weak stationarity of $(u_t, v_t)'$ (i.e. $H(r) = H$ almost everywhere), we essentially recover the baseline model with $B_{H,u} = \sigma_u W_u$, where W_u is standard Wiener process with correlation $\omega = \mathrm{E}\left(W_u(1)W_v(1)\right)$, $B_{H,v} = \sigma_v \left(\sum_{j\geq 0} b_j\right) W_v$ and $J_{c,H,v} = J_c$.

2.1 Variable addition-based inference

Building on Choi's (1993) results, Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) proposed a simple approach to robustify Granger causality tests by including redundant regressors. The approach can be adapted to our situation[4] by testing the hypothesis $\beta = 0$ in the regression

$$y_t = \beta x_{t-1} + \phi x_{t-2} + u_t \tag{5}$$

$$= \beta v_{t-1} + \psi x_{t-2} + u_t\ , \tag{6}$$

where under the null hypothesis $\beta = 0$ and $\psi = \beta\varrho + \phi = 0$. The restriction on ψ is however ignored when testing the hypothesis. It is well known (e.g. Sims et al., 1990) that tests of coefficients attached to stationary variables (here: $v_{t-1} = x_{t-1} - \varrho x_{t-2}$)

[4]Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) considered the case that the variables are $I(1)$ (i.e. $c = 0$), whereas Bauer and Maynard (2012) present some results for the VA method under the local-to-unity assumptions with $c < 0$.

possess standard (normal or χ^2) limiting distributions. Thus, the t-statistic for testing the null $\beta = 0$ in (6) is asymptotically normally distributed no matter what value ϱ has.

In the present framework, this test suffers from a severe loss of power resulting from the fact that the highly persistent regressor x_{t-1} is essentially replaced by the white noise series v_{t-1} in (6); see Theorem 1 below. To circumvent the problem, we suggest a modified variable addition approach, where x_{t-1} is replaced by a (highly persistent) process. As an example consider the decomposition

$$\begin{aligned} x_t &= (1-\alpha L)_+^{-1}\Delta x_t + (1-\alpha)(1-\alpha L)_+^{-1} x_{t-1} \\ &:= z_t + \zeta_t, \end{aligned} \tag{7}$$

where $z_t = \Delta x_t + \alpha \Delta x_{t-1} + \cdots + \alpha^{t-2}\Delta x_2$ with $|\alpha| < 1$. The advantage of using (7) instead of the decomposition $x_t = \Delta x_t + x_{t-1}$ (i.e. $\alpha = 0$) employed in the original VA approach is that $z_{t-1} = (1-\alpha L)_+^{-1}\Delta x_{t-1}$ more closely mimics the behavior of x_{t-1} if α is close to one.

In what follows we generalize this approach by considering the augmented test regression of the form

$$y_t = \beta z_{t-1} + \psi \zeta_{t-1} + u_t, \tag{8}$$

where z_{t-1} is suitably chosen (see the assumption below) and $\zeta_{t-1} = x_{t-1} - z_{t-1}$. As before we ignore the restriction $\psi = \beta$ (i.e. $\phi = 0$) when testing the null hypothesis $\beta = 0$ by employing the familiar t-statistic for $\beta = 0$.

By constructing z_t such that it is indeed less persistent than the regressor x_t, standard inference can be conducted using the OLS estimator of β in (8). At the same time, z_t should be as persistent as possible to enhance the power properties. The following assumption summarizes the properties required for the variable z_t.

Assumption 2 *For some $0 \le \delta < 1/2$, it holds that*

$$V_{T,z} := \frac{1}{T^{1+2\delta}} \sum_{t=2}^{T} z_{t-1}^2 \Rightarrow V_z \qquad \text{and} \qquad V_{T,zu} := \frac{1}{T^{1+2\delta}} \sum_{t=2}^{T} z_{t-1}^2 u_t^2 \Rightarrow V_{zu}$$

where V_z and V_{zu} are positive and stochastically bounded,

$$(i) \quad \frac{1}{T^{1.5+\delta}} \sum_{t=2}^{T} z_{t-1} x_{t-1} \xrightarrow{p} 0,$$

and

$$(ii) \quad \frac{1}{T^{1/2+\delta}\sqrt{V_{zu}}} \sum_{t=2}^{T} z_{t-1} u_t \Rightarrow \mathcal{Z}$$

with $\mathcal{Z} \sim \mathcal{N}(0,1)$.

Note that condition (i) is satisfied by stationary variables, e.g. $z_t = \Delta x_t$ in the original VA approach. In contrast, it is obviously violated by letting $z_t = x_t$ whenever x_t is a (nearly) integrated process. Condition (i) may thus be interpreted as quantifying how much "less persistent" the variable z_t should be for the modified VA to work. The other assumptions are high-level regularity conditions ensuring e.g. the application of a central limit theorem.

When allowing for heteroskedasticity, Eicker-White standard errors (Eicker, 1967; White, 1980) need to be employed. The respective t-statistic is given by

$$t_{va}^{w} = \frac{\widehat{\beta}_{va} - \beta_0}{s.e_w\left(\widehat{\beta}_{va}\right)}, \tag{9}$$

where $\widehat{\beta}_{va}$ is the OLS estimate of β in (8) and

$$s.e_w^2\left(\widehat{\beta}_{va}\right) = \frac{\sum\limits_{t=2}^{T} \widetilde{z}_{t-1}^2 \widehat{u}_t^2}{\left(\sum\limits_{t=2}^{T} \widetilde{z}_{t-1}^2\right)^2}$$

where $\widetilde{z}_{t-1}$ is the residual from a projection on ζ_{t-1},

$$\widetilde{z}_{t-1} = z_{t-1} - \zeta_{t-1}\left(\sum_{t=2}^{T} \zeta_{t-1}^2\right)^{-1} \sum_{t=2}^{T} \zeta_{t-1} z_{t-1},$$

and $\widehat{u}_t$ denotes the OLS residual $\widehat{u}_t = y_t - \widehat{\beta}_{ls} x_{t-1}$ from (1).

The obvious question is how to construct the less persistent proxy z_t from the regressor x_t. Clearly, Assumption 2 allows for a wide range of candidate variables. For instance any (asymptotically) stationary variable is admissible, but stationarity is not necessary. In what follows we consider the following "natural" examples:

1. **Short memory:** $z_{t-1} = (1 - \bar{\alpha}L)_+^{-1} \Delta x_{t-1} = \Delta x_{t-1} + \bar{\alpha}\Delta x_{t-2} + \cdots + \bar{\alpha}^{t-2}\Delta x_1$ with $|\bar{\alpha}| < 1$ and thus $\delta = 0$.

2. **Mild integration:** $z_{t-1} = (1-\alpha_T L)_+^{-1} \Delta x_{t-1}$ for $\alpha_T = 1 - a/T^{\eta}$, where $a > 0$, $0 < \eta < 1$ leading to $\delta = \eta/2$.

3. **Fractional integration:** $z_{t-1} = \Delta_+^{1-d^*} x_{t-1}$ for some $d^* \in (0, 1/2)$, with $\delta = 0$.

4. **Long differences:** $z_{t-1} = x_{t-1} - x_{t-k_T}$ for $k_T = \min\{[KT^{\nu}], t-1\}$ for some $0 < \nu < 1$, with $\delta = \nu/2$.

The null distribution (i.e. for $b = 0$) and the local power (for $b \neq 0$) for the class of test statistics characterized by Assumption 2 are given in the following theorem.

Theorem 1 *Let y_t and x_t be generated as in* (1) *and* (2) *with $\varrho = 1 - {}^{c}/_{T}$. Under Assumptions 1 – 2 and the local alternative $\beta = {}^{b}/_{T^{\delta+0.5}}$, the statistic t^w_{va} defined in* (9) *has the limiting distribution*

$$t^w_{va} \xrightarrow{d} b\frac{V_z}{\sqrt{V_{zu}}} + \mathcal{Z}\,,$$

where $\mathcal{Z}$ represents a standard normal random variable.

Proof: *See the Appendix.*

Remark 1 *In the homoskedastic case where* $\mathrm{Var}\,(u_t) = \sigma^2_u\ \forall t$ *and $V_{zu} = \sigma^2_u V_z$, the ordinary t-statistic based on the standard errors*

$$s.e\left(\widehat{\beta}_{va}\right) = \widehat{\sigma}_u \left(\sum_{t=2}^{T} z^2_{t-1} - \sum_{t=2}^{T} z_{t-1}\zeta_{t-1}\left(\sum_{t=2}^{T}\zeta^2_{t-1}\right)^{-1}\sum_{t=2}^{T}\zeta_{t-1}z_{t-1}\right)^{-1/2}$$

may be used. Under the conditions of Theorem 1 the limiting distribution of $t_{va} = \left(\widehat{\beta}_{va} - \beta_0\right)/s.e\left(\widehat{\beta}_{va}\right)$ results as

$$t_{va} \xrightarrow{d} b\ \frac{\sqrt{V_z}}{\sigma_u} + \mathcal{Z},$$

that is, V_{zu} is replaced by $\sigma^2_u V_z$, as discussed in the Appendix.

It is important to note that the local power of the VA test does not depend on the persistence of the regressor (represented by the parameter c). The following corollary provides the details for the examples considered above. For simplicity we focus on i.i.d. innovations.

Corollary 1 *Let $(u_t, v_t)' \overset{iid}{\sim} (\mathbf{0}, \Sigma)$ with finite moments of order $4 + \epsilon$ for some $\epsilon > 0$ implying $B_{H,u} = \sigma_u W_u$, $B_{H,v} = \sigma_v W_v$ and $J_{c,H,v} = \sigma_v J_c$. Furthermore, let $G(r)$ be a standard Wiener process independent of W_u and W_v.*

1. *If $z_{t-1} = (1 - \bar{\alpha}L)^{-1}_{+}\Delta x_{t-1}$, $|\bar{\alpha}| < 1$ fixed, we have $V_{zu} = \sigma^2_u V_z$, $V_z = \frac{\sigma^2_v}{1-\bar{\alpha}^2}$ and $\delta = 0$. For $\beta = {}^{b}/\sqrt{T}$, it follows that*

$$t^w_{va} \xrightarrow{d} b\,\frac{\sigma_v}{\sigma_u\sqrt{1-\bar{\alpha}^2}} + G(1).$$

2. *If $z_{t-1} = (1 - \alpha_T L)^{-1}_{+}\Delta x_{t-1}$, with $\alpha_T = 1 - {}^{a}/_{T^{\eta}}$ then $V_{zu} = \sigma^2_u V_z$, $V_z = \frac{\sigma^2_v}{2a}$ and $\delta = {}^{\eta}/2$. For $\beta = {}^{b}/_{T^{1/2+\eta/2}}$, it follows that*

$$t^w_{va} \xrightarrow{d} b\,\frac{\sigma_v}{\sigma_u\sqrt{2a}} + G(1).$$

3. *If* $z_{t-1} = \Delta_+^{1-d^*} x_{t-1}$ *for some* $d^* \in (0, 1/2)$*, then* $V_{zu} = \sigma_u^2 V_z$*,* $V_z = \sigma_v^2 \frac{\Gamma(1-2d^*)}{\Gamma^2(1-d^*)}$ *and* $\delta = 0$*. For* $\beta = b/T^{1/2}$*, it follows that*

$$t_{va}^w \xrightarrow{d} b \frac{\sigma_v}{\sigma_u} \sqrt{\frac{\Gamma(1-2d^*)}{\Gamma^2(1-d^*)}} + G(1).$$

4. *If* $z_{t-1} = x_{t-1} - x_{t-k_T}$*, with* $k_T = \min\{[KT^\nu], t-1\}$ *then* $V_{zu} = \sigma_u^2 V_z$*,* $V_z = K\sigma_v^2$ *and* $\delta = \nu/2$*. For* $\beta = b/T^{1/2+\nu/2}$*, it follows that*

$$t_{va}^w \xrightarrow{d} b \frac{\sigma_v}{\sigma_u} \sqrt{K} + G(1).$$

Proof: *See the Appendix.*

In *case 1* of a short-memory VA scheme (i.e. if $\bar{\alpha}$ is a fixed parameter) the local power of the test is monotonically increasing in $|\bar{\alpha}|$. It follows that choosing $\bar{\alpha}$ close to unity can to some extent close the power gap between the original VA test (obtained by setting $\bar{\alpha} = 0$) and the power of the (infeasible) t test based on (4).

If $\alpha_T = 1 - a/T^\eta$, x_t is mildly (or moderately) integrated in the sense of Phillips and Magdalinos (2007) and Magdalinos and Phillips (2009). Since $0 < \eta < 1$ and α_T converges slowly to one, one can account for the trade-off between size control and power.

When z_t has (stationary) long memory (*case 3*), the local power of the test is monotonically increasing in d^* and can again enhance the power properties of the VA procedure as $\Gamma(1-2d^*) \to \infty$ for $d^* \to 1/2$. The test fails however to possess power against alternatives vanishing faster than $T^{-1/2}$.

Finally, the long difference VA scheme (*case 4*) has power against local alternatives with an improved rate $T^{-(1+\nu)/2}$, which is faster than the rate for the long memory variant and the same as the mildly integrated regressor for $\nu = \eta$. Note that the same trade-off as for the moderately integrated case arises.

Since the innovations $(u_t, v_t)'$ have been restricted for Corollary 1 to be conditionally homoskedastic, the ordinary standard errors might have been used yielding a simpler and asymptotically equivalent test statistic.

It is interesting to note that the limiting distribution of $T^{1/2+\delta}\left(\widehat{\beta}_{va} - \beta\right)$ is not normal in general. It follows from the proof of Theorem 1 that the leading term in the numerator is $T^{-1/2-\delta} \sum_{t=2}^T z_{t-1} u_t$ which, upon normalization with $\sqrt{V_{zu}}$, is normally distributed. The denominator, however, has as leading term $T^{1+2\delta} \sum_{t=2}^T z_{t-1}^2$ whose limit V_z need not be deterministic and therefore $\widehat{\beta}_{va}$ need not be asymptotically normally distributed. For the examples we consider in Corollary 1, V_z and V_{zu} are indeed deterministic and the respective asymptotic distribution of $\widehat{\beta}_{va}$ is asymptotically normal. Note however the reduced convergence rate compared to the T-consistent OLS estimator.

2.2 Inference based on instrumental variables

Following Phillips and Magdalinos (2009) we now consider test statistics based on instrumental variable (IV) estimators given by

$$\widehat{\beta}_{iv} = \frac{\sum\limits_{t=2}^{T} z_{t-1} y_t}{\sum\limits_{t=2}^{T} z_{t-1} x_{t-1}} \tag{10}$$

and the associated t-statistic using Eicker-White heteroskedasticity consistent standard errors

$$t_{iv}^{w} = \frac{\sum\limits_{t=2}^{T} z_{t-1} y_t}{\sqrt{\sum\limits_{t=2}^{T} z_{t-1}^2 \widehat{u}_t^2}} \, , \tag{11}$$

where $\widehat{u}_t = y_t - \widehat{\beta}_{ls} x_{t-1}$ and $\widehat{\beta}_{ls}$ is the OLS estimator of β in (1).[5]

Comparing this testing strategy with the VA approach proposed in the previous section it is apparent that the IV test neglects the remainder ζ_{t-1} in (8) when instrumenting x_{t-1} by the variable z_{t-1}. Note also that the VA test considered in Theorem 1 is equivalent to an IV test when using the residual of a regression of z_{t-1} on ζ_{t-1} as instrument. This suggests that valid instruments should obey Assumption 2. In order to describe the behavior of the IV statistic under sequences of local alternatives, we require a slightly stronger version of Assumption 2:

Assumption 3 *(Type-I instruments) For some $\delta \geq 0$ and $0 \leq \vartheta < 1/2$, it holds that*

$$V_{T,zu} \; := \; \frac{1}{T^{1+2\delta}} \sum_{t=2}^{T} z_{t-1}^2 u_t^2 \; \Rightarrow \; V_{zu} \, ,$$

where V_{zu} is positive and stochastically bounded, and the vector $(z_t, z_{t-1}u_t, u_t, v_t)'$ satisfies the invariance principle

$$\begin{pmatrix} \frac{1}{T^{1/2+\vartheta+\delta}} \sum_{t=1}^{[rT]} z_t \\ \frac{1}{T^{1/2+\delta}\sqrt{V_{T,zu}}} \sum_{t=2}^{[rT]} z_{t-1} u_t \\ \frac{1}{T^{1/2}} \sum_{t=1}^{[rT]} u_t \\ \frac{1}{T^{1/2}} \sum_{t=1}^{[rT]} v_t \end{pmatrix} \Rightarrow \begin{pmatrix} Z(r) \\ G_I(r) \\ B_{H,u}(r) \\ B_{H,v}(r) \end{pmatrix}$$

[5] As the OLS estimator is T-consistent, the residuals involve a smaller estimation error than the residuals employing the IV estimator. Note that asymptotic properties of the test are not affected by the estimation error of the residuals, as long as the estimator for β is consistent at a rate higher than $\sqrt{T}$.

with $Z(r)$ and $G_I(r)$ càdlàg processes such that $T^{-1-\delta-\vartheta}\sum_{t=1}^{T}\left(\sum_{j=1}^{t-1} z_j\right) v_t \Rightarrow \int_0^1 Z(r)\mathrm{d}B_{H,v}(r)+\lambda$ for some real λ and $G_I(1)\sim\mathcal{N}(0,1)$.

The additional parameter ϑ is closely related to the mean reverting behavior of the instrument z_t and crucially affects the local power of the IV-based test. This is in stark contrast to the modified variable addition approach, where only δ plays a role; see Theorem 2 below.

For $z_{t-1} = \Delta x_{t-1}$, the parameter λ equals the one-sided long-run variance of v_t, $\lambda = \sum_{j\geq 1} \mathrm{E}\left(v_t v_{t-j}\right)$. The parameter λ is relevant only under the alternative hypothesis and appears whenever v_t is correlated with the lags of z_t. Should v_t be a martingale difference sequence and z_{t-1} in the respective information set, then $\lambda = 0$ and the weak limit typically holds.

The interpretation of z_{t-1} is, like in the VA approach, that of an instrument having a lower persistence than the regressor x_{t-1}. The weak convergence $\frac{1}{T^{1/2+\vartheta+\delta}}\sum_{j=1}^{[rT]} z_j \Rightarrow Z(r)$ replaces condition (ii) of Assumption 2. In fact it implies it, as argued in the proof of Corollary 1, so it can be interpreted as a quantification of "reduced persistence" of the variable z_t as well.

Assumption 2 limits the persistence of the instruments. It is however possible to employ instruments with a similar persistence as x_{t-1}. In this case we require the instruments to be strictly exogenous with respect to u_t. The corresponding IV estimators actually achieve optimal convergence rates while still allowing us to draw robust inference on β.

Assumption 4 *(Type-II instruments) For some $\delta \geq 0$, the sequence z_t satisfies the invariance principle*

$$\frac{1}{T^\delta} z_{[rT]} \;\Rightarrow\; \dot{Z}(r)$$

jointly with weak convergence of $T^{-1/2}\sum_{t=1}^{[rT]}(u_t, v_t)'$, such that

$$V_{T,zu} \;:=\; \frac{1}{T^{1+2\delta}}\sum_{t=2}^{T} z_{t-1}^2 u_t^2 \;\Rightarrow\; V_{zu}$$

and

$$\frac{1}{T^{1/2+\delta}}\sum_{t=1}^{T} z_{t-1}u_t \Rightarrow \int_0^1 \dot{Z}(r)\mathrm{d}B_{H,u}(r) \sim \mathcal{MN}\left(0, V_{zu}\right),$$

where $\mathcal{MN}(0, V_{zu})$ denotes a mixed normally distributed random variable with expectation zero and (stochastic) covariance matrix V_{zu}.

We call instruments obeying Assumption 3 *type-I instruments*, and instruments obeying Assumption 4 *type-II instruments.* Type-I instruments are allowed to be weakly exogenous with respect to u_t but are less persistent than x_t. Type-II instruments include nonstationary variables as well as deterministic functions of time. For a type-II instrument it is basically required that the limiting process $\dot{Z}(r)$ is independent of the disturbances in (1). In a nutshell, valid instruments should be either free of stochastic trends (Assumption 3) or strictly exogenous (Assumption 4).

It may be surprising to learn that instruments like a trend or a trigonometric function turn out to be very powerful instruments although such instruments do not reveal any specific information about the regressor. The key insight is that, whenever the continuous-time limit of the suitably normalized predictor possesses a Loève-Karhunen expansion (see e.g. Phillips, 1998), type-II instruments correlate with elements of this expansion and thus the R^2 from a regression of x_t on z_t is larger than zero.

Some examples of readily available type-II instruments are the following:

1. **Independent random walk:** $z_t = (1-L)_+^{-1} w_t$, where w_t is an $iid(0, \sigma_w^2)$ sequence, independent of u_t and v_t (and whose use can be traced back to Park, 1990);

2. **Deterministic functions:** For example $z_t = t$ or $z_t = \sin(\pi t/2T)$ (cf. Phillips, 1998);

3. **Cauchy instrument:** $z_t = \text{sign}\,(x_t)$ (cf. So and Shin, 1999).

Remark 2 *The two types of instruments are formally encompassed in the more general class satisfying*

(i) the normalized partial sums of z_t possess a weak limit,

$$\frac{1}{T^{1/2+\vartheta+\delta}} \sum_{j=2}^{[rT]} z_{j-1} \Rightarrow Z\,(r)$$

jointly with weak convergence of $T^{-1/2} \sum_{t=1}^{[rT]} (u_t, v_t)'$ such that

$$V_{T,zu} \;:=\; \frac{1}{T^{1+2\delta}} \sum_{t=2}^{T} z_{t-1}^2 u_t^2 \;\Rightarrow\; V_{zu}$$

and

$$\frac{1}{T^{1+\delta+\vartheta}} \sum_{t=2}^{T} \left(\sum_{j=1}^{t-1} z_j \right) v_t \Rightarrow \int_0^1 Z(r) \mathrm{d}B_{H,v}(r) + \lambda$$

(ii) there exists a càdlàg process $X(s)$ such that

$$\frac{1}{\sqrt{T^{2\delta+1}}} \sum_{t=2}^{[rT]} z_{t-1} u_t \Rightarrow \int_0^r \mathcal{F}\,(Z(r))\,\mathrm{d}X\,(r)$$

with $\mathcal{F}\,(\cdot)$ a functional depending on the nature of the instrument, where *(iii)* *the functional* $\int_0^1 \mathcal{F}(Z(r))\,\mathrm{d}X(r)$ *has a zero-mean mixed Gaussian distribution with variance* V_{zu}.

For type-I instruments, $\mathcal{F}(Z(r)) = \sqrt{V_{zu}}$ *does not depend on* r, *and* $X(r) = G(r)$ *almost everywhere. In the case of type-II instruments,* $Z(r) = \int_0^r \dot{Z}(s)\mathrm{d}s$ *leading to* $\mathcal{F}(Z(r)) = \frac{\mathrm{d}Z(r)}{\mathrm{d}r}$ *and* $X(s) \equiv B_{H,u}$. *For our type-II examples,* $\lambda = 0$, *but this need not always be the case: a random walk whose increments correlate with* v_t, *but not with* u_t, *is exogenous in the framework of this paper, yet* λ *may still be non-zero when* v_t *is serially correlated*

For both classes of instruments, δ acts as a scaling parameter and the parameter ϑ characterizes the mean reversion of z_t. Note that $z_t = x_t$ would imply a value of $1/2$ for both δ and ϑ, but $G(1)$ would not possess a standard normal distribution for $z_t = x_t$. Type-II instruments, exhibiting indeed $\vartheta = 1/2$, circumvent the possible nonnormality by imposing strict exogeneity: the weak convergence of the normalized partial sums of the product $z_{t-1}u_t$ is the analog of the usual exogeneity condition of instrument variables.

The distribution of the IV test statistic under a suitable sequence of local alternatives is presented in

Theorem 2 *Let* y_t *and* x_t *be generated as in* (1) *and* (2) *with* $\varrho = 1 - c/T$ *where Assumption 1 holds. Assume further that* z_t *satisfies either Assumption 3 or Assumption 4. Under the sequence of local alternatives* $\beta = b/T^{0.5+\vartheta}$, *the limiting distribution of the IV test statistic* t_{iv}^w *from* (11) *is,*

1. *for type-I instruments,*
$$t_{iv}^w \xrightarrow{d} b\,\frac{R_{zx}^c}{\sqrt{V_{zu}}} + G_I(1),$$
and,

2. *for type-II instruments,*
$$t_{iv}^w \xrightarrow{d} b\,\frac{R_{zx}^c}{\sqrt{V_{zu}}} + G_{II}(1)$$
with $V_{zu} = \int_0^1 \left(\dot{Z}(r)\right)^2 \mathrm{d}[B_{H,u}](r)$ *and* $[B_{H,u}](r)$ *the quadratic variation process of* $B_{H,u}(r)$,

where $G_I(1)$ *and* $G_{II}(1) = \left(\int_0^1 \dot{Z}^2(r)\mathrm{d}[B_{H,u}](r)\right)^{-1/2} \int_0^1 \dot{Z}(r)\mathrm{d}B_{H,u}(r)$ *are both standard normal, and* R_{zx}^c *is given by*

$$\frac{1}{T^{1+\vartheta+\delta}} \sum_{t=2}^{T} z_{t-1}x_{t-1} \Rightarrow R_{zx}^c := \left(Z(1)\,J_{c,H,v}(1) - \int_0^1 Z(r)\,\mathrm{d}J_{c,H,v}(r) - \lambda\right)$$

(with $\mathrm{d}J_c(r)$ *shorthand for* $\mathrm{d}B_{H,v}(r) - c\,J_{c,H,v}(r)\,\mathrm{d}r$*) for type-I instruments, and by*

$$\frac{1}{T^{1+\vartheta+\delta}}\sum_{t=2}^{T} z_{t-1}x_{t-1} \;\Rightarrow\; R_{zx}^{c} \;:=\; \int_0^1 J_{c,H,v}(r)\,\dot{Z}(r)\,\mathrm{d}r$$

for type-II instruments.

Proof: *See the Appendix.*

Note that under the null hypothesis ($b = 0$) the IV test statistic has a standard normal limiting distribution irrespective of the value of the parameter c. Furthermore it turns out that the local power of the IV test crucially depends on the parameter ϑ. From the the proof of Theorem 2 it follows for the sample correlation between z_t and x_t:

$$\frac{\sum\limits_{t=1}^{T} z_t x_t}{\sqrt{\sum\limits_{t=1}^{T} z_t^2}\sqrt{\sum\limits_{t=1}^{T} x_t^2}} \;=\; O_p\left(T^{\vartheta-0.5}\right).$$

The larger the parameter ϑ is, the more the instrument is correlated with the regressor, resulting in an improved asymptotic power. For type-I instruments $\vartheta < 0.5$, but we can come arbitrary close to the optimal $1/T$ rate of the distorted OLS t-statistic. For type-II instruments the optimal rate $1/T$ can be achieved, however at the expense of some more restrictive assumption on the limit of $\sum z_{t-1}u_t$ (see Assumption 4).

Remark 3 *For type-II instruments, it is argued in the proof that, if* $\mathrm{Var}(u_t) = \sigma_u^2\ \forall t$ *(i.e. weak stationarity of* u_t *allowing for conditional heteroskedasticity), one has* $V_{zu} = \sigma_u^2 \int_0^1 \left(\dot{Z}(r)\right)^2 \mathrm{d}r$ *and the usual standard errors can be used, s.e* $\left(\widehat{\beta}_{iv}\right) = \widehat{\sigma}_u \left(\sum\limits_{t=2}^{T} z_{t-1}^2\right)^{-1/2}$. *For type-I instruments, just like in the VA case, ordinary standard errors s.e* $\left(\widehat{\beta}_{iv}\right)$ *should be applied only in case of conditional homoskedasticity.*

Remark 4 *In the compact notation of Remark 2,*

$$t_{iv}^{w} \;\overset{d}{\to}\; b\,\frac{R_{zx}^{c}}{\sqrt{V_{zu}}} + \frac{1}{\sqrt{V_{zu}}}\int_0^1 \mathcal{F}(Z(r))\,\mathrm{d}X(r),$$

where the distribution of the second summand on the r.h.s. is standard normal and R_{zx}^{c} *is given by*

$$\frac{1}{T^{1+\vartheta+\delta}}\sum_{t=2}^{T} z_{t-1}x_{t-1} \;\Rightarrow\; R_{zx}^{c} \;:=\; \left(Z(1)\,J_{c,H,v}(1) - \int_0^1 Z(r)\,\mathrm{d}J_{c,H,v}(r) - \lambda\right)$$

with $Z(r) = \int_0^r \dot{Z}(s)\mathrm{d}s$ *again. For type-II instruments, integration by parts leads also to* $\int_0^1 J_{c,H,v}(r)\,\mathrm{d}Z(r)$ *as equivalent representation of* R_{zx}^c.

For concreteness we now discuss the special cases mentioned above in the following corollaries of the theorem.

Corollary 2 (Type-I instruments) *Let* $(u_t, v_t)' \overset{iid}{\sim} (\mathbf{0}, \Sigma)$ *with finite moments of order* $4+\epsilon$ *for some* $\epsilon > 0$ *such that* $B_{H,u} = \sigma_u W_u$, $B_{H,v} = \sigma_v W_v$, $J_{c,H,v} = \sigma_v J_c$, *and* $G_I(r)$ *is a standard Wiener process independent of* J_c *and* W_v.

1. **Short-memory instrument** *If* $z_{t-1} = (1-\bar{\alpha}L)_+^{-1}\Delta x_{t-1}$, *then* $\delta = 0$, $\vartheta = 0$, $V_{zu} = \sigma_u^2 V_z$ *with* $V_z = \frac{\sigma_v^2}{1-\bar{\alpha}^2}$, $\lambda = 0$ *and* $Z(r) = \frac{\sigma_v}{1-\bar{\alpha}} J_c(r)$; *furthermore, with* $\beta = b/T^{1/2}$,

$$t_{iv}^w \overset{d}{\to} b\,\frac{\sigma_v\sqrt{1+\bar{\alpha}}}{\sigma_u\sqrt{1-\bar{\alpha}}}\left(J_c(1)^2 - \int_0^1 J_c(r)\,\mathrm{d}J_c(r)\right) + G_I(1).$$

2. **Mildly integrated instrument** *If* $z_{t-1} = (1-\alpha_T L)_+^{-1}\Delta x_{t-1}$, *with* $\alpha_T = 1 - a/T^\eta$ *then* $\delta = \eta/2$, $\vartheta = \eta/2$, $Z(r) = \frac{\sigma_v}{a} J_c(r)$, $V_{zu} = \sigma_u^2 V_z$ *with* $V_z = \frac{\sigma_v^2}{2a}$, *and* $\lambda = 0$; *furthermore, with* $\beta = b/T^{1/2+\eta/2}$,

$$t_{iv}^w \overset{d}{\to} b\,\frac{\sigma_v\sqrt{2}}{\sigma_u\sqrt{a}}\left(J_c(1)^2 - \int_0^1 J_c(r)\,\mathrm{d}J_c(r)\right) + G_I(1).$$

3. **Fractionally integrated instrument** *If* $z_t = \Delta_+^{1-d^*} x_t$ *for some* $d^* \in (0, 1/2)$, *then* $\delta = 0$, $\vartheta = d^*$, $V_{zu} = \sigma_u^2 V_z$ *with* $V_z = \sigma_v^2 \frac{\Gamma(1-2d^*)}{\Gamma^2(1-d^*)}$, $\lambda = 0$ *and* $Z(r) = \sigma_v J_c^{d^*+1}(r)$; *furthermore, with* $\beta = b/T^{1/2+d^*}$,

$$t_{iv}^w \overset{d}{\to} b\,\frac{\sigma_v}{\sigma_u\sqrt{\frac{\Gamma(1-2d^*)}{\Gamma^2(1-d^*)}}}\left(J_c^{d^*+1}(1)\,J_c(1) - \int_0^1 J_c^{d^*+1}(r)\,\mathrm{d}J_c(r)\right) + G_I(1)$$

   where $J_c^{d^*+1}(r) = W_v^{d^*+1}(r) - c\int_0^r e^{-c(r-s)} W_v^{d^*+1}(s)\,\mathrm{d}s$ *is the Ornstein-Uhlenbeck process driven by the fractional Brownian motion (c.f. Kleptsyna and Le Breton, 2002)* $W_v^{d^*+1}$, *and* $G_I(r)$ *is a standard Wiener process independent of* W_v *(and thus of* J_c *and* $J_c^{d^*+1}$ *as well).*

4. **Long difference instrument** *If* $z_t = x_{t-1} - x_{t-k_T}$, *with* $k_T = \min\{[KT^\nu], t-1\}$ *then* $\delta = \nu/2$, $\vartheta = \nu/2$, $Z(r) = K\sigma_v J_c(r)$, $V_{zu} = \sigma_u^2 V_z$ *with* $V_z = K\sigma_v^2$, *and* $\lambda = 0$; *furthermore, with* $\beta = b/T^{1/2+\nu/2}$,

$$t_{iv}^w \overset{d}{\to} b\,\frac{\sigma_v}{\sigma_u}\sqrt{K}\left(J_c(1)^2 - \int_0^1 J_c(r)\,\mathrm{d}J_c(r)\right) + G_I(1).$$

Proof: *See the Appendix.*

The test based on mildly integrated instruments can be seen as a special case of the IVX approach of Phillips and Magdalinos (2009). The latter approach applies to cointegrating systems with nearly or mildly integrated common trends, of which our baseline predictive regression model is a special case. The fact that the predictive regression framework assumes weakly exogenous regressors facilitates the use of the IVX method as no bias correction is required.

Corollary 3 (Type-II instruments) *Let* $(u_t, v_t)' \overset{iid}{\sim} (\mathbf{0}, \Sigma)$ *with finite moments of order* $4 + \epsilon$ *for some* $\epsilon > 0$ *implying* $B_{H,u} = \sigma_u W_u$*,* $B_{H,v} = \sigma_v W_v$ *and* $J_{c,H,v} = \sigma_v J_c$*.*

1. **Independent random walk** *If* $z_t = \sum_{j=1}^{t} w_{t-1}$ *with* w_t *an* $iid(0, \sigma_w^2)$ *sequence independent of the sequence* $(u_t, v_t)'$*, then* $\delta = 1/2$*,* $\vartheta = 1/2$*,* $\dot{Z}(r) = \sigma_w W_\perp(r)$ *with* $W_\perp$ *a standard Wiener process independent of* W_u *and* W_v*; furthermore, with* $\beta = b/T$*,*

$$t_{iv}^{w} \xrightarrow{d} b\frac{\sigma_v}{\sigma_u}\frac{\int_0^1 W_\perp(r)J_c(r)\mathrm{d}r}{\sqrt{\int_0^1 W_\perp^2(r)\mathrm{d}r}} + \frac{\int_0^1 W_\perp(r)\mathrm{d}W_u(r)}{\sqrt{\int_0^1 W_\perp^2(r)\mathrm{d}r}}$$

with the second summand on the r.h.s. being standard normally distributed.

2. **Linear trend** *If* $z_t = t$*, then* $\delta = 1$*,* $\vartheta = 1/2$*,* $\dot{Z}(r) = r$ *or* $Z(r) = \int_0^r s\mathrm{d}s = r^2/2$*; furthermore, with* $\beta = b/T$*,*

$$t_{iv}^{w} \xrightarrow{d} b\frac{\sigma_v\sqrt{3}}{\sigma_u}\int_0^1 rJ_c(r)\,\mathrm{d}r + \sqrt{3}\int_0^1 r\mathrm{d}W_u(r)$$

with the second summand on the r.h.s. being standard normally distributed.

3. **Cauchy instrument** *If* $z_t = \mathrm{sign}(x_{t-1})$*, then* $\delta = 0$*,* $\vartheta = 1/2$ *and* $\dot{Z}(r) = \mathrm{sign}(J_c(s))$*; furthermore, with* $\beta = b/T$*,*

$$t_{iv}^{w} \xrightarrow{d} b\frac{\sigma_v}{\sigma_u}\int_0^1 |J_c(r)|\,\mathrm{d}r + \int_0^1 \mathrm{sign}(J_c(r))\,\mathrm{d}W_u(r)$$

where the second summand on the r.h.s. is standard normally distributed.

Proof: *See the Appendix.*

Under the null hypothesis $b = 0$, we thus obtain a standard normal limiting null distribution for all IV statistics. Under the local alternative $b \neq 0$, the power depends also on the local-to-unity parameter c, which is an important difference to the VA tests using *the same* variable z_t. Furthermore, the IV test relying on a random walk, a linear time trend or the Cauchy instrument, has power against $^1/_T$-sequences of local alternatives, whereas the asymptotic power of the other tests is equal to the size for such alternatives. This suggests that, in finite samples, the IV test with a linear time trend or the Cauchy instrument is more powerful relative to IV tests employing stationary or moderately integrated instruments for large T.

Remark 5 *Some care is required when the limit R_{zx}^c is random and can take both positive and negative values with positive probabilities. Namely, the power of one-sided tests can be affected in a rather unpredictable way. As an example consider the IV estimator resulting from $z_t = t$ where $R_{zx}^c = \frac{\sqrt{3}}{\sigma_u} \int_0^1 r J_c(r) \mathrm{d}r$ under the conditions of Corollary 3. Since R_{zx}^c is distributed symmetrically about zero, it follows that $\mathrm{E}\left(t_\beta^{iv}\right) = 0$ no matter what the value of the drift parameter b is. Hence, the power of the one-sided test converges to $^1/_2$ as $b \to \infty$ and the test is inconsistent. On the other hand, the power of the two-sided test tends to unity as $b \to \infty$ or $b \to -\infty$ yielding a consistent test against alternatives of the form $\beta = {}^b/_{T^{1-\delta}}$ for all $\delta > 0$.*

Similarly to the VA case, the IV estimator $\widehat{\beta}_{iv}$ need not be normally distributed. Although this is the case for our type-I examples from Corollary 2, it does not hold in general. Moreover, asymptotic normality is unlikely to hold for type-II instruments, for which we have

$$T\left(\widehat{\beta}_{iv} - \beta\right) \Rightarrow \frac{\int_0^1 \dot{Z}(r) \mathrm{d}B_{H,u}(r)}{\int_0^1 \dot{Z}(r) B_{H,v} \mathrm{d}r}.$$

Should e.g. z_{t-1} be a deterministic function of time, both integrals are normally distributed with zero mean; therefore, their ratio is not normally distributed.

2.3 VA versus IV

It is only natural to ask, which procedure is to be preferred in applied work, IV or VA? At a first glance, type-II instruments should be preferred, as they achieve optimal local power. But this comes at the cost of imposing strict exogeneity – which has consequences, should the regressors be stationary. We shall in fact propose a solution to this problem relying on overidentified IV-based inference in the following section. For the remainder of the section, we compare VA-based inference to IV-based procedures employing the same variable z_{t-1}.

Under the null, the IV and VA tests relying on the same variable are asymptotically equivalent, as the difference between the VA and IV test statistics vanishes as $T \to \infty$ in probability. Under the alternatives considered here, however, there can be severe differences, the most obvious one being the additional condition on weak convergence of the partial sums of z_t required for the IV case. But it is not the additional condition of Assumption 3 that is the most relevant for the purposes of this subsection. Namely, the power against local alternatives is governed by the parameter δ for the VA procedure, and by ϑ for the IV procedure. The performance can therefore be quite different even when using the same variable. This is best seen in the case where z_{t-1} is fractionally integrated: the predictability test has power against $T^{-1/2}$ local alternatives in the case of the VA statistic, whereas, if using z_{t-1} as instrument, the local power may come arbitrarily close to the optimal rate T^{-1}. Thus, if $\delta \neq \vartheta$, the decision between IV and VA is straightforward.

But how do IV and VA compare when $\vartheta = \delta$ (for example, when z_t is short memory or mildly integrated)? The answer is not clear cut. By examining the limiting distributions, it is clear that V_z has to be compared to R_{zx}^c. The latter obviously depends on c whereas the former does not, which suggests that the actual value of c influences the ranking. Moreover, for some of the examples considered above, V_z is deterministic while R_{zx}^c is not.

3 Extensions

3.1 Combining instruments

Using one single instrument (the just-identified case) may be problematic. For example, using a linear trend as instrument does not exploit any specific information about the regressor beside the (near) integration of x_t. This implies that the instrument is weak (in the sense that it is uncorrelated with the regressor) if x_t is stationary. Therefore, the IV test has only trivial power in such a case. Moreover, using a time trend as instrument is rather arbitrary as any other function of time (e.g. a sine function or the square of t) may be used instead.

This suggests that the IV test may be improved by employing additional instruments. The two-stage least-squares [2SLS] IV t-statistic in the case of Eicker-White heteroskedasticity-robust standard errors is given by

$$t_{2S}^{w} = \frac{\left(\sum_{t=2}^{T} x_{t-1}\mathbf{z}_{t-1}'\right)\left(\sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)^{-1}\left(\sum_{t=2}^{T} \mathbf{z}_{t-1}y_t\right)}{\sqrt{\left(\sum_{t=2}^{T} x_{t-1}\mathbf{z}_{t-1}'\right)\left(\sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)^{-1}\left(\sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}'\widehat{u}_t^2\right)\left(\sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)^{-1}\left(\sum_{t=2}^{T} \mathbf{z}_{t-1}x_{t-1}\right)}}, \quad (12)$$

where $\mathbf{z_t}$ is an m-dimensional vector of instruments.

Unfortunately, the null distribution of this test statistic is not invariant to the local-to-unity parameter c in general. To illustrate the problems involved consider the following example which is based on the first two terms of the Loève-Karhunen expansion of the Wiener process (cf. Phillips, 1998).

Example 1 *Let* $\mathbf{z}_t = \left(\sin\frac{\pi t}{2T}, \sin\frac{3\pi t}{2T}\right)'$ *and* $(y_t, x_t)'$ *be generated as in* (1) *and* (2)*, where*

$(u_t, v_t)' \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \Sigma)$. *Under the null hypothesis* $\beta = 0$, *the IV statistic* t^w_{2S} *from* (12) *possesses the following asymptotic null distribution:*

$$t^w_{2S} \overset{d}{\to} \frac{\int_0^1 \sin\frac{\pi r}{2}\, J_c(r)\,\mathrm{d}r \cdot \int_0^1 \sin\frac{\pi r}{2}\,\mathrm{d}W_u(r) + \int_0^1 \sin\frac{3\pi r}{2}\, J_c(r)\,\mathrm{d}r \cdot \int_0^1 \sin\frac{3\pi r}{2}\,\mathrm{d}W_u(r)}{\sqrt{2\left[\left(\int_0^1 \sin\frac{\pi r}{2}\, J_c(r)\,\mathrm{d}r\right)^2 + \left(\int_0^1 \sin\frac{3\pi r}{2}\, J_c(r)\,\mathrm{d}r\right)^2\right]}}$$

where $(\sigma_u W_u(r), \sigma_v W_v(r))'$ *is the weak limit of* $T^{-1/2}\sum_{t=1}^{[rT]}(u_t, v_t)'$. *See the Appendix for details.*

The statistic t^w_{2S} *has a standard normal limiting distribution if its numerator is mixed Gaussian. This holds in our model for the special case of uncorrelated shocks* u_t *and* v_t.[6] *In the just identified case,* z_{t-1} *is a scalar and the term* $\sum_{t=2}^{T} x_{t-1}z_{t-1}$ *cancels out (up to the sign). Hence, the resulting statistic* (12) *has a standard normal limiting null distribution. But in the overidentified case these terms do not cancel out anymore, resulting in a nonstandard distribution of* t^{2S}_{β}.

It is nevertheless possible to combine instruments for some interesting special cases. To establish the asymptotic properties of the 2SLS t-statistic in such situations, a multivariate version of Assumptions 3 and 4 is given in

Assumption 5 *Let* $D_T = \mathrm{diag}\left(T^{\delta_1}, \ldots, T^{\delta_m}\right)$ *and* $K_T = \mathrm{diag}\left(T^{1/2+\vartheta_1}, \ldots, T^{1/2+\vartheta_m}\right)$. *The vector of instruments* $\mathbf{z}_t = (z_{1t}, \ldots, z_{mt})'$ *satisfies*

(i) the invariance principle

$$K_T^{-1} D_T^{-1} \sum_{j=2}^{[rT]} \mathbf{z}_{j-1} \;\Rightarrow\; \mathbf{Z}(r)$$

(jointly with weak convergence of $T^{-1/2}\sum_{t=1}^{[rT]}(u_t, v_t)'$*) such that* $\mathbf{Z}(r) = (Z_1(r), \ldots, Z_m(r))'$ *has linearly independent elements,*

$$\frac{1}{T} D_T^{-1}\left(\sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{z}'_{t-1}u_t^2\right) D_T^{-1} \Rightarrow V_{\mathbf{z}u} \quad \text{as well as} \quad \frac{1}{T} D_T^{-1}\left(\sum_{t=2}^{T} \mathbf{z}_{t-1}\mathbf{z}'_{t-1}\right) D_T^{-1} \Rightarrow V_{\mathbf{z}}$$

and

$$\frac{1}{\sqrt{T}} K_T^{-1} D_T^{-1} \sum_{t=2}^{T}\left(\sum_{j=1}^{t-1}\mathbf{z}_j\right) v_t \;\Rightarrow\; \int_0^1 \mathbf{Z}(r)\,\mathrm{d}B_{H,v} + \boldsymbol{\lambda},$$

[6] Note that this finding is invariant to the covariance between the limits of the normalized instruments since the two trigonometric functions are orthogonal in L_2 space over $[0,1]$. Moreover, since the 2SLS t-statistic is identical for all linear combination of the instruments, an orthogonalization of the instruments does not affect the limiting distribution of the test statistic.

(ii) There exists a multivariate process $X(r)$ such that

$$\frac{1}{\sqrt{T}} D_T^{-1} \sum_{t=2}^{T} \mathbf{z}_{t-1} u_t \;\Rightarrow\; \int_0^1 \operatorname{diag}\left(\mathcal{F}_1\left(\mathbf{Z}(r)\right), \ldots, \mathcal{F}_m\left(\mathbf{Z}(r)\right)\right) \mathrm{d}\mathbf{X}(r),$$

with m functions $\mathcal{F}_1, \ldots, \mathcal{F}_m$ as defined in Remark 2, and

(iii) the functional $\mathcal{F} = \int_0^1 \operatorname{diag}\left(\mathcal{F}_1\left(\mathbf{Z}(r)\right), \ldots, \mathcal{F}_m\left(\mathbf{Z}(r)\right)\right) \mathrm{d}\mathbf{X}(r)$ has a zero-mean mixed Gaussian distribution with covariance matrix $V_{\mathbf{z}u}$ such that $V_{\mathbf{z}}\mathcal{F}$ is mixed Gaussian as well.

In the following theorem we assume that there exists an instrument with the property $\vartheta_i > \vartheta_j$ for all $j \neq i$, that is, there is only one instrument that is characterized by the maximum value of ϑ_i. An example is that one instrument is a linear time trend, whereas all other instruments are of type I.

Following Remark 5, we focus on two-sided testing in the remainder of the paper. We posit that the square of the 2SLS t-statistic has a chi-square limiting distribution with one degree of freedom in this case.

Theorem 3 *Let y_t and x_t be generated as in* (1) *and* (2) *with* (3) *where Assumption 1 holds. Consider a vector of m instruments $\mathbf{z}_t = (z_{1t}, \ldots, z_{mt})'$ satisfying Assumption 5 with associated constants $\delta_1, \ldots, \delta_m$ and $\vartheta_1, \ldots, \vartheta_m$. If there exist an instrument z_{it} such that $\vartheta_i > \vartheta_j$ for all $j \neq i$, the 2SLS t-statistic defined in* (12) *has the limiting null distribution*

$$\left(t_{2S}^{w}\right)^2 \overset{d}{\to} \chi^2\left(1\right) .$$

Proof: *See the Appendix.*

The intuition behind the result is that the instrument with the largest value of ϑ_i dominates the asymptotic properties in the sense that under the null hypothesis the 2SLS t-statistic is asymptotically equivalent to the just-identified IV statistic considered in Subsection 2.2. This prevents situations like the one discussed in Example 1.

Notwithstanding the asymptotic equivalence of the over-identified and the just-identified IV statistics, it is quite appealing to employ the over-identified IV statistic. The following example based on the proposed combination of a type-I and type-II instruments illustrates the benefits of this approach.

Example 2 *Let $\mathbf{z}_t = (z_{I,t}, z_{II,t})'$ where $z_{I,t}$ obeys Assumption 3 and $z_{II,t}$ obeys Assumption 4. Let y_t and x_t be generated as in* (1) *and* (2) *with* (3)*, where $(u_t, v_t)' \overset{iid}{\sim} N(\mathbf{0}, \Sigma)$.*

Under the null hypothesis $\beta = 0$ the IV statistic using $\mathbf{z}_t$ as instruments has a χ^2 asymptotic null distribution with one degree of freedom irrespective of whether x_t is nearly integrated with arbitrary c or stationary. (The result is a corollary of Theorem 3 in the near-integrated case, and straightforward to prove in the stationary case.)

Note further that the test has power in a $1/T$ neighborhood of the null when x_t is nearly integrated, and in a $1/\sqrt{T}$ neighborhood when x_t is stationary (under the additional condition that $z_{I,t}$ correlates with x_t, which is fulfilled, e.g., by the examples in Corollary 2). When x_t is nearly integrated, the type-I instrument is asymptotically irrelevant, whereas, when x_t is stationary, it is the type-II instrument that is irrelevant. In this sense, the 2SLS procedure asymptotically "picks" the instrument suitable for the given degree of persistence of the regressor.

As a counterexample we note that a combination of the Cauchy instrument and a linear trend has a nonstandard limiting distribution: the crucial value of ϑ is identical for both instruments (and both are type-II instruments). Since the two sine functions from Example 1 have the same $\vartheta = 1/2$, Theorem 3 does not apply either.

It is however possible to construct Anderson-Rubin [AR] type statistics for more general combinations of instruments than specified in Theorem 3, including e.g. two or more sine functions. In the heteroskedasticity-robust form, the test statistic is given by

$$AR = \left(\sum_{t=1}^{T} y_t \mathbf{z}_{t-1}'\right) \left(\sum_{t=1}^{T} \mathbf{z}_{t-1}\mathbf{z}_{t-1}' \widehat{u}_t^2\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{z}_{t-1} y_t\right) \tag{13}$$

with $\widehat{u}_t$ OLS residuals as before, and is equivalent to the LM statistic of the hypothesis $\gamma = 0$ in the regression $y_t = \boldsymbol{\gamma}'\mathbf{z}_{t-1} + e_t$. It is interesting to note that, in the just-identified case where the number of instruments is equal to the number of regressors, the test statistic is identical to the square of the ordinary IV statistic given in (11). The limiting distribution for a general class of instruments is presented in

Theorem 4 *Let y_t and x_t be generated as in* (1) *and* (2) *with* (3) *where Assumption 1 holds and let $\mathbf{z}_t$ be an $m \times 1$ vector of instruments obeying Assumption 5 with $m \geq 1$. Then, under the sequence of local alternatives $\beta = b/T^{\max(0.5+\vartheta_i)}$ and Assumption 5, the AR statistic from* (13) *has the following limiting distribution as $T \to \infty$:*

$$AR \xrightarrow{d} \chi^2 \left(m, b^2 \sum\sum_{\vartheta_i, \vartheta_j = \max \vartheta} \left[V_{\mathbf{z}u}^{-1}\right]_{i,j} R_{zx,i}^c R_{zx,j}^c \right)$$

where $\chi^2(m, \kappa)$ represents a noncentral χ^2 distribution with m degrees of freedom and noncentrality parameter κ.

Proof: *See the Appendix.*

Note that imposing Assumption 5 rules out the joint use of the Cauchy instrument and a linear trend, since the corresponding vector $\mathbf{z}_t$ would violate Assumption 5 (iii). On the other hand, several powers of the trend, several independent random walks, or trigonometric functions with different frequencies can be used.

An important drawback of the AR approach is that the critical values increase with the number of instruments m. Thus, in order to improve the power of the test in finite samples, the additional instruments have to be sufficiently informative to compensate the higher critical value. Note further that only the most powerful instrument(s) (that is the instruments with the maximal value of ϑ) determine(s) the asymptotic power of the AR statistic.

3.2 Deterministic components

Let us examine the predictive regression (1) augmented by an additive deterministic component of the form

$$\mu_t = \sum_{\ell=1}^{L} \tau_\ell f_\ell(t) = \boldsymbol{\tau}'\mathbf{f}_t \ ,$$

where $\mathbf{f}_t$ obeys the conditions of Assumption 5 as a type-II variable. Accordingly, this allows for more general deterministic components than a non-zero mean, e.g. polynomial trends or dummy variables.

We may also include deterministic terms in the data generating process of x_t. Let $E(x_t) = \boldsymbol{\tau}_x'\mathbf{f}_t^x$. If, for example, a constant is included in (2), then $\mathbf{f}_t^x = (1, t)'$. In this case the vector $\mathbf{f}_t$ is constructed such that $\mathbf{f}_t^x$ lies in the space of $\mathbf{f}_t$. For instance, if both (1) and (2) include a constant, then $\mathbf{f}_t = (1, t)'$ should be chosen, although we know that $\tau_2 = 0$. The (irrelevant) time trend is included in (2) to remove the time trend from the regressor whenever it is (nearly) integrated.

By expanding the test equation for the VA test accordingly we obtain

$$y_t = \boldsymbol{\tau}'\mathbf{f}_t + \beta z_{t-1} + \psi\zeta_{t-1} + u_t. \tag{14}$$

Since, under Assumptions 2 and 5, the variable z_t is asymptotically orthogonal to both ζ_t and $\mathbf{f}_t$, it is not difficult to show that augmentation of the testing equation with deterministic terms does not change the asymptotic results presented in Theorem 1; cf. Sims et al. (1990) again. For the 2SLS IV approach, it is sufficient to ensure that all elements of the vector $(\mathbf{f}_t, z_t)$ obey Assumption 5 *jointly*. The corresponding result is presented in

Theorem 5 *Assume that y_t and x_t are generated as in* (1) *and* (2) *with* (3) *where Assumption 1 holds. Furthermore, let the vector $(\mathbf{f}_t', z_{t-1})'$ obey Assumption 5. The limiting distribution of the IV statistic for the null hypothesis $\beta = 0$ in model $y_t = \boldsymbol{\tau}'\mathbf{f}_t + \beta x_{t-1} + u_t$*

is given by

$$(t_{iv}^w)^2 \xrightarrow{d} \chi^2(1).$$

Proof: *straightforward and omitted.*

Note, however, that Assumption 5 (i) rules out the use of the linear trend as instrument when a time trend is already present in the data and has to be accounted for. Assumption 5 (iii) furthermore prohibits the use of the Cauchy instrument[7] but we may, for instance, use $z_t = \sin\frac{\pi t}{2T}$ instead of a linear time trend.

For the Anderson-Rubin statistic several type-II instruments may be used. Let $\widetilde{y}_t$ denote the residuals from an OLS regression of y_t on $\mathbf{f}_t$ and $\widetilde{u}_t$ is the residual from y_t on $\mathbf{f}_t$ and x_{t-1}. The Anderson-Rubin statistic is computed as in (13) where y_t and $\widehat{u}_t^2$ is replaced by $\widetilde{y}_t$ and $\widetilde{u}_t^2$, and its limiting null distribution remains χ^2 with m degrees of freedom.

3.3 Multiple predictors

Should there be several regressors $\mathbf{x}_t \in \mathbb{R}^K$,

$$y_t = \boldsymbol{\tau}'\mathbf{f}_t + \boldsymbol{\beta}'\mathbf{x}_{t-1} + u_t, \tag{15}$$

it is straightforward to use the proposed approaches. Moreover, the extension to several predictor variables for different degrees of persistence for each of the predictors is straightforward. Let $^1/_T\Gamma$ be a (not necessarily diagonal) $K \times K$ matrix with all eigenvalues within the unit circle for all T. The univariate autoregressive process (2) is replaced by the multivariate analog

$$\mathbf{x}_t = \left(I_K - \frac{1}{T}\Gamma\right)\mathbf{x}_{t-1} + \mathbf{v}_t\,, \tag{16}$$

and a multivariate version of Assumption 1 is available for the vector $(u_t, \mathbf{v}_t')'$. In the case of the VA method we adapt a generalized version of (14) yielding

$$y_t = \boldsymbol{\tau}'\mathbf{f}_t + \boldsymbol{\beta}'\mathbf{z}_{t-1} + \boldsymbol{\psi}'\boldsymbol{\zeta}_{t-1} + u_t, \tag{17}$$

where each element of the vector $\mathbf{z}_{t-1}$ is constructed as in the univariate case, and $\boldsymbol{\zeta}_t = \mathbf{x}_t - \mathbf{z}_t$. If Assumption 5 holds jointly for $\mathbf{z}_t$ and $\mathbf{f}_t$, the previous results on the properties of the modified VA procedure are easily established in the multivariate case as well. The test statistic is the usual Wald statistic for the null hypothesis

[7] If the data only exhibits a nonzero mean, recursive demeaning of the regressor and forward demeaning of the dependent variable lead to normality of the IV test statistic based on the sign instrument.

$\boldsymbol{\beta} = \mathbf{0}$ where the Eicker-White heteroskedasticity-consistent covariance matrix estimator $\widehat{V}_w\left(\widehat{\boldsymbol{\beta}}_{va}\right)$ is employed to accommodate heteroskedastic errors. This test statistic is denoted by $\mathcal{T}^w_{va,K} = \widehat{\boldsymbol{\beta}}'_{va}\, \widehat{V}_w^{-1}\left(\widehat{\boldsymbol{\beta}}_{va}\right) \widehat{\boldsymbol{\beta}}_{va}$, and its asymptotic null properties are established in the following

Theorem 6 *Assume that y_t and $\mathbf{x}_t$ are generated as in* (1) *and* (16)*, where the vector $(u_t, \mathbf{v}'_t)'$ satisfies a multivariate version of Assumption 1. The vector $(\mathbf{f}'_t, \mathbf{z}'_{t-1})'$ obeys Assumption 5 and $\mathbf{z}_t$ contains only type-I instruments. Under the null hypothesis $\boldsymbol{\beta} = \mathbf{0}$, the limiting distribution of the Wald-type statistic is given by*

$$\mathcal{T}^w_{va,K} \xrightarrow{d} \chi^2(K).$$

Proof: *see the Appendix.*

A analogous result applies to the IV approach using type-I instruments. Some care is needed, though, for type-II instruments. While the Anderson-Rubin statistic can be used in a straightforward manner, the 2SLS approach is not without potential pitfalls. First, an obvious requirement is that $(\mathbf{f}'_t, \mathbf{z}'_{t-1})'$ is not multicollinear. For example, if $\mathbf{f}_t$ includes a linear time trend, the trend is obviously ruled out as a type-II instrument. Second, the problems with combining instruments (see Example 1) are also relevant in the multivariate context. This suggests that one may only use one type-II instrument for each regressor. Theorem 7 below shows that using K type-II instruments for K predictors leads to a χ^2_K asymptotic null distribution. As before, the Wald statistic is defined as

$$\mathcal{T}^w_{iv,K} = \widehat{\boldsymbol{\beta}}'_{iv}\, \widehat{V}^{-1}\left(\widehat{\boldsymbol{\beta}}_{iv}\right) \widehat{\boldsymbol{\beta}}_{iv}\,, \tag{18}$$

where $\widehat{\boldsymbol{\beta}}_{iv}$ denotes the IV estimator of $\boldsymbol{\beta}$ in (15) and $\widehat{V}_w\left(\widehat{\boldsymbol{\beta}}_{iv}\right)$ is the corresponding diagonal block of the heteroskedasticity-consistent IV covariance matrix estimator (the precise expression is given in the proof of the following theorem).

Theorem 7 *Under the assumptions of Theorem 6, the limiting distribution of the IV statistic from* (18) *for model* (15) *is given by*

$$\mathcal{T}^w_{iv,K} \xrightarrow{d} \chi^2(K).$$

Proof: *see the Appendix.*

Under the conditions of Theorem 3, it follows that the type-II instruments (i.e. the instruments with the largest value of ϑ_i matter asymptotically. Thus, adding further type-I instruments to obtain an overidentified 2SLS IV estimator does not raise difficulties. Just like in the univariate case, adding type-I instruments may improve finite-sample power in particular if x_t is stationary (to escape the weak instruments problem).

4 Monte Carlo experiments

We now present some simulation evidence comparing the size and power properties of alternative tests. Data are generated according to model (1) and (2) with serially independent bivariate normal u_t and v_t, where an additional constant is included in (1). Accordingly, all tests are applied to the demeaned series. Instead of the parameter c we present the resulting autoregressive parameter $\varrho = 1 - {}^{c}\!/_{T}$, where the sample size is $T = 250$. The parameter of interest is computed as $\beta = {}^{b}\!/_{T}$ and, therefore, b measures the deviation from the null hypothesis relative to the sample size. Table 1 reports rejection rates for two-sided tests based on 10 000 replications of the model with nominal size 10%. The correlation between u_t and v_t is $\omega = 0.9$ and both errors have unit variance.

The following test statistics are considered. The statistic π^* represents the test proposed by Jansson and Moreira (2006) and "CY-Q" refers to the Bonferroni Q statistic of Campbell and Yogo (2006).[8] The original variable addition test proposed by Dolado and Lütkepohl (1996) (i.e. the test is performed on the difference $z_{t-1} = \Delta x_{t-1}$) is denoted by "$\text{VA}_{\bar{\alpha}=0}$" and the modified VA test suggested in Corollary 1.2 is labeled as "VA_{mild}", where the autoregressive parameter used to construct $\Delta\xi_{t-1}$ is computed as $\alpha_T = 1 - {}^{12.5}\!/_{T^{0.8}}$. The test statistic employing long differences of the regressor as instrument with truncation parameter $k_T = [0.2 \cdot T^{0.85}]$ is denoted by "IV_{k_T}". The IV test using $\text{sign}(x_{t-1})$ as instrument (with forward demeaning for the dependent variable and recursive demeaning for the regressor) is indicated as "IV_{sign}", whereas the IV tests based on fractionally integrated instrument $\Delta_+^{1/2} x_{t-1}$ and a sine function[9] with $\sin({}^{\pi t}\!/_{T})$ are labeled "$\text{IV}_{d^*={1/2}}$" and "IV_{sin}", respectively. Finally, "IV_{comb}" indicates an over-identified IV estimator combining the sine function with the fractional instrument (cf. Theorem 3).

The simulation results for a significance level of 10% are reported in Table 1. They show negligible size distortions for all VA and IV tests that are within the simulation error of the rejection frequencies. Only the VA test with moderately integrated regressor tends to reject slightly too often. With respect to the power of the tests, substantial differences emerge. As predicted by Theorem 1, the original VA test has very poor power relative to the other tests whenever ϱ is close to unity. The modified statistic VA_{mild} proposed in Corollary 1 (2.) dramatically improves the power of the VA test. The relative performance of the IV tests also depends on the persistence of the regressor. While the sine function

[8] The authors are grateful to Michael Jansson for sharing the corresponding computer codes and to Motohiro Yogo for making them available (`https://sites.google.com/site/motohiroyogo/home/publications/Predict_Programs.zip`).

[9] The IV test with a trend as instrument performs slightly worse and is omitted.

Table 1: Comparison of alternative robust test procedures when the regressor is nearly integrated (nominal size: 10%)

$\varrho = 1.00$									
b	CY-Q	π^*	$\text{VA}_{\bar{\alpha}=0}$	VA_{mild}	IV_{k_T}	IV_{sign}	$\text{IV}_{d^*=1/2}$	IV_{sin}	IV_{comb}
0	5.6	10.1	11.1	13.3	12.5	10.3	11.1	9.9	11.2
5	44.5	50.2	10.1	12.8	15.8	19.1	13.6	37.7	34.5
10	77.1	55.7	13.5	22.8	33.6	35.7	33.4	61.4	65.7
15	91.2	58.4	18.7	39.5	51.8	48.5	52.4	73.4	83.7
20	96.7	54.6	27.3	57.6	66.1	58.6	66.9	79.5	91.3
$\varrho = 0.98$									
b	CY-Q	π^*	$\text{VA}_{\bar{\alpha}=0}$	VA_{mild}	IV_{k_T}	IV_{sign}	$\text{IV}_{d^*=1/2}$	IV_{sin}	IV_{comb}
0	5.8	8.1	11.0	11.4	10.7	9.3	10.2	10.6	10.8
5	20.5	17.7	11.0	14.4	17.8	17.8	17.0	29.4	30.9
10	52.7	23.8	14.3	26.4	37.6	30.5	34.0	45.1	55.6
15	75.6	25.9	20.6	44.6	59.2	44.7	54.8	55.6	75.1
20	89.6	33.9	29.5	62.8	76.1	56.5	69.7	63.2	86.6
$\varrho = 0.96$									
b	CY-Q	π^*	$\text{VA}_{\bar{\alpha}=0}$	VA_{mild}	IV_{k_T}	IV_{sign}	$\text{IV}_{d^*=1/2}$	IV_{sin}	IV_{comb}
0	6.0	6.2	10.5	10.4	10.5	10.0	9.4	9.9	10.4
5	15.2	11.9	11.1	14.5	16.4	16.8	15.5	21.9	26.3
10	41.0	16.5	14.6	26.4	33.4	30.5	30.1	34.4	47.3
15	63.4	18.8	21.4	43.1	53.5	45.0	48.4	44.4	66.5
20	79.9	21.0	30.5	61.6	70.0	58.9	64.7	51.9	80.4
$\varrho = 0.92$									
b	CY-Q	π^*	$\text{VA}_{\bar{\alpha}=0}$	VA_{mild}	IV_{k_T}	IV_{sign}	$\text{IV}_{d^*=1/2}$	IV_{sin}	IV_{comb}
0	6.5	6.1	10.7	10.4	10.6	10.0	10.1	9.8	11.4
5	12.4	7.3	11.3	13.8	14.7	15.0	13.9	16.9	20.0
10	28.9	10.7	15.4	24.0	26.3	26.1	23.7	24.4	35.7
15	48.0	13.6	22.1	40.8	43.6	40.4	39.6	31.5	54.3
20	65.6	15.5	30.5	57.4	61.4	55.1	55.3	37.3	68.9

Note: Rejection frequencies are computed from 10 000 replications of model (1) – (2) allowing for a nonzero constant. The mutual correlation of the errors is $\omega = 0.9$. The nominal size is 0.1, and the sample size is $T = 250$. The null hypothesis is $b = 0$.

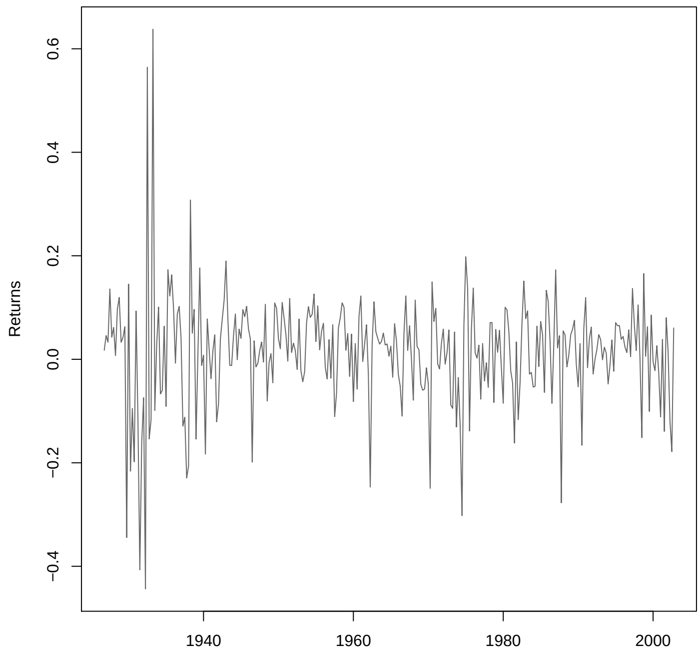
performs well as an instrument if x_t is close to $I(1)$, the performance of the IV_{sin} test deteriorates severely if the regressor approaches stationarity (still, it is more powerful than π^*). The reason is that the deterministic sine function is uncorrelated with the stationary AR process x_t and, therefore, the weak instrument problem arises. The IV estimator that combines the sine function and the fractional instrument performs best among all VA and IV statistics. The power of this test is comparable to the CY-Q test which has been characterized to be efficient in this context (cf. Campbell and Yogo, 2006).

In a set of additional Monte Carlo experiments we have also investigated the relative performance of various Anderson-Rubin statistics proposed in (13). It turned out that the AR test generally suffers from significant size distortions in small samples, in particular when more than two instruments are employed. Moreover the power tends to be smaller compared to the 2SLS statistic IV_{comb}. We therefore do not recommend the AR test for empirical applications and in order to save space we do not present the results here.[10]

5 Predictability of quarterly U.S. stock index returns

To illustrate the methods discussed in the previous sections, we re-analyze the predictability of U.S. equity data. The predicted series are quarterly NYSE/AMEX value-weighted index data (1926 Q4 to 2002 Q4) from the Center for Research in Security Prices analyzed in Campbell and Yogo (2006); as predictors we employ the (log) dividend yield, the (log) earnings-price ratio, a risk-free interest rate and a yield spread.[11] The series are plotted in Figures 1 and 2.

Figure 1: Quarterly NYSE/AMEX value-weighted index returns, 1926Q4–2002Q4



For the potential predictors, we also report the estimated autoregressive coefficient $\widehat{\varrho}$ and the sample correlation between the prediction error $\widehat{u}_t$ and the residual of the autoregression $\widehat{v}_t$ in Table 2. For two predictors, the log dividend yield and the log earnings-price ratio, we find high persistence and a very large negative correlation, therefore, we expect the OLS t-statistic to be unreliable.

We perform OLS predictive regressions with each possible predictor alone, and a joint regression with all four predictors. All regressions include an intercept. The results are

[10] The respective results are available from the authors on request.

[11] Obtained from `https://sites.google.com/site/motohiroyogo/home/publications/Predict_Data.xls`; we thank Motohiro Yogo for making the replication data available.

Figure 2: Potential predictor series, 1926Q4–2002Q4

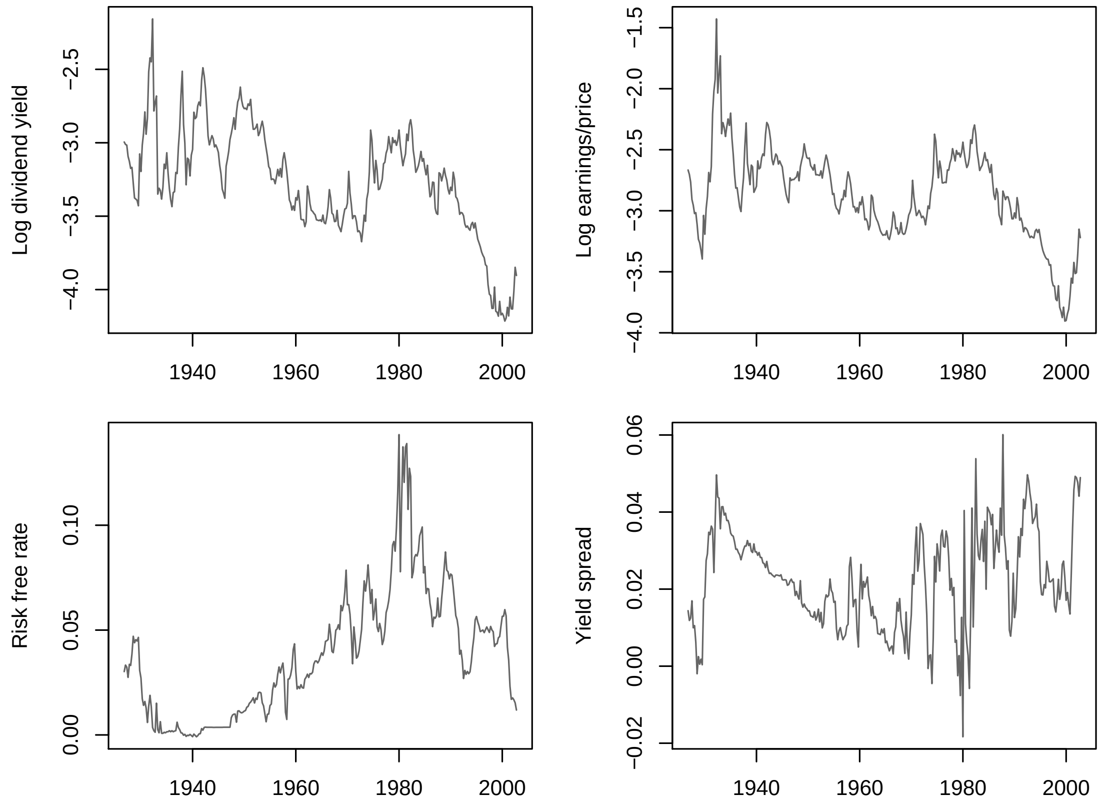


Table 2: Predictors: Some stylized facts

	Dividend yield	Earnings/price	Risk-free rate	Yield spread
$\widehat{\varrho}$	0.9634	0.9578	0.9654	0.7999
$\widehat{\sigma}_{uv}/\widehat{\sigma}_u\widehat{\sigma}_v$	−0.9422	−0.9861	−0.0501	−0.1194

presented in Table 3. It turns out that, in the problematic cases where size distortions are expected, the OLS t statistics tend to reject. In the joint regression, only the t-statistics for the earnings-price ratio is significant.

Applying the robust methods reveals interesting differences. On the one hand, employing long-difference instruments (the VA or the IV approach) rejects the null hypothesis $\beta = 0$ for the earnings-price based on the usual t-statistic (assuming homoskedasticity). On the other hand, neither test is able to reject the null hypothesis when the heteroskedasticity-robust variant of the t-statistic is applied to the four indicators. Thus, for the period 1926-2002, there is no robust evidence in favor of the predictability of quarterly U.S. index returns by using financial variables like the earnings-price ratio or the dividend yield.

Lettau and Van Nieuwerburgh (2008) argue however that it is the *stationary* component of the dividend yield or the earnings-price ratio that matters. They identify two

Table 3: OLS estimation results

Intercept	Dividend yield	Earnings/price	Risk-free rate	Yield spread
0.1247* (0.0544)	0.0342* (0.0166)	–	–	–
0.1489* (0.0470)	–	0.0473* (0.0163)	–	–
0.0221* (0.0098)	–	–	−0.2332 (0.2020)	-
0.0009 (0.0126)	–	–	–	0.5556 (0.4951)
0.1138* (0.0551)	−0.0505 (0.0410)	0.0894* (0.0402)	−0.2097 (0.2193)	−0.0345 (0.5631)

Note: * indicates significance at the 5%. Ordinary standard errors (assuming homoskedastic errors) are presented in parentheses.

breaks in the mean, in 1955 and in 1995, for the yearly dividend yield, and find that the demeaned series (that is by removing the means over 1926-1954, 1955-1994 and 1995-2002) does indeed help to predict yearly returns. We therefore included the break-adjusted predictors in our empirical investigation, focusing on the log earnings-price ratio, as this variable appeared the more promising candidate in our previous analysis. We adopt the break dates identified by Lettau and Van Nieuwerburgh (2008). Since our data is quarterly, we locate the breaks at the first quarter of 1955 and 1995. The demeaned predictor is depicted in Figure 3. As the demeaned series are still persistent with an autoregressive root of 0.905, the OLS t-statistic may still suffer from size distortions. As in Lettau and Van Nieuwerburgh (2008) we found that the null hypothesis is rejected but if the test statistic is adjusted for heteroskedasticity the t-statistics become insignificant, with the exception IV statistic based on long-differences that marginally rejects at a significance level of 0.05. This result corroborates earlier findings that it is important to account for heteroskedastic errors when testing financial time series.

6 Concluding remarks

The paper discussed alternative inferential procedures for testing the predictive power of variables with unknown persistence. It is known that the asymptotic null distributions of standard test statistics such as the usual OLS-based t-test may suffer from severe size distortions if the regressor is (nearly) integrated. The proposed procedures are robust in the sense that size control does not come at essential power losses. Furthermore the proposed tests are robust against a general forms of heteroskedasticity.

Concretely, we discuss a class of modified variable addition tests and consider their local power under sequences of local alternatives. We found that the original version of the VA tests may suffer from dramatic loss of power relative to other existing tests. By choosing a more appropriate variable decomposition, the loss of power may however be

Table 4: Results based on robust test procedures

	Dividend yield	Earnings/price (E/P)	Risk-free rate	Yield spread	E/P (break-adj.)
		VA test: mildly integrated			
$\widehat{\beta}$	0.0052	0.0383	–0.3522	0.3416	0.1143
t_{va}	0.1516	1.2476	–0.7227	0.5328	3.4840*
t_{va}^{w}	0.0828	0.4631	–0.6888	0.4148	2.9368*
		VA test: long differences			
$\widehat{\beta}$	0.0372	0.0531	–0.2940	0.3796	0.0660
t_{va}	1.9683*	3.0239*	–1.1011	0.5603	2.0754*
t_{va}^{w}	1.3778	1.5965	–0.9245	0.6002	1.3348
		IV test: mildly integrated			
$\widehat{\beta}$	0.0115	0.0393	–0.3625	0.3158	0.0598
t_{iv}	0.4020	1.3824	–0.6963	0.4704	1.8571
t_{iv}^{w}	0.2269	0.5576	–0.6840	0.3757	0.7909
		IV test: long differences			
$\widehat{\beta}$	0.0400	0.0620	–0.3840	0.7585	0.1090
t_{iv}	1.6667	2.6840*	–0.8084	1.0459	3.2812*
t_{iv}^{w}	1.2991	1.5938	–0.7998	0.8409	1.9894*
		IV test: combined (sine, fractional)			
$\widehat{\beta}$	0.0169	0.0308	–0.1467	0.3989	0.0701
t_{iv}	0.7788	1.3756	–0.5840	0.5380	2.0378*
t_{iv}^{w}	0.4495	0.6016	–0.4353	0.4693	0.9616

Note: This table present the result of 5 predictive regressions for the indicated variables. t_{va} and t_{iv} indicate the ordinary t-statistics based on the assumption of homoskedastic errors. t_{va}^{w} and t_{iv}^{w} denote the respective statistics based on the Eicker-White heteroskedasticity robust standard errors. * indicates significance at the 5%.

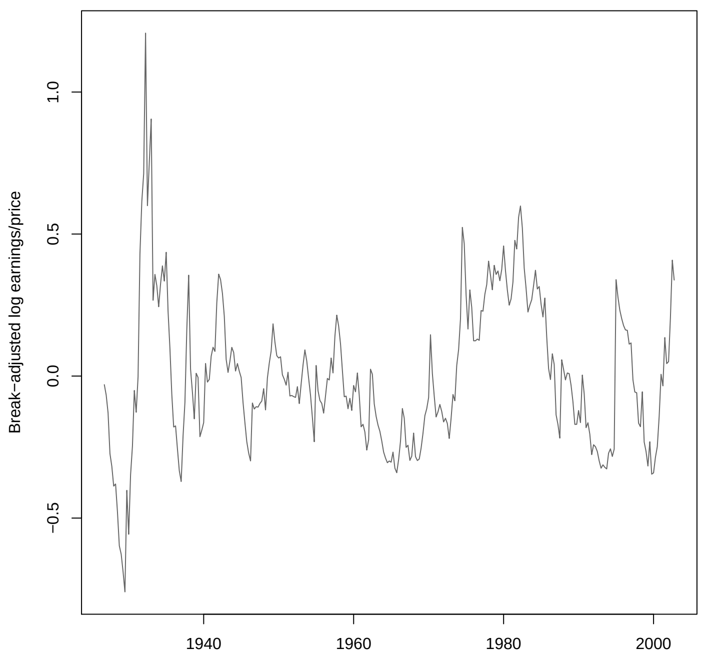


Figure 3: Break-adjusted predictor (log earnings-price ratio), 1926Q4–2002Q4

reduced to a minimum. Instrumental variable inference offers an alternative simple and robust solution. In fact, when choosing instruments properly, the IV tests have power in the same neighborhood of the null as the infeasible OLS-based test: instruments as simple as a linear trend or the sign of the regressor fit into this category. Another useful class of instruments consists of mean-reverting, yet highly persistent variables. Such instrumental variables can e.g. be obtained by filtering the differenced regressor and are useful when the regressor approaches the stationarity region. All instruments lead to standard inference irrespective of the degree of persistence of the regressor. Moreover, combining instruments by 2SLS or by using the Anderson-Rubin test statistic may further improve on the power of the tests.

The proposed test statistics perform well in finite samples. Our Monte Carlo experiments find for the baseline model that the VA procedure and the combination of a stationary instrument and a linear trend appear most promising for applied work: both control size while having power comparable to that of the Bonferroni Q test proposed by Cambell and Yogo (2006) which are based however on a more restrictive model framework.

The application to U.S. stock market returns suggest that accounting for heteroskedasticity of the error process is crucial for valid inference. Applying our robust framework questions the finding of earlier work that earnings-price ratios or dividend yields are useful predictors for stock price returns.

Appendix

Proof of Theorem 1

The heteroskedasticity robust t statistic is written as

$$\begin{aligned} t_{va}^{w} &= \frac{1}{\sqrt{N_T}} \left(\sum_{t=2}^{T} z_{t-1} u_t - \sum_{t=2}^{T} z_{t-1} \zeta_{t-1} \left(\sum_{t=2}^{T} \zeta_{t-1}^2 \right)^{-1} \sum_{t=2}^{T} \zeta_{t-1} u_t \right) \\ &\quad + \frac{\beta}{\sqrt{N_T}} \left(\sum_{t=2}^{T} z_{t-1}^2 - \sum_{t=2}^{T} z_{t-1} \zeta_{t-1} \left(\sum_{t=2}^{T} \zeta_{t-1}^2 \right)^{-1} \sum_{t=2}^{T} \zeta_{t-1} z_{t-1} \right) \end{aligned}$$

where

$$\begin{aligned} N_T &= \sum_{t=2}^{T} z_{t-1}^2 \widehat{u}_t^2 - 2 \sum_{t=2}^{T} z_{t-1} \zeta_{t-1} \left(\sum_{t=2}^{T} \zeta_{t-1}^2 \right)^{-1} \sum_{t=2}^{T} \zeta_{t-1} z_{t-1} \widehat{u}_t^2 \\ &\quad + \sum_{t=2}^{T} z_{t-1} \zeta_{t-1} \left(\sum_{t=2}^{T} \zeta_{t-1}^2 \right)^{-1} \sum_{t=2}^{T} \zeta_{t-1}^2 \widehat{u}_t^2 \left(\sum_{t=2}^{T} \zeta_{t-1}^2 \right)^{-1} \sum_{t=2}^{T} \zeta_{t-1} z_{t-1}. \end{aligned}$$

Let us now analyze the asymptotic behavior of the various sample moments in the above expression. It follows from Assumption 2 that

$$\sum_{t=2}^{T} z_{t-1} \zeta_{t-1} = \sum_{t=2}^{T} z_{t-1} x_{t-1} - \sum_{t=2}^{T} z_{t-1}^2 = o_p\left(T^{1.5+\delta}\right),$$

since $\delta < 1/2$. Recall that $T^{-1/2} x_{[rT]} \Rightarrow J_{c,H,v}(r)$, so the continuous mapping theorem (CMT) implies that

$$\frac{1}{T^2} \sum_{t=2}^{T} \zeta_{t-1}^2 = \frac{1}{T^2} \sum_{t=2}^{T} x_{t-1}^2 - \frac{2}{T^2} \sum_{t=2}^{T} x_{t-1} z_{t-1} + \frac{1}{T^2} \sum_{t=2}^{T} z_{t-1}^2 \Rightarrow \int_0^1 J_{c,H,v}^2(r) \mathrm{d}r,$$

and, therefore,

$$\left(\sum_{t=2}^{T} \zeta_{t-1}^2 \right)^{-1} = O_p\left(\frac{1}{T^2}\right).$$

Since $x_{t-1} u_t$ has the martingale difference property it follows that $\operatorname{Var}\left(\sum_{t=2}^{T} x_{t-1} u_t\right) = \sum_{t=2}^{T} \mathrm{E}\left(x_{t-1}^2 u_t^2\right) \leq \sqrt{\mathrm{E}\left(x_{t-1}^4\right) \mathrm{E}\left(u_t^4\right)}$; the uniform boundedness of $\mathrm{E}\left(u_t^4\right)$ is guaranteed by Assumption 1, while the Beveridge-Nelson decomposition (c.f. Phillips and Solo (1992)) indicates that the leading term of $\mathrm{E}\left(x_{t-1}^4\right)$ is

$$\sum_{j=1}^{t-1} \sum_{k=1}^{t-1} \sum_{l=1}^{t-1} \sum_{m=1}^{t-1} \mathrm{E}\left(\bar{v}_j \bar{v}_k \bar{v}_l \bar{v}_m\right).$$

The above expectations are nonzero if the largest two indices are equal, so

$$\sum_{j=1}^{t-1} \sum_{k=1}^{t-1} \sum_{l=1}^{t-1} \sum_{m=1}^{t-1} \mathrm{E}\left(\bar{v}_j \bar{v}_k \bar{v}_l \bar{v}_m\right) = \mathrm{E}\left(\bar{v}_1^4\right) + \sum_{j=2}^{t-1} \left(\sum_{k=1}^{j-1} \sum_{l=1}^{j-1} \mathrm{E}\left(\bar{v}_k \bar{v}_l \bar{v}_j^2\right) \right).$$

Using the summability condition on $\mathrm{E}\left(\widetilde{v}_k\widetilde{v}_l\widetilde{v}_j^2\right)$ from Assumption 1 and the boundedness of H, we conclude that $\mathrm{E}\left(x_{t-1}^4\right) = O\left(T^2\right)$ leading to $\sum_{t=2}^{T} x_{t-1}u_t = O_p(T)$ and thus

$$\sum_{t=2}^{T} \zeta_{t-1}u_t = \sum_{t=2}^{T} x_{t-1}u_t - \sum_{t=2}^{T} z_{t-1}u_t = O_p\left(T\right)$$

by using $\sum_{t=2}^{T} z_{t-1}u_t = O_p\left(T^{1/2+\delta}\right)$.

Moving on to the analysis of terms containing residuals, note that weak convergence of $T^{-1/2}\,x_{[rT]}$ to Ornstein-Uhlenbeck process implies that $\sup_{2\leq t\leq T}|x_{t-1}| = O_p\left(T^{1/2}\right)$ and $\widehat{\beta}_{ls} - \beta = O_p\left(T^{-1}\right)$. Furthermore, $\sup_{2\leq t\leq T}|u_t| = o_p\left(T^{1/2}\right)$ and $\sup_{2\leq t\leq T} u_t^2 = o_p\left(T^{1/2}\right)$ given the uniform boundedness of $\mathrm{E}\left(|u_t|^{4+\epsilon}\right)$ for some $\epsilon > 0$. This leads to

$$\begin{aligned}\sum_{t=2}^{T} z_{t-1}^2\widehat{u}_t^2 &= \sum_{t=2}^{T} z_{t-1}^2u_t^2 - 2\left(\widehat{\beta}_{ls} - \beta\right)\sum_{t=2}^{T} z_{t-1}^2x_{t-1}u_t + \left(\widehat{\beta}_{ls} - \beta\right)^2\sum_{t=2}^{T} z_{t-1}^2x_{t-1}^2\\ &= \sum_{t=2}^{T} z_{t-1}^2u_t^2 + o_p\left(T^{1+2\delta}\right)\end{aligned}$$

since

$$\begin{aligned}\left|\left(\widehat{\beta}_{ls} - \beta\right)\sum_{t=2}^{T} z_{t-1}^2x_{t-1}u_t\right| &\leq \left|\widehat{\beta}_{ls} - \beta\right|\left(\sup_{2\leq t\leq T}|x_{t-1}u_t|\right)\sum_{t=2}^{T} z_{t-1}^2\\ &\leq \left|\widehat{\beta}_{ls} - \beta\right|\sum_{t=2}^{T} z_{t-1}^2\sup_{2\leq t\leq T}|x_{t-1}|\sup_{2\leq t\leq T}|u_t|\\ &= o_p\left(T^{1+2\delta}\right)\end{aligned}$$

and

$$0 \leq \left(\widehat{\beta}_{ls} - \beta\right)^2\sum_{t=2}^{T} z_{t-1}^2x_{t-1}^2 \leq \left(\widehat{\beta}_{ls} - \beta\right)^2\left(\sup_{2\leq t\leq T}|x_{t-1}|\right)^2\sum_{t=2}^{T} z_{t-1}^2 = O_p\left(T^{2\delta}\right).$$

Similarly,

$$\begin{aligned}\sum_{t=2}^{T} \zeta_{t-1}z_{t-1}\widehat{u}_t^2 &= \sum_{t=2}^{T} x_{t-1}z_{t-1}\widehat{u}_t^2 - \sum_{t=2}^{T} z_{t-1}^2\widehat{u}_t^2 = O_p\left(T^{1.5+\delta}\right) + O_p\left(T^{1+2\delta}\right)\\ &= O_p\left(T^{1.5+\delta}\right)\end{aligned}$$

by using

$$\left|\sum_{t=2}^{T} x_{t-1}z_{t-1}\widehat{u}_t^2\right| \leq \sum_{t=2}^{T}\left|x_{t-1}z_{t-1}u_t^2\right| + 2\left|\widehat{\beta}_{ls} - \beta\right|\sum_{t=2}^{T} x_{t-1}^2\left|z_{t-1}u_t\right| + \left(\widehat{\beta}_{ls} - \beta\right)^2\sum_{t=2}^{T}\left|x_{t-1}^3z_{t-1}\right|$$

with

$$0 \leq \sum_{t=2}^{T}\left|x_{t-1}z_{t-1}u_t^2\right| \leq \sqrt{\sum_{t=2}^{T} x_{t-1}^2u_t^2\sum_{t=2}^{T} z_{t-1}^2u_t^2} \leq \sqrt{\left(\sup_{2\leq t\leq T} x_{t-1}^2\right)\sum_{t=2}^{T} u_t^2\sum_{t=2}^{T} z_{t-1}^2u_t^2} = O_p\left(T^{1/2}T^{1/2}T^{1/2+\delta}\right)$$

and

$$0 \leq \left|\widehat{\beta}_{ls} - \beta\right| \sum_{t=2}^{T} x_{t-1}^2 \left|z_{t-1} u_t\right| \leq \left|\widehat{\beta}_{ls} - \beta\right| \left(\sup_{2\leq t\leq T} x_{t-1}^2\right) \sum_{t=2}^{T} \left|z_{t-1} u_t\right| \leq \left|\widehat{\beta}_{ls} - \beta\right| \left(\sup_{2\leq t\leq T} x_{t-1}^2\right) \sqrt{T \sum_{t=2}^{T} z_{t-1}^2 u_t^2}$$
$$= O_p\left(T^{-1} T T^{1/2} T^{1/2+\delta}\right) = O_p\left(T^{1+\delta}\right).$$

Furthermore,

$$\begin{aligned} 0 \leq \left(\widehat{\beta}_{ls} - \beta\right)^2 \sum_{t=2}^{T} \left|x_{t-1}^3 z_{t-1}\right| &\leq \left(\widehat{\beta}_{ls} - \beta\right)^2 \left(\sup_{2\leq t\leq T} \left|x_{t-1}\right|^3\right) \sum_{t=2}^{T} \left|z_{t-1}\right| \\ &\leq \left(\widehat{\beta}_{ls} - \beta\right)^2 \left(\sup_{2\leq t\leq T} \left|x_{t-1}\right|^3\right) \sqrt{T \sum_{t=2}^{T} z_{t-1}^2} \\ &= O_p\left(T^{-2} T^{1.5} T^{1+\delta}\right) = O_p\left(T^{1/2+\delta}\right). \end{aligned}$$

Using $0 \leq \sum_{t=2}^{T} x_{t-1}^2 u_t^2 \leq \sqrt{\sum_{t=2}^{T} x_{t-1}^4 \sum_{t=2}^{T} u_t^4} = O_p\left(T^2\right)$, $\sum_{t=2}^{T} x_{t-1}^3 u_t \leq \sqrt{\sum_{t=2}^{T} x_{t-1}^6 \sum_{t=2}^{T} u_t^2} = O_p\left(T^{2.5}\right)$ and $\sum_{t=2}^{T} x_{t-1}^4 = O_p\left(T^3\right)$, we obtain

$$\sum_{t=2}^{T} x_{t-1}^2 \widehat{u}_t^2 = \sum_{t=2}^{T} x_{t-1}^2 u_t^2 - 2\left(\widehat{\beta}_{ls} - \beta\right) \sum_{t=2}^{T} x_{t-1}^3 u_t + \left(\widehat{\beta}_{ls} - \beta\right)^2 \sum_{t=2}^{T} x_{t-1}^4 = O_p\left(T^2\right)$$

and

$$\sum_{t=2}^{T} \zeta_{t-1}^2 \widehat{u}_t^2 = \sum_{t=2}^{T} x_{t-1}^2 \widehat{u}_t^2 - 2\sum_{t=2}^{T} x_{t-1} z_{t-1} \widehat{u}_t^2 + \sum_{t=2}^{T} z_{t-1}^2 \widehat{u}_t^2 = O_p\left(T^2\right).$$

Collecting all results we finally obtain

$$\begin{aligned} t_{va}^w &= \frac{\frac{1}{T^{1/2+\delta}} \sum_{t=2}^{T} z_{t-1} u_t}{\sqrt{\frac{1}{T^{1+2\delta}} \sum_{t=2}^{T} z_{t-1}^2 u_t^2}} + b\, \frac{\frac{1}{T^{1+2\delta}} \sum_{t=2}^{T} z_{t-1}^2}{\sqrt{\frac{1}{T^{1+2\delta}} \sum_{t=2}^{T} z_{t-1}^2 u_t^2}} + o_p(1), \\ &\xrightarrow{d} \mathcal{Z} + b\, \frac{V_z}{\sqrt{V_{zu}}} \end{aligned}$$

establishing the result.

Details for Remark 1

Applying standard regression algebra we obtain

$$\begin{aligned} t_{va} &= \frac{\sum_{t=2}^{T} z_{t-1} u_t - \sum_{t=2}^{T} z_{t-1}\zeta_{t-1} \left(\sum_{t=2}^{T} \zeta_{t-1}^2\right)^{-1} \sum_{t=2}^{T} \zeta_{t-1} u_t}{\widehat{\sigma}_u \sqrt{\sum_{t=2}^{T} z_{t-1}^2 - \sum_{t=2}^{T} z_{t-1}\zeta_{t-1} \left(\sum_{t=2}^{T} \zeta_{t-1}^2\right)^{-1} \sum_{t=2}^{T} \zeta_{t-1} z_{t-1}}} \\ &\quad + \frac{\beta}{\widehat{\sigma}_u} \sqrt{\sum_{t=2}^{T} z_{t-1}^2 - \sum_{t=2}^{T} z_{t-1}\zeta_{t-1} \left(\sum_{t=2}^{T} \zeta_{t-1}^2\right)^{-1} \sum_{t=2}^{T} \zeta_{t-1} z_{t-1}}\,. \end{aligned}$$

We know from the proof of Theorem 1 that

$$\sum_{t=2}^{T} z_{t-1}\zeta_{t-1} = o_p\left(T^{1.5+\delta}\right),$$

$$\sum_{t=2}^{T} \zeta_{t-1}^2 = O_p\left(T^{-2}\right)$$

and

$$\sum_{t=2}^{T} \zeta_{t-1}u_t = O_p\left(T\right).$$

Thus

$$t_{va} = \frac{\frac{1}{T^{1/2+\delta}}\sum_{t=2}^{T} z_{t-1}u_t}{\widehat{\sigma}_u\sqrt{\frac{1}{T^{1+2\delta}}\sum_{t=2}^{T} z_{t-1}^2}} + \frac{b}{\widehat{\sigma}_u}\sqrt{\frac{1}{T^{1+2\delta}}\sum_{t=2}^{T} z_{t-1}^2} + o_p\left(1\right),$$

as required for the result.

Proof of Corollary 1

To begin, note that $\left(\frac{1}{\sqrt{T}}\sum_{j=1}^{[rT]} u_t\,,\,\frac{1}{\sqrt{T}}\sum_{j=1}^{[rT]} v_t\right)' \Rightarrow \left(\sigma_u W_u\left(r\right)\,,\,\sigma_v W_v\left(r\right)\right)'$ where the r.h.s. is a bivariate Brownian motion with covariance Σ, and $\frac{1}{\sqrt{T}}x_{[rT]} \Rightarrow \sigma_v J_c\left(r\right)$; see e.g. Phillips and Durlauf (1986) and Phillips (1987).

We show that each of the variables satisfies Assumption 3 and the condition $T^{-1-2\delta}\sum_{t=2}^{T} z_{t-1}^2 \Rightarrow V_z$, which implies that the discussed variables obey Assumption 2. We take the route over Assumption 3 in order to avoid proving very similar results twice. Note that Assumption 3 involves the additional parameter $\vartheta < 1/2$ which does not play an explicit role for the VA procedure, but simplifies some of the derivations significantly.

For all four choices of z_t, the condition $\frac{1}{T^{1/2+\vartheta+\delta}}\sum_{j=1}^{[rT]} z_j \Rightarrow Z(r)$ implies $\frac{1}{T^{1.5+\delta}}\sum_{t=2}^{T} z_{t-1}x_{t-1} \xrightarrow{p} 0$. Concretely, it is shown in the proof of Theorem 2 that

$$\frac{1}{T^{\vartheta+\delta+1}}\sum_{t=2}^{T} z_{t-1}x_{t-1} \Rightarrow R_{zx}^c,$$

the key ingredient in establishing convergence to R_{zx}^c being precisely the weak convergence $\frac{1}{T^{1/2+\vartheta+\delta}}\sum_{j=1}^{[rT]} z_j \Rightarrow Z(r)$. Then, given that $\vartheta < 1/2$, $\frac{1}{T^{1.5+\delta}}\sum_{t=2}^{T} z_{t-1}x_{t-1} \xrightarrow{p} 0$ follows immediately. Also, the i.i.d. property of $(u_t, v_t)'$ and the weak convergence of the partial sums of z_t implies

$$\frac{1}{T^{1+\delta+\vartheta}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t-1} z_j\right)v_t \Rightarrow \int_0^1 Z(r)\mathrm{d}B_{H,v}(r)$$

(Hansen, 1992) such that $\lambda = 0$.

1. Short-memory We write

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{[rT]} z_j = \frac{1}{\sqrt{T}}\sum_{j=0}^{[rT]-1} \bar{\alpha}^j x_{[rT]-j}.$$

Employing the Beveridge-Nelson decomposition and rearranging the sum terms (see also the proof for the fractional instrument below) leads to

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{[rT]} z_j \Rightarrow \frac{\sigma_v}{1-\bar{\alpha}} J_c(r) = Z(r).$$

Note that $z_t = v_t^* - \frac{c}{T}x_{t-1}^*$ where $v_t^* = (1-\bar{\alpha}L)_+^{-1}v_t$ and $x_t^* = (1-\bar{\alpha}L)_+^{-1}x_t$; v_t^* is (asymptotically) $I(0)$ whereas x_{t-1} is nearly integrated. It follows that

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T} z_{t-1}^2 &= \frac{1}{T}\sum_{t=2}^{T}(v_{t-1}^*)^2 + O_p(T^{-1}) \\ &\overset{p}{\to} \frac{\sigma_v^2}{1-\bar{\alpha}^2} = V_z.\end{aligned}$$

Furthermore,

$$\frac{1}{T}\sum_{t=2}^{T} z_{t-1}^2 u_t^2 = \sigma_u^2 \frac{1}{T}\sum_{t=2}^{T}(v_{t-1}^*)^2 + \frac{1}{T}\sum_{t=2}^{T}(v_{t-1}^*)^2(u_t^2-\sigma_u^2) + o_p(1)$$

implying convergence in probability to $\sigma_u^2\sigma_v^2/(1-\bar{\alpha}^2) = V_{zu}$, since $\mathrm{E}(u_t^2-\sigma_u^2|u_{t-1}, v_{t-1}, \ldots) = 0$ under the assumptions of the corollary. This also implies that $\delta = 0$ which, together with the $\sqrt{T}$ normalization for the partial sums of z_t, leads to $\vartheta = 0$.

Note that the difference between v_t^* and $v_t/(1-\bar{\alpha}L)$ vanishes fast enough as $t \to \infty$ and that $T^{-1/2}x_{t-2}^* u_t$ is a martingale difference sequence with bounded variance, so

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=2}^{T} z_{t-1}u_t &= \frac{1}{\sqrt{T}}\sum_{t=2}^{T} v_{t-1}^* u_t - \frac{c}{T}\sum_{t=2}^{T}\frac{x_{t-2}^*}{\sqrt{T}}u_t \\ &= \frac{1}{\sqrt{T}}\sum_{t=2}^{T} v_{t-1}^* u_t + o_p(1).\end{aligned}$$

Furthermore, $v_{t-1}^* u_t$ and u_t are uncorrelated since v_{t-1}^* is independent of u_t. For the same reason $(v_{t-1}^* u_t)$ and v_t are uncorrelated too. A multivariate invariance principle applies to the cumulated sums of $(v_{t-1}^* u_t, u_t, v_t)'$ under our conditions. Joint convergence with $Z(r)$ follows immediately given that $\frac{\sigma_v}{1-\bar{\alpha}} J_c(r) = Z(r)$.

2. Mild integration We need to establish the behavior of four terms,

$$\frac{1}{T^{1/2+\eta}}\sum_{t=1}^{[rT]} z_t, \quad V_{T,z} = \frac{1}{T^{1+\eta}}\sum_{t=2}^{T} z_{t-1}^2, \quad \frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{[rT]} z_{t-1}u_t \quad \text{and} \quad V_{T,zu} = \frac{1}{T^{1+\eta}}\sum_{t=2}^{T} z_{t-1}^2 u_t^2.$$

Clearly, $\delta = \eta/2$ from the normalization of $V_{T,z}$ and thus $\vartheta = \eta/2$ as well.

For establishing the behavior of the first term, note that

$$\frac{1}{T^{1/2+\eta}}\sum_{t=1}^{[rT]} z_t = \frac{1}{T^{1/2+\eta}}\sum_{j=0}^{[rT]-1} \alpha_T^j x_{[rT]-j}.$$

Following Phillips and Solo (1992), it is straightforward to obtain that

$$\sum_{j=1}^{t} z_j = \left(\sum_{j=0}^{t-1} \alpha_T^j \right) x_t + \sum_{j=1}^{t-1} \alpha_T^j \left(x_{t-j} - x_t \right).$$

Note that $\mathrm{E}\left| x_t - x_{t-j} \right| \leq C\sqrt{j}$, leading with $0 \leq \alpha_T < 1$ to

$$\mathrm{E}\left| \frac{1}{T^{1/2+\eta}} \sum_{j=1}^{t-1} \alpha_T^j \left(x_{t-j} - x_t \right) \right| \leq \frac{C}{T^{1/2+\eta}} \sum_{j=1}^{t-1} \sqrt{j}\, \alpha_T^j \leq \frac{C}{T^{1/2+\eta}} \sum_{t=1}^{T} \sqrt{t}\, \alpha_T^t.$$

To obtain an upper bound for the latter, write

$$\frac{C}{T^{1/2+\eta}} \sum_{t=1}^{T} \sqrt{t}\, \alpha_T^t = \frac{C}{T^{1/2+\eta}} \sum_{t=1}^{h_T} \sqrt{t}\, \alpha_T^t + \frac{C}{T^{1/2+\eta}} \sum_{t=h_T+1}^{T} \sqrt{t}\, \alpha_T^t$$

with $h_T = CT^\lambda$ for some $\eta < \lambda < 1$. Note that

$$\alpha_T^{T^\lambda} = \left(1 - \frac{a}{T^\eta}\right)^{T^\lambda} = \left(\left(1 - \frac{a}{T^\eta}\right)^{-T^\eta/a} \right)^{-a\, T^\lambda / T^\eta}.$$

For any $\eta < \lambda \leq 1$ and $T > T_0$, there exists a constant $C_* > 1$ such that

$$\alpha_T^{T^\lambda} \leq C_*^{-a\, T^{\lambda-\eta}} \to 0.$$

Thus, since $\alpha_T^{h_T+1} \to 0$, it holds that

$$\begin{aligned} 0 \leq \frac{C}{T^{1/2+\eta}} \sum_{t=1}^{h_T} \sqrt{t}\, \alpha_T^t \quad &\leq \quad \sqrt{\frac{h_T}{T}} \frac{C}{T^\eta} \sum_{t=1}^{h_T} \alpha_T^t = \sqrt{\frac{h_T}{T}} \frac{C}{T^\eta} \frac{1 - \alpha_T^{h_T+1}}{1 - \alpha_T} \\ &\leq \quad C\sqrt{\frac{h_T}{T}} \to 0. \end{aligned}$$

Furthermore,

$$0 \leq \frac{C}{T^{1/2+\eta}} \sum_{t=h_T+1}^{T} \sqrt{t}\, \alpha_T^t \leq \frac{C\alpha_T^{h_T}}{T^{1/2+\eta}} \sum_{t=h_T+1}^{T} \sqrt{t} \leq C\alpha_T^{h_T} T^{1-\eta}.$$

Recall that $\alpha_T^{h_T}$ vanishes at a rate exponential in $T^{\lambda-\eta}$, so $\alpha_T^{h_T} \left(T^{\lambda-\eta}\right)^{\frac{1-\eta}{\lambda-\eta}} \to 0$ (given that $\frac{1-\eta}{\lambda-\eta}$ is fixed) leading to

$$\frac{1}{T^{1/2+\eta}} \sum_{t=1}^{[rT]} z_t = \frac{1}{T^{1/2+\eta}} \frac{1 - \alpha_T^{[rT]}}{1 - \alpha_T} x_{[rT]} + o_p(1).$$

Using

$$\frac{1}{T^{1/2+\eta}} \frac{1 - \alpha_T^{[rT]}}{1 - \alpha_T} x_{[rT]} = \frac{1}{a\, T^{1/2}} x_{[rT]} - \frac{\alpha_T^{[rT]}}{a\, T^{1/2}} x_{[rT]}$$

and $\alpha_T^{[rT]} \to 0$ yields

$$\frac{1}{T^{1/2+\eta}} \sum_{t=1}^{[rT]} z_t \Rightarrow \frac{\sigma_v}{a} J_c(r) = Z(r)$$

on $[r, 1]$ for all $0 < r < 1$. The weak convergence is extended to $[0, 1]$ by noting that $\frac{1}{T^{1/2+\eta}} \sum_{t=1}^{[rT]} z_t = o_p(1)$ as $r \to 0$.

The asymptotic behavior of the second expression was derived by Phillips and Magdalinos (2007, Equation (4)),

$$\frac{1}{T^{1+\eta}}\sum_{t=1}^{T} z_t^2 \to \frac{\sigma_v^2}{2a} = V_z.$$

When deriving the limit of the third term, $\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{[rT]} z_{t-1}u_t$, we resort to arguments similar to those of the proof in Lemma 3.1 of Phillips and Magdalinos (2009). Their Lemma 3.1 is not directly applicable since they require $\eta > 1/2$. Our predictive regression model exhibits however weak exogeneity, which they do not assume, and we are thus able to weaken the condition to $0 < \eta < 1$.

Let $z_t = \sum_{j=0}^{t-1}\alpha_T^j \Delta x_{t-j} = \sum_{j=0}^{t-1}\alpha_T^j v_{t-j} + \frac{c}{T}\sum_{j=0}^{t-1}\alpha_T^j x_{t-1-j}$, and define, as in Phillips and Magdalinos (2009),

$$z_t^{(0)} = \sum_{j=0}^{t-1}\alpha_T^j v_{t-j} = \alpha_T z_{t-1}^{(0)} + v_t$$

where $z_0^{(0)} = 0$, and $\psi_t = \frac{c}{T}\sum_{j=0}^{t-1}\alpha_T^j x_{t-1-j}$ with $x_0 = 0$ for simplicity, such that

$$\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T} z_{t-1}u_t = \frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T} z_{t-1}^{(0)}u_t + \frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\psi_{t-1}u_t.$$

Using Lemma 3.2 of Phillips and Magdalinos (2009), the first term on the r.h.s. converges to the required normal distribution

$$\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T} z_{t-1}^{(0)}u_t \Rightarrow \mathcal{N}\left(0, \sigma_u^2 V_z\right),$$

and we show in the following that the second term vanishes for any $0 < \eta < 1$.

Note that $\psi_{t-1}u_t$ possesses the md property, so it is an uncorrelated sequence and

$$\operatorname{Var}\left(\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\psi_{t-1}u_t\right) = \frac{1}{T^{1+\eta}}\sum_{t=2}^{T}\operatorname{Var}\left(\psi_{t-1}u_t\right).$$

Then, due to the independence of $(u_t, v_t)'$,

$$\begin{aligned}\operatorname{Var}\left(\psi_{t-1}u_t\right) &= \sigma_u^2 \operatorname{Var}\left(\psi_{t-1}\right) = \sigma_u^2 \operatorname{E}\left(\left(\frac{c}{T}\sum_{j=0}^{t-1}\alpha_T^j x_{t-1-j}\right)^2\right). \\ &= \frac{c^2\sigma_u^2}{T^2}\sum_{j=0}^{t-1}\sum_{k=0}^{t-1}\alpha_T^j\alpha_T^k \operatorname{E}\left(x_{t-1-j}x_{t-1-k}\right);\end{aligned}$$

since $\operatorname{E}\left(x_{t-1-j}x_{t-1-k}\right) \le CT$, it follows that

$$\operatorname{Var}\left(\psi_{t-1}u_t\right) \le C\frac{1}{T^2}\sum_{j=0}^{t-1}\sum_{k=0}^{t-1}\alpha_T^j\alpha_T^k = C\frac{1}{T^2}\left(\sum_{j=0}^{t-1}\alpha_T^j\right)^2.$$

With $\sum_{j=0}^{t-1}\alpha_T^j = \frac{1-\alpha_T^t}{1-\alpha_T} = O\left(T^\eta\right)$, it follows that

$$\operatorname{Var}\left(\psi_{t-1}u_t\right) \le CT^{2\eta-2}.$$

Thus,

$$0 < \mathrm{Var}\left(\frac{1}{T^{1/2+\eta/2}}\sum_{t=2}^{T}\psi_{t-1}u_t\right) \leq C\frac{T^{2\eta-2}}{T^{1+\eta}} = CT^{\eta-1} \to 0$$

for any $0 < \eta < 1$ as required.

For the fourth term, write

$$\frac{1}{T^{1+\eta}}\sum_{t=2}^{T} z_{t-1}^2 u_t^2 = \sigma_u^2 \frac{1}{T^{1+\eta}}\sum_{t=2}^{T}(z_{t-1}^0)^2 + \frac{1}{T^{1+\eta}}\sum_{t=2}^{T}(z_{t-1}^0)^2(u_t^2 - \sigma_u^2) + o_p(1).$$

The limit $V_{zu} = \sigma_u^2 V_z$ follows if $\frac{1}{T^{1+\eta}}\sum_{t=2}^{T}(z_{t-1}^0)^2(u_t^2-\sigma_u^2) \xrightarrow{p} 0$ since $(z_{t-1}^0)^2(u_t^2-\sigma_u^2)$ has the md property under the assumptions of the corollary. Hence it suffices to show that $\mathrm{Var}\left(T^{-\eta}(z_{t-1}^0)^2\right)$ is uniformly bounded. This is the case when $T^{-2\eta}\,\mathrm{E}\left((z_{t-1}^0)^4\right)$ is itself uniformly bounded and

$$\mathrm{E}\left((z_{t-1}^0)^4\right) = \sum_{j=1}^{t-1}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}\sum_{m=1}^{t-1}\alpha_T^j\alpha_T^k\alpha_T^l\alpha_T^m\,\mathrm{E}\left(v_{t-j}v_{t-k}v_{t-l}v_{t-m}\right).$$

With $v_t \sim iid(0,\sigma^2)$, the expectation $\mathrm{E}\left(v_{t-j}v_{t-k}v_{t-l}v_{t-m}\right)$ is nonzero only if the indices j, k, l and m are pairwise equal, and is also uniformly bounded, so for all t

$$\mathrm{E}\left((z_{t-1}^0)^4\right) = O\left(\sum_{j=1}^{T}\sum_{k=1}^{T}\alpha_T^{2j}\alpha_T^{2k}\right) = O\left(\left(\sum_{j=1}^{T}\alpha_T^{2j}\right)^2\right).$$

The result follows with $\sum_{j=1}^{T}\alpha_T^{2j} = \frac{1-\alpha_T^{2T}}{1-\alpha_T^2} = O\left(T^\eta\right)$ since $\alpha_T = 1 - \frac{2a}{T^\eta} - \frac{a^2}{T^{2\eta}} \approx 1 - \frac{2a}{T^\eta}$.

3. Fractional integration Note first that

$$\sum_{j=1}^{[rT]}\Delta_+^{1-d^*}x_j = \Delta_+^{-d^*}x_{[rT]};$$

assume for simplicity that $x_0 = 0$ and note further that

$$\Delta_+^{-d^*}x_t = \sum_{j=0}^{t-1}\delta_j^* x_{t-j} = \sum_{j=0}^{t-1}\delta_j^*\sum_{i=0}^{t-j-1}\varrho^i v_{t-j-i}.$$

After rearranging the sums we obtain

$$\Delta_+^{-d^*}x_t = \sum_{j=0}^{t-1}\varrho^j\Delta_+^{-d^*}v_{t-j}.$$

Then,

$$\sum_{j=0}^{t-1}\varrho^j\Delta_+^{-d^*}v_{t-j} = \sum_{j=0}^{t-1}\left(1-\frac{c}{T}\right)^j\left(\sum_{k=j}^{t-1}\Delta_+^{-d^*}v_{t-k} - \sum_{k=j+1}^{t-1}\Delta_+^{-d^*}v_{t-k}\right);$$

Let $\sum_1^0 \cdot = 0$; summation by parts thus leads to

$$\begin{aligned}\sum_{j=0}^{t-1}\left(1-\frac{c}{T}\right)^j \Delta_+^{-d^*} v_{t-j} &= \sum_{j=0}^{t-1}\Delta_+^{-d^*} v_{t-j} + \sum_{k=t-1}^{1}\left(\left(1-\frac{c}{T}\right)^{t-k} - \left(1-\frac{c}{T}\right)^{t-k-1}\right)\sum_{k=j}^{t-1}\Delta_+^{-d^*} v_{t-k} \\ &= \sum_{j=0}^{t-1}\Delta_+^{-d^*} v_{t-j} - \frac{c}{T}\sum_{k=1}^{t-1}\left(\left(1-\frac{c}{T}\right)^{-\frac{T}{c}}\right)^{-\frac{c}{T}(t-k-1)}\sum_{j=0}^{k-1}\Delta_+^{-d^*} v_{k-j}.\end{aligned}$$

Using $\left(\left(1-\frac{c}{T}\right)^{-\frac{T}{c}}\right)^{-\frac{c}{T}([rT]-[sT]-1)} \Rightarrow e^{-c(r-s)}$ and $\frac{1}{T^{d^*-1/2}}\Delta_+^{-d^*} v_{[rT]} \Rightarrow \sigma_v W_v^{d^*+1}(r)$ for $d \in (0.5, 1.5)$ (Marinucci and Robinson, 2000), the required convergence

$$Z(r) = \sigma_v J_c^{1+d^*}(r)$$

follows with the CMT along the lines of Phillips (1987, Lemma 1a). Furthermore,

$$z_{t-1}^2 = \left(\Delta_+^{-d^*} v_{t-1}\right)^2 - 2\frac{c}{\sqrt{T}}\Delta_+^{-d^*} v_{t-1}\frac{x_{t-2}}{\sqrt{T}} + \frac{c^2}{T}\left(\Delta_+^{-d^*}\frac{x_{t-2}}{\sqrt{T}}\right)^2$$

and note that Lemma 2 (17) in (Dolado et al., 2002) implies that

$$\frac{1}{T}\sum\left(\Delta_+^{-d^*} v_{t-1}\right)^2 \xrightarrow{p} \sigma_v^2\frac{\Gamma(1-2d^*)}{\Gamma^2(1-d^*)}.$$

Then, $\left\|T^{-1/2}x_{t-2}\right\|_2 \le C$ so Minkowski's norm inequality yields

$$\left\|\Delta_+^{-d^*} T^{-1/2} x_{t-2}\right\|_2 \le \sum_{j=1}^{t-2}|\delta_j^*|\left\|T^{-1/2}x_j\right\|_2 \le Ct^{d^*}\sum_{j=1}^{t-2}\frac{1}{j}$$

where δ_j^* are the coefficients of the filter $\Delta_+^{-d^*}$. Using the logarithmic approximation for the harmonic sum, we obtain equivalently

$$\mathrm{E}\left(\left|\Delta_+^{-d^*} T^{-1/2} x_{t-2}\right|^2\right) \le Ct^{2d^*}\log^2(t-2),$$

leading to

$$\mathrm{E}\left(\left|\sum_{t=2}^{T}\left(\Delta_+^{-d^*}\frac{x_{t-2}}{\sqrt{T}}\right)^2\right|\right) \le CT^{1+2d^*+2\epsilon}$$

for some $0 < \epsilon < (1/2 - d^*)/2$. Thus, with $d^* + \epsilon < 1/2$,

$$\frac{1}{T}\sum_{t=2}^{T}\frac{c^2}{T}\left(\Delta_+^{-d^*}\frac{x_{t-2}}{\sqrt{T}}\right)^2 \xrightarrow{p} 0.$$

Moreover, the Cauchy-Schwarz inequality leads to

$$\mathrm{E}\left(\frac{1}{\sqrt{T}}\left|\Delta_+^{-d^*} v_{t-1}\Delta_+^{-d^*}\frac{x_{t-2}}{\sqrt{T}}\right|\right) \le \frac{1}{\sqrt{T}}\left\|\Delta_+^{-d^*} v_{t-1}\right\|_2\left\|\Delta_+^{-d^*}\frac{x_{t-2}}{\sqrt{T}}\right\|_2 \le CT^{d^*-1/2}\log T$$

with $\|\cdot\|_2$ the L_2-norm of a random variable, $\sqrt{\mathrm{E}(|\cdot|^2)}$, so we have

$$\frac{1}{T}\sum_{t=2}^{T} z_{t-1}^2 \xrightarrow{p} \sigma_v^2\frac{\Gamma(1-2d^*)}{\Gamma^2(1-d^*)}$$

and $V_z = \sigma_v^2 \frac{\Gamma(1-2d^*)}{\Gamma^2(1-d^*)} = V_z$. Thus, $\delta = 0$ and $\vartheta = d^*$.

For $V_{T,zu}$ we obtain have along the same lines

$$\frac{1}{T}\sum_{t=2}^{T} z_{t-1}^2 u_t^2 = \frac{1}{T}\sum_{t=2}^{T} (\Delta_+^{-d^*} v_{t-1})^2 u_t^2 + o_p(1)$$

such that the desired limit arises if $\frac{1}{T}\sum_{t=2}^{T}(\Delta_+^{-d^*} v_{t-1})^2 \left(u_t^2 - \sigma_u^2\right) \xrightarrow{p} 0$. This follows from the md property of the summands if $\mathrm{E}\left((\Delta_+^{-d^*} v_{t-1})^4\right)$ is uniformly bounded. It follows that

$$\mathrm{E}\left((\Delta_+^{-d^*} v_{t-1})^4\right) = \sum_{j=1}^{t-1}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}\sum_{m=1}^{t-1} \delta_j^* \delta_k^* \delta_l^* \delta_m^* \, \mathrm{E}\left(v_{t-j} v_{t-k} v_{t-l} v_{t-m}\right).$$

Due to the i.i.d. property of v_t, the expectation is nonzero only for pairwise equal indices, thus

$$\mathrm{E}\left((\Delta_+^{-d^*} v_{t-1})^4\right) = \sum_{j=1}^{t-1}\sum_{k=1}^{t-1} (\delta_j^*)^2 (\delta_k^*)^2 \, \mathrm{E}\left(v_{t-j}^2 v_{t-k}^2\right).$$

Since the coefficients δ_j^* are square summable and $\mathrm{E}\left(v_{t-j}^2 v_{t-k}^2\right)$ is uniformly bounded it follows from the Cauchy-Schwarz inequality that the fourth moment of $\Delta_+^{-d^*} v_{t-1})^4$ is is indeed uniformly bounded, as required.

Then, similarly to the case of the short-memory instrument,

$$\frac{1}{\sqrt{T}} \sum z_{t-1} u_t = \frac{1}{\sqrt{T}} \sum_{t=2}^{T} (\Delta_+^{-d^*} v_{t-1}) u_t + o_p(1)$$

and $(\Delta_+^{-d^*} v_{t-1}) u_t$ is uncorrelated with u_t as well as with v_t. Since the fourth moment of $\Delta_+^{-d^*} v_{t-1}$ were shown to be bounded,it follows that the third moment of $z_{t-1} u_t$. is bounded as well. Hence, a multivariate invariance principle applies. Finally, $\Delta_+^{-d^*} x_{[rT]}$ being driven by v_t, $G(r)$, is independent of W_u and W_v.

4. Long differences

Let w.l.o.g. $k_T = [KT^\nu]$ and $x_t = 0$ for $t \le 0$. With this convention, we have

$$\begin{aligned}\frac{1}{k_T\sqrt{T}} \sum_{j=1}^{[rT]} z_j &= \frac{1}{k_T\sqrt{T}} \sum_{j=1}^{[rT]} (x_j - x_{j-k_T}) = \frac{1}{k_T\sqrt{T}} \sum_{k=0}^{k_T-1} x_{[rT]-k} \\ &= \frac{1}{k_T\sqrt{T}} k_T x_{[rT]} + \frac{1}{k_T\sqrt{T}} \sum_{k=1}^{k_T-1} \left(x_{[rT]-k} - x_{[rT]}\right).\end{aligned}$$

Then, $\sum_{k=1}^{k_T-1}\left(x_{[rT]-k} - x_{[rT]}\right) = \sum_{k=1}^{k_T-1}(k - k_T)\,\Delta x_{[rT]-k+1} = O_p\left(k_T^{1.5}\right)$, so

$$\frac{1}{k_T\sqrt{T}} \sum_{j=1}^{[rT]} z_j = \frac{1}{\sqrt{T}} x_{[rT]} + O_p\left(\sqrt{\frac{k_T}{T}}\right)$$

with the O_p term not depending on r, leading in turn to

$$\frac{1}{T^{1/2+\delta+\vartheta}} \sum_{j=1}^{[rT]} z_j \Rightarrow K\sigma_v J_c(r) = Z(r)$$

with $\delta = \vartheta = \nu/2$.

Let us now examine

$$\begin{aligned}\frac{1}{k_T T}\sum_{j=1}^{[rT]} z_j^2 &= \frac{1}{k_T T}\sum_{j=1}^{[rT]}\left(\sum_{k=0}^{k_T-1}\Delta x_{j-k}\right)^2\\ &= \frac{1}{k_T T}\sum_{j=1}^{[rT]}\sum_{k=0}^{k_T-1}\left(\Delta x_{j-k}\right)^2+\frac{1}{k_T T}\sum_{j=1}^{[rT]}\sum\sum_{1\le k\ne l\le k_T-1}\Delta x_{j-k}\Delta x_{j-l}\end{aligned}$$

For the first sum term, we obtain that

$$\begin{aligned}\frac{1}{k_T T}\sum_{j=1}^{[rT]}\sum_{k=0}^{k_T-1}\left(\Delta x_{j-k}\right)^2 &= \frac{1}{T}\sum_{j=1}^{[rT]}\left(\Delta x_j\right)^2+o_p\left(1\right)=\frac{1}{T}\sum_{j=1}^{[rT]}v_j^2+o_p\left(1\right)\\ &= \sigma_v^2\frac{[rT]}{T}+\frac{1}{T}\sum_{j=1}^{[rT]}\left(v_j^2-\sigma_v^2\right)+o_p\left(1\right)\end{aligned}$$

converging to the desired limit due to the martingale difference property of $v_j^2-\sigma_v^2$. The second sum is rearranged to obtain

$$\frac{1}{k_T T}\sum_{j=1}^{[rT]}\sum\sum_{1\le k\ne l\le k_T-1}\Delta x_{j-k}\Delta x_{j-l}=\frac{1}{k_T T}\sum_{j=1}^{[rT]}\Delta x_j\left(\sum_{k=1}^{k_T-1}\left(k_T-k\right)\Delta x_{j-k}\right)+o_p\left(1\right),$$

and recall that $\Delta x_j = v_j-(c/T)x_{j-1}$. Tedious, yet straightforward evaluations lead to the conclusion that $\frac{c}{k_T T^2}\sum_{j=1}^{[rT]}x_{j-1}\left(\sum_{k=1}^{k_T-1}\left(k_T-k\right)\Delta x_{j-k}\right)=o_p\left(1\right)$, so

$$\frac{1}{k_T T}\sum_{j=1}^{[rT]}\sum\sum_{1\le k\ne l\le k_T-1}\Delta x_{j-k}\Delta x_{j-l}=\frac{1}{k_T T}\sum_{j=1}^{[rT]}v_j\left(\sum_{k=1}^{k_T-1}\left(k_T-k\right)\Delta x_{j-k}\right)+o_p\left(1\right).$$

Furthermore, $v_j\left(\sum_{k=1}^{k_T-1}\left(k_T-k\right)\Delta x_{j-k}\right)$ are the elements of an md array, and thus

$$\mathrm{Var}\left(\sum_{j=1}^{[rT]}v_j\left(\sum_{k=1}^{k_T-1}\left(k_T-k\right)\Delta x_{j-k}\right)\right)=\sum_{j=1}^{[rT]}\mathrm{Var}\left(v_j\left(\sum_{k=1}^{k_T-1}\left(k_T-k\right)\Delta x_{j-k}\right)\right);$$

with

$$\mathrm{Var}\left(v_j\left(\sum_{k=1}^{k_T-1}\left(K_T-k\right)\Delta x_{j-k}\right)\right)\le Ck_T^3,$$

Consequently we obtain

$$\mathrm{Var}\left(\sum_{j=1}^{[rT]}v_j\left(\sum_{k=1}^{k_T-1}\left(k_T-k\right)\Delta x_{j-k}\right)\right)\le Ck_T^3[rT].$$

Summing up,

$$\frac{1}{k_T T}\sum_{j=1}^{[rT]}\sum\sum_{1\le k\ne l\le k_T-1}\Delta x_{j-k}\Delta x_{j-l}=O_p\left(\sqrt{\frac{k_T}{T}}\right)+o_p\left(1\right)=o_p\left(1\right)$$

yields

$$\frac{1}{T^{1+2\delta}}\sum_{j=1}^{[rT]}z_j^2\Rightarrow K\sigma_v^2 r$$

which implies that

$$V_z = K\sigma_v^2.$$

Finally, apply a functional central limit theorem for md arrays (Davidson, 1994, Theorem 27.14) to establish the limiting behavior of

$$\frac{1}{\sigma_u \sigma_v \sqrt{k_T T}} \sum_{t=2}^{[rT]} z_{t-1} u_t.$$

To this end, we need to check

1. that $\frac{1}{\sigma_u^2 \sigma_v^2 k_T T} \sum_{t=2}^{T} z_{t-1}^2 u_t^2 \xrightarrow{p} 1$,
2. that $\max_t \left| \frac{1}{\sigma_u \sigma_v \sqrt{k_T T}} z_{t-1} u_t \right| \xrightarrow{p} 0$, and that
3. $\frac{1}{\sigma_u^2 \sigma_v^2 k_T T} \sum_{t=2}^{[rT]} \mathrm{Var}\left(z_{t-1} u_t\right) \xrightarrow{p} r$.

It is straightforward to check condition 1 since

$$\frac{1}{Tk_T} \sum_{t=2}^{[rT]} z_{t-1}^2 u_t^2 = \frac{\sigma_u^2}{Tk_T} \sum_{t=2}^{[rT]} z_{t-1}^2 + \frac{1}{Tk_T} \sum_{t=2}^{[rT]} z_{t-1}^2 \left(u_t^2 - \sigma_u^2\right),$$

with the first summand converging to $r\sigma_v^2 \sigma_u^2$ and the second vanishing due to the md property of the elements in the sum. This also establishes the existence of the limit of $V_{T,zu}$ with $V_{zu} = \sigma_u^2 V_z$.

The second condition follows from

$$\mathrm{E} \left| \frac{z_{t-1} u_t}{\sigma_u \sigma_v \sqrt{k_T}} \right|^3 < \infty,$$

which is implied by the fact that the kurtosis of the i.i.d. sequences u_t and v_t is bounded. The global homoskedasticity condition is easily established exploiting the independence of u_t and v_{t-j} for any $j > 0$. The increments $z_{t-1} u_t$ are obviously uncorrelated with u_t or v_t, so $G(r)$ is indeed independent of W_u and W_v.

Proof of Theorem 2

Begin by letting $S_t = \sum_{j=1}^{t} z_j$ denote the partial sums of z_t with $S_0 = 0$, so $T^{-1/2-\vartheta-\delta} S_{[rT]} \Rightarrow Z(r)$, where $Z(r) = \int_0^r \dot{Z}(s) \mathrm{d}s$ for type-II instruments. Then

$$\begin{aligned} \sum_{t=2}^{T} z_{t-1} x_{t-1} &= \sum_{t=2}^{T} \left(S_{t-1} - S_{t-2}\right) x_{t-1} = S_{T-1} x_{T-1} - S_0 x_0 - \sum_{t=2}^{T-1} S_{t-1} \Delta x_t \\ &= S_{T-1} x_{T-1} - \sum_{t=2}^{T-1} S_{t-1} v_t + \frac{c}{T} \sum_{t=2}^{T-1} S_{t-1} x_{t-1} \end{aligned}$$

and the weak convergence

$$\frac{1}{T^{\vartheta+\delta+1}} \sum_{t=2}^{T} z_{t-1} x_{t-1} \Rightarrow \sigma_v \left(Z(1)\, J_{c,H,v}(1) - \int_0^1 Z(r) \mathrm{d}B_{H,v}(r) - \lambda + c \int_0^1 Z(r)\, J_{c,H,v}(r) \mathrm{d}r \right) = R_{zx}^c$$

follows with Assumptions 3 or 4 and the CMT.

The t-statistic is then given under the local alternative by

$$\begin{aligned} t_{iv}^{w} &= \frac{\sum_{t=2}^{T} z_{t-1}u_t}{\sqrt{\sum_{t=2}^{T} z_{t-1}^2 \widehat{u}_t^2}} + \frac{b}{T^{1/2+\vartheta}} \frac{\sum_{t=2}^{T} z_{t-1}x_{t-1}}{\sqrt{\sum_{t=2}^{T} z_{t-1}^2 \widehat{u}_t^2}} \\ &= \frac{\frac{1}{T^{\delta+1/2}}\sum_{t=2}^{T} z_{t-1}u_t}{\sqrt{\frac{1}{T^{2\delta+1}}\sum_{t=2}^{T} z_{t-1}^2 u_t^2 + o_p(1)}} + b\,\frac{\frac{1}{T^{\delta+\vartheta+1}}\sum_{t=2}^{T} z_{t-1}x_{t-1}}{\sqrt{\frac{1}{T^{2\delta+1}}\sum_{t=2}^{T} z_{t-1}^2 u_t^2 + o_p(1)}}. \end{aligned} \quad (19)$$

Since we know from the proof of Theorem 1 that $\sum_{t=2}^{T} z_{t-1}^2 \widehat{u}_t^2 = \sum_{t=2}^{T} z_{t-1}^2 u_t^2 + o_p\left(T^{2\delta+1}\right)$ for type-I instruments, and, for type-II instruments,

$$\frac{1}{T^{2\delta+1}}\sum_{t=2}^{T} z_{t-1}^2 \widehat{u}_t^2 = \frac{1}{T^{2\delta+1}}\sum_{t=2}^{T} z_{t-1}^2 u_t^2 - \frac{2\left(\widehat{\beta}_{ls}-\beta\right)}{T^{2\delta+1}}\sum_{t=2}^{T} z_{t-1}^2 x_{t-1}u_t + \frac{\left(\widehat{\beta}_{ls}-\beta\right)^2}{T^{2\delta+1}}\sum_{t=2}^{T} z_{t-1}^2 x_t^2$$

with $\widehat{\beta}_{ls}-\beta = O_p\left(T^{-1}\right)$ and

$$\frac{1}{T^{2\delta+1}}\sum_{t=2}^{T} z_{t-1}^2 x_{t-1}u_t \le \sqrt{\frac{1}{T^{4\delta+2}}\sum_{t=2}^{T} z_{t-1}^4 x_{t-1}^2 \sum_{t=2}^{T} u_t^2} = O_p\left(\sqrt{T}\right) \quad \frac{1}{T^{2\delta+2}}\sum_{t=2}^{T} z_{t-1}^2 x_t^2 \Rightarrow \int_0^1 \left(\dot{Z}(r)\right)^2 \left(J_{c,H,v}(r)\right)^2 \mathrm{d}r.$$

Due to Assumptions 3 or 4, $\frac{1}{T^{2\delta+1}} \sum_{t=2}^{T} z_{t-1}^2 u_t^2 \xrightarrow{p} V_{zu}$ and, as shown above, $\frac{1}{T^{\vartheta+\delta+1}} \sum_{t=2}^{T} z_{t-1}x_{t-1} \Rightarrow R_{zx}^c$. Therefore, the second ratio in (19) converges to $b\,R_{zx}^c/\sqrt{V_{zu}}$ due to the CMT. Moreover, given convergence to $G_I(1)$ from Assumption 3 or the limiting (mixed) normality from Assumption 4, we have for both types of instruments

$$\frac{\frac{1}{T^{\delta+1/2}}\sum_{t=2}^{T} z_{t-1}u_t}{\sqrt{\frac{1}{T^{2\delta+1}}\sum_{t=2}^{T} z_{t-1}^2 \widehat{u}_t^2}} \xrightarrow{d} \mathcal{N}(0,1).$$

For type-II instruments, we have

$$\frac{1}{T^{\delta+1/2}}\sum_{t=2}^{T} z_{t-1}u_t \Rightarrow \int_0^1 \dot{Z}(r)\mathrm{d}B_{H,u}(r)$$

which is mixed Gaussian, so V_{zu} is given by the quadratic variation of the Ito integral over $[0,1]$, $V_{zu} = \int_0^1 \dot{Z}^2(r)\mathrm{d}[B_{H,u}](r)$. Should u_t be weakly stationary, $B_{H,u} \equiv \sigma_u W_u(r)$ with $W_u(r)$ a standard Wiener process, leading to

$$\int_0^1 \dot{Z}^2(r)\mathrm{d}[B_{H,u}](r) = \sigma_u^2 \int_0^1 \dot{Z}^2(r)\mathrm{d}r.$$

The CMT implies $T^{-1-2\delta}\sum_{t=2}^{T} z_{t-1}^2 \Rightarrow \int_0^1 \dot{Z}^2(r)\mathrm{d}r$ and the usual standard errors are equivalent to the Eicker-White standard errors asymptotically.

Proof of Corollary 2

Each of the instruments satisfies Assumption 3; see the proof of Corollary 1. Theorem 2 then applies with R_{zx}^c derived from the $Z(r)$ and V_z implied by the chosen instrument.

Proof of Corollary 3

We show that each of the instruments satisfy Assumption 4.

1. Random walk instrument Weak convergence of the partial sums of w_t, jointly with the one of the partial sums of u_t and v_t, follows from Phillips and Durlauf (1986) under our assumptions. By exploiting the independence of w_t of u_t and v_t, it is straightforward to establish the desired mixed normality as required for the result.

2. Linear trend With $z_t = t$ we have that

$$\frac{1}{T^2}\sum_{t=2}^{[rT]} z_{t-1} \Rightarrow \int_0^r s\mathrm{d}s = \frac{r^2}{2} = Z(r)$$

and

$$\frac{1}{T^3}\sum z_{t-1}^2 \to \frac{1}{3},$$

so $V_z = 1/3$. Then,

$$\frac{1}{T^{1.5}}\sum z_{t-1}u_t \Rightarrow \sigma_u \int_0^1 r\mathrm{d}W_u(r)$$

where the integral is Gaussian with variance $1/3$, leading to a $\mathcal{N}\left(0, \sigma_u^2 V_z\right)$ limiting distribution.

The CMT provides

$$\frac{1}{T}\sum z_{t-1}x_{t-1} \Rightarrow \sigma_v \int_0^1 rJ_c(r)\,\mathrm{d}r$$

so

$$R_{zx}^c = \sigma_v \int_0^1 rJ_c(r)\,\mathrm{d}r$$

as required.

3. Cauchy instrument Using Christopeit's (2009) results, we have

$$\operatorname{sign}(x_{t-1}) \Rightarrow \operatorname{sign}(J_c(r)).$$

It follows that (Hansen, 1992)

$$\frac{1}{T^{1/2}}\sum_{t=2}^{T}\operatorname{sign}(x_{t-1})\,u_t \Rightarrow \int_0^1 \operatorname{sign}(J_c(r))\,\mathrm{d}W_u(r).$$

Moreover, we obviously have

$$\frac{1}{T}\sum_{t=1}^{T} z_{t-1}^2 = 1,$$

such that $V_z = 1$. The md sequence $\operatorname{sign}(x_{t-1})\,u_t$ has variance σ_u^2 and finite kurtosis. From White (2001, Corollary 5.26) it follows

$$\frac{1}{T^{1/2}}\sum \operatorname{sign}(x_{t-1})\,u_t \xrightarrow{d} \mathcal{N}\left(0, \sigma_u^2\right) \sim \mathcal{N}\left(0, \sigma_u^2 V_z\right)$$

as required. Finally, the CMT implies that

$$\frac{1}{T^{1.5}}\sum z_{t-1}x_{t-1} \xrightarrow{d} \sigma_v \int_0^1 |J_c(r)|\,\mathrm{d}r.$$

Thus $R_{zx}^c = \sigma_v \int_0^1 |J_c(r)|\,\mathrm{d}r$.

Details for Example 1

Let the vector of instruments be $\mathbf{z}_t = \left(\sin\pi\frac{t}{2T}\ , sin3\pi\frac{t}{2T}\right)'$. The (2SLS) t-statistic is

$$t_{2S}^{w} = \frac{\sum_{t=2}^{T} x_{t-1}\mathbf{z}_{t-1}'\left(\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)^{-1}\sum_{t=2}^{T}\mathbf{z}_{t-1}y_t}{\sqrt{\sum_{t=2}^{T} x_{t-1}\mathbf{z}_{t-1}'\left(\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)^{-1}\left(\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\widehat{u}_t^2\right)\left(\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)^{-1}\sum_{t=2}^{T}\mathbf{z}_{t-1}x_t}}.$$

Due to the orthogonality of the instruments we have

$$\frac{1}{T}\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}' \to \frac{1}{2}I_2.$$

Furthermore, using

$$\frac{1}{T}\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'x_{t-1}u_t \Rightarrow \sigma_u\sigma_v\int_0^1\begin{pmatrix}\sin^2\left(\pi\frac{r}{2}\right) & \sin\left(\pi\frac{r}{2}\right)\sin\left(3\pi\frac{r}{2}\right)\\ \sin\left(\pi\frac{r}{2}\right)\sin\left(3\pi\frac{r}{2}\right) & \sin^2\left(3\pi\frac{r}{2}\right)\end{pmatrix}J_c\left(r\right)\mathrm{d}W_u(r)\,,$$

and

$$\frac{1}{T}\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'u_t^2 = \sigma_u^2\frac{1}{T}\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}' + \frac{1}{T}\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\left(u_t^2-\sigma_u^2\right)$$

where the second term on the r.h.s. vanishes since $\left(u_t^2-\sigma_u^2\right)$ is i.i.d. we obtain

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\widehat{u}_t^2 &= \frac{1}{T}\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'u_t^2 - 2\left(\widehat{\beta}_{ls}-\beta\right)\frac{1}{T}\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'x_{t-1}u_t + \left(\widehat{\beta}_{ls}-\beta\right)^2\frac{1}{T}\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'x_{t-1}^2\\ &\overset{p}{\to} \frac{\sigma_u^2}{2}I_2\end{aligned}$$

Hence, under the null,

$$t_{2S}^{w} = \frac{1}{\sqrt{2}}\frac{\frac{1}{T^{3/2}}\sum_{t=2}^{T}x_{t-1}\mathbf{z}_{t-1}'\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\mathbf{z}_{t-1}u_t}{\sigma_u\sqrt{\frac{1}{T^{3/2}}\sum_{t=2}^{T}x_{t-1}\mathbf{z}_{t-1}'\frac{1}{T^{1.5}}\sum_{t=2}^{T}\mathbf{z}_{t-1}x_{t-1}}} + o_p\left(1\right).$$

Using

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\mathbf{z}_{t-1}u_t &\overset{d}{\to} \sigma_u\begin{pmatrix}\int_0^1\sin\left(\pi\frac{r}{2}\right)\mathrm{d}W_u\left(r\right)\\ \int_0^1\sin\left(3\pi\frac{r}{2}\right)\mathrm{d}W_u\left(r\right)\end{pmatrix}\\ \frac{1}{T^{1.5}}\sum_{t=2}^{T}\mathbf{z}_{t-1}x_{t-1} &\overset{d}{\to} \sigma_v\begin{pmatrix}\int_0^1\sin\left(\pi\frac{r}{2}\right)J_c\left(r\right)\mathrm{d}r\\ \int_0^1\sin\left(3\pi\frac{r}{2}\right)J_c\left(r\right)\mathrm{d}r\end{pmatrix}\end{aligned}$$

the result follows with the CMT.

Proof of Theorem 3

Under the null hypothesis and Assumption 5 the 2SLS test statistic (12) can be written as

$$t_{2S}^{w} = \frac{M_{xz}D_T^{-1}(D_T^{-1}M_{zz}D_T^{-1})^{-1}\sqrt{T}\cdot D_T^{-1}M_{zu}}{\sqrt{M_{xz}D_T^{-1}(D_T^{-1}M_{zz}D_T^{-1})^{-1}(D_T^{-1}M_{zzuu}D_T^{-1})(D_T^{-1}M_{zz}D_T^{-1})^{-1}D_T^{-1}M_{zx}}}\,,$$

where $M_{xz} = T^{-1} \sum_{t=2}^{T} x_{t-1} \mathbf{z}'_{t-1}$, all other matrices $M_{\bullet\bullet}$ are defined in a similar manner, and $M_{zzuu} = T^{-1} \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{z}'_{t-1} \widehat{u}_t^2$. Assume that the instruments in the vector z_t are arranged with respect to ϑ_i such that z_{1t} denotes the instrument with the largest value $\vartheta_1 > 0$ with $\vartheta_1 > \vartheta_i$ for $i = 2, \ldots, m$. Since Assumption 5 implies Assumption 3 or 4 individually, it follows from Theorem 2 that

$$\frac{1}{T^{\vartheta_1+1}} D_T^{-1} M_{zx} \overset{p}{\to} \left(R_{zx,1}^c, 0, \ldots, 0\right)'.$$

Following the proof of Theorem 2, we obtain

$$D_T^{-1} \left(\frac{1}{T} \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{z}'_{t-1} \widehat{u}_t^2 \right) D_T^{-1} = D_T^{-1} \left(\frac{1}{T} \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{z}'_{t-1} u_t^2 \right) D_T^{-1} + o_p(1).$$

Assumption 5 then leads to

$$t_{2S}^w \overset{d}{\to} \operatorname{sign}\left(R_{zx,1}^c\right) \frac{[V_{\mathbf{z}}^{-1}]_{1,\bullet}\, \mathcal{MN}(0, V_{\mathbf{z}u})}{\sqrt{[V_{\mathbf{z}}^{-1}]_{1,\bullet} V_{\mathbf{z}u} \, [V_{\mathbf{z}}^{-1}]'_{1,\bullet}}},$$

where $[V_{\mathbf{z}}^{-1}]_{1\bullet}$ denotes the first row of the matrix $V_{\mathbf{z}}^{-1}$. Hence $[V_{\mathbf{z}}^{-1}]_{1,\bullet}\, \mathcal{MN}(0, V_{\mathbf{z}u})$ is mixed Gaussian too as stated by Assumption 5, and the ratio $\frac{[V_{\mathbf{z}}^{-1}]_{1,\bullet}\, \mathcal{N}(0, V_{\mathbf{z}u})}{\sqrt{[V_{\mathbf{z}}^{-1}]_{1,\bullet} V_{\mathbf{z}u} [V_{\mathbf{z}}^{-1}]'_{1,\bullet}}}$ is standard normal. Since $\operatorname{sign}\left(R_{zx,1}^c\right)$ may be random and dependent of the ratio (e.g. for type-II instruments), t_{2S}^w itself is not always standard normal. But it follows immediately that its square $(t_{2S}^w)^2$ is chi-squared distributed with one degree of freedom as required.

Proof of Theorem 4

The result follows with the property of quadratic forms of m-dimensional multivariate normal distributions $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ that $\mathbf{x}'\Sigma^{-1}\mathbf{x} \sim \chi^2\left(m, \boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}\right)$ by using Assumption 5 (iii),

$$\frac{1}{\sqrt{T}} D_T \sum z_{t-1} u_t \overset{d}{\to} \mathcal{N}\left(0; V_{\mathbf{z}u}\right),$$

together with the convergence to $R_{zx,i}^c$ for each instrument as in Theorem 2, and by noting that

$$\frac{1}{T^{\max_j \vartheta_j + \delta_i + 1}} \sum z_{i,t-1} x_{t-1} \overset{d}{\to} R_{zx,i}^c$$

only if $\max_j \vartheta_j = \vartheta_i$ (the latter limit being zero otherwise).

Proof of Theorem 6

The estimators $\widehat{\boldsymbol{\tau}}$ and $\widehat{\boldsymbol{\beta}}_{va}$ are given by

$$\begin{pmatrix} \widehat{\boldsymbol{\tau}} \\ \widehat{\boldsymbol{\beta}}_{va} \end{pmatrix} = M_T^{-1} P_T,$$

where the FWL theorem indicates that

$$P_T = \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t u_t \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} u_t \end{pmatrix} - \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \boldsymbol{\zeta}'_{t-1} \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \boldsymbol{\zeta}'_{t-1} \end{pmatrix} \left(\sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \boldsymbol{\zeta}'_{t-1} \right)^{-1} \sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} u_t$$

and

$$M_T = \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \mathbf{f}_t' & \sum_{t=2}^{T} \mathbf{f}_t \mathbf{z}_{t-1}' \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{f}_t' & \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{z}_{t-1}' \end{pmatrix} - \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \boldsymbol{\zeta}_{t-1}' \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \boldsymbol{\zeta}_{t-1}' \end{pmatrix} \left(\sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \boldsymbol{\zeta}_{t-1}' \right)^{-1} \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \boldsymbol{\zeta}_{t-1}' \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \boldsymbol{\zeta}_{t-1}' \end{pmatrix}'.$$

The HC covariance matrix estimator is given by

$$M_T^{-1} Q_T M_T^{-1},$$

where

$$\begin{aligned} Q_T &= \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \mathbf{f}_t' \widehat{u}_t^2 & \sum_{t=2}^{T} \mathbf{f}_t \mathbf{z}_{t-1}' \widehat{u}_t^2 \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{f}_t' \widehat{u}_t^2 & \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{z}_{t-1}' \widehat{u}_t^2 \end{pmatrix} - \\ &\quad - \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \boldsymbol{\zeta}_{t-1}' \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \boldsymbol{\zeta}_{t-1}' \end{pmatrix} \left(\sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \boldsymbol{\zeta}_{t-1}' \right)^{-1} \begin{pmatrix} \sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \mathbf{f}_t' \widehat{u}_t^2 & \sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \mathbf{z}_{t-1}' \widehat{u}_t^2 \end{pmatrix} \\ &\quad - \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \boldsymbol{\zeta}_{t-1}' \widehat{u}_t^2 \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \boldsymbol{\zeta}_{t-1}' \widehat{u}_t^2 \end{pmatrix} \left(\sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \boldsymbol{\zeta}_{t-1}' \right)^{-1} \begin{pmatrix} \sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \mathbf{f}_t' & \sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \mathbf{z}_{t-1}' \end{pmatrix} \\ &\quad + \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \boldsymbol{\zeta}_{t-1}' \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \boldsymbol{\zeta}_{t-1}' \end{pmatrix} \left(\sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \boldsymbol{\zeta}_{t-1}' \right)^{-1} \sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \boldsymbol{\zeta}_{t-1}' \widehat{u}_t^2 \left(\sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \boldsymbol{\zeta}_{t-1}' \right)^{-1} \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \boldsymbol{\zeta}_{t-1}' \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \boldsymbol{\zeta}_{t-1}' \end{pmatrix}'. \end{aligned}$$

The steps of the proof are quite similar to that of Theorem 1. Specifically, by using similar arguments we obtain

$$\left(\sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \boldsymbol{\zeta}_{t-1}' \right)^{-1} = O_p\left(T^{-2}\right)$$

and

$$\sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} u_t = O_p\left(T\right).$$

Moreover,

$$\frac{1}{T^{1.5}} D_T^{-1} \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \boldsymbol{\zeta}_{t-1}' \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \boldsymbol{\zeta}_{t-1}' \end{pmatrix} = \frac{1}{T^{1.5}} D_T^{-1} \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \mathbf{x}_{t-1}' \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{x}_{t-1}' \end{pmatrix} - \frac{1}{T^{1.5}} D_T^{-1} \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \mathbf{z}_{t-1}' \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{z}_{t-1}' \end{pmatrix} = o_p(1)$$

since, under the null, $y_t = u_t$ and

$$\frac{1}{\sqrt{T}} D_T^{-1} P_T = \frac{1}{\sqrt{T}} D_T^{-1} \sum_{t=2}^{T} \begin{pmatrix} \mathbf{f}_t \\ \mathbf{z}_{t-1} \end{pmatrix} u_t + o_p(1).$$

Using again the arguments of the proof of Theorem 1, we obtain

$$\sum_{t=2}^{T} \boldsymbol{\zeta}_{t-1} \boldsymbol{\zeta}_{t-1}' \widehat{u}_t^2 = O_p\left(T^2\right),$$

$$\frac{1}{T} D_T^{-1} \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \mathbf{f}_t' \widehat{u}_t^2 & \sum_{t=2}^{T} \mathbf{f}_t \mathbf{z}_{t-1}' \widehat{u}_t^2 \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{f}_t' \widehat{u}_t^2 & \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{z}_{t-1}' \widehat{u}_t^2 \end{pmatrix} D_T^{-1} = \frac{1}{T} D_T^{-1} \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \mathbf{f}_t' u_t^2 & \sum_{t=2}^{T} \mathbf{f}_t \mathbf{z}_{t-1}' u_t^2 \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{f}_t' u_t^2 & \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{z}_{t-1}' u_t^2 \end{pmatrix} D_T^{-1} + o_p(1)$$

and

$$\frac{1}{T^{1.5}} D_T^{-1} \begin{pmatrix} \sum_{t=2}^{T} \mathbf{f}_t \boldsymbol{\zeta}_{t-1}' \widehat{u}_t^2 \\ \sum_{t=2}^{T} \mathbf{z}_{t-1} \boldsymbol{\zeta}_{t-1}' \widehat{u}_t^2 \end{pmatrix} = O_p\left(1\right)$$

leading to

$$\frac{1}{T}D_T^{-1}Q_TD_T^{-1} = \frac{1}{T}D_T^{-1}\left(\begin{array}{cc} \sum_{t=2}^T \mathbf{f}_t\mathbf{f}_t' u_t^2 & \sum_{t=2}^T \mathbf{f}_t\mathbf{z}_{t-1}' u_t^2 \\ \sum_{t=2}^T \mathbf{z}_{t-1}\mathbf{f}_t' u_t^2 & \sum_{t=2}^T \mathbf{z}_{t-1}\mathbf{z}_{t-1}' u_t^2 \end{array}\right) D_T^{-1} + o_p(1)$$

taking us to the desired limiting distribution of $\mathcal{T}_{va,K}$.

Proof of Theorem 7

Denote the diagonal block of D_T corresponding to $\mathbf{f}_t$ by $D_{\mathbf{f}T}$, and the diagonal block of D_T corresponding to $\mathbf{z}_{t-1}$ by $D_{\mathbf{z}T}$. Then, standard results for IV regression yield

$$\widehat{\boldsymbol{\beta}}_{iv} = M_T^{-1}P_T$$

with

$$\begin{aligned} \frac{1}{\sqrt{T}}D_{\mathbf{z}T}^{-1}P_T &= \frac{1}{\sqrt{T}}D_{\mathbf{z}T}^{-1}\sum_{t=2}^T \mathbf{z}_{t-1}y_t - \frac{1}{T}D_{\mathbf{z}T}^{-1}\sum_{t=2}^T \mathbf{z}_{t-1}\mathbf{f}_t' D_{\mathbf{f}T}^{-1}\left(\frac{1}{T}D_{\mathbf{f}T}^{-1}\sum_{t=2}^T \mathbf{f}_t\mathbf{f}_t' D_{\mathbf{f}T}^{-1}\right)^{-1}\frac{1}{\sqrt{T}}D_{\mathbf{f}T}^{-1}\sum_{t=2}^T \mathbf{f}_t y_t\,, \\ M_T &= \sum_{t=2}^T \mathbf{z}_{t-1}\mathbf{x}_{t-1} - \sum_{t=2}^T \mathbf{z}_{t-1}\mathbf{f}_t'\left(\sum_{t=2}^T \mathbf{f}_t\mathbf{f}_t'\right)^{-1}\sum_{t=2}^T \mathbf{f}_t\mathbf{x}_{t-1} \end{aligned}$$

and

$$\widehat{V}\left(\widehat{\boldsymbol{\beta}}_{iv}\right) = M_T^{-1}Q_T\left(M_T^{-1}\right)',$$

where

$$\begin{aligned} \frac{1}{T}D_{\mathbf{z}T}^{-1}Q_TD_{\mathbf{z}T}^{-1} &= \frac{1}{T}D_{\mathbf{z}T}^{-1}\sum_{t=2}^T \mathbf{z}_{t-1}\mathbf{z}_{t-1}'\widehat{u}_t^2 D_{\mathbf{z}T}^{-1} \\ &\quad - \frac{1}{T}D_{\mathbf{z}T}^{-1}\sum_{t=2}^T \mathbf{z}_{t-1}\mathbf{f}_t' D_{\mathbf{f}T}^{-1}\left(\frac{1}{T}D_{\mathbf{f}T}^{-1}\sum_{t=2}^T \mathbf{f}_t\mathbf{f}_t' D_{\mathbf{f}T}^{-1}\right)^{-1}\frac{1}{T}D_{\mathbf{f}T}^{-1}\sum_{t=2}^T \mathbf{f}_t\mathbf{z}_{t-1}'\widehat{u}_t^2 D_{\mathbf{z}T}^{-1} \\ &\quad - \frac{1}{T}D_{\mathbf{z}T}^{-1}\sum_{t=2}^T \mathbf{z}_{t-1}\mathbf{f}_t'\widehat{u}_t^2 D_{\mathbf{f}T}^{-1}\left(\frac{1}{T}D_{\mathbf{f}T}^{-1}\sum_{t=2}^T \mathbf{f}_t\mathbf{f}_t' D_{\mathbf{f}T}^{-1}\right)^{-1}\frac{1}{T}D_{\mathbf{f}T}^{-1}\sum_{t=2}^T \mathbf{f}_t\mathbf{z}_{t-1}' D_{\mathbf{z}T}^{-1} \\ &\quad + \frac{1}{T}D_{\mathbf{z}T}^{-1}\sum_{t=2}^T \mathbf{z}_{t-1}\mathbf{f}_t' D_{\mathbf{f}T}^{-1}\left(\frac{1}{T}D_{\mathbf{f}T}^{-1}\sum_{t=2}^T \mathbf{f}_t\mathbf{f}_t' D_{\mathbf{f}T}^{-1}\right)^{-1}\frac{1}{T}D_{\mathbf{f}T}^{-1}\sum_{t=2}^T \mathbf{f}_t\mathbf{f}_{t-1}'\widehat{u}_t^2 D_{\mathbf{f}T}^{-1}\times \\ &\qquad \times\left(\frac{1}{T}D_{\mathbf{f}T}^{-1}\sum_{t=2}^T \mathbf{f}_t\mathbf{f}_t' D_{\mathbf{f}T}^{-1}\right)^{-1}\frac{1}{T}D_{\mathbf{f}T}^{-1}\sum_{t=2}^T \mathbf{f}_t\mathbf{z}_{t-1}' D_{\mathbf{z}T}^{-1}. \end{aligned}$$

Consequently,

$$\mathcal{T}_{iv,K}^{w} = \frac{1}{\sqrt{T}}P_T' D_{\mathbf{z}T}^{-1}\left(\frac{1}{T}D_{\mathbf{z}T}^{-1}Q_TD_{\mathbf{z}T}^{-1}\right)^{-1}\frac{1}{\sqrt{T}}D_{\mathbf{z}T}^{-1}P_T.$$

Since under the null $y_t = u_t$,

$$\frac{1}{\sqrt{T}}\sum_{t=2}^T\left(\begin{array}{c} D_{\mathbf{f}T}^{-1}\mathbf{f}_t y_t \\ D_{\mathbf{z}T}^{-1}\mathbf{z}_{t-1}y_t \end{array}\right) \Rightarrow \mathcal{MN}\left(0, V_{\mathbf{fz}u}\right),$$

it follows by Assumption 5 and the fact that $\mathbf{f}_t$ is deterministic that

$$\frac{1}{\sqrt{T}} P_T \Rightarrow \mathcal{MN}\left(0, R' V_{\mathbf{fz}u} R\right)$$

with $R' = \left(-V_{\mathbf{zf}} V_{\mathbf{ff}}^{-1}, I_K\right)$, where $V_{\mathbf{zf}}$ is the (weak) limit of $\frac{1}{T} D_{\mathbf{z}T}^{-1} \sum_{t=2}^{T} \mathbf{z}_{t-1} \mathbf{f}_t' D_{\mathbf{f}T}^{-1}$ and $V_{\mathbf{ff}}$ that of $\frac{1}{T} D_{\mathbf{f}T}^{-1} \sum_{t=2}^{T} \mathbf{f}_t \mathbf{f}_t' D_{\mathbf{f}T}^{-1}$ and is thus deterministic. But this is exactly the (weak) limit of $\frac{1}{T} D_{\mathbf{z}T}^{-1} Q_T D_{\mathbf{z}T}^{-1}$, as can be easily checked using the arguments employed in the proof of Theorems 1 and 2, and the result follows.

Note that the result only holds when M_T cancels out, which is only the case when K instruments are used.

References

Bauer, D. and A. Maynard (2012). Persistence-Robust Surplus-Lag Granger Causality Testing. *Journal of Econometrics 169*(2), 293–300.

Campbell, J. Y. and R. J. Shiller (1988). Stock prices, earnings, and expected dividends. *The Journal of Finance 43*(3), 661–676.

Campbell, J. Y. and M. Yogo (2006). Efficient tests of stock return predictability. *Journal of Financial Economics 81*(1), 27–60.

Cavaliere, G. (2004). Unit root tests under time-varying variances. *Econometric Reviews 23*(3), 259–292.

Cavaliere, G., A. Rahbek, and A. M. R. Taylor (2010). Testing for Co-Integration in Vector Autoregressions with Non-Stationary Volatility. *Journal of Econometrics 158*(1), 7–24.

Cavanagh, C. L., G. Elliott, and J. H. Stock (1995). Inference in Models with Nearly Integrated Regressors. *Econometric Theory 11*(5), 1131–1147.

Chen, W. W. and R. S. Deo (2009). Bias reduction and likelihood based almost-exactly sized hypothesis testing in predictive regressions using the restricted likelihood. *Econometric Theory 25*(5), 1143<–1179.

Choi, I. (1993). Asymptotic Normality of the Least-Squares Estimates for Higher Order Autoregressive Integrated Processes with Some Applications. *Econometric Theory 9*(2), 263–282.

Christopeit, N. (2009). Weak Convergence of Nonlinear Transformations of Integrated Processes: The Multivariate Case. *Econometric Theory 25*(5), 1180–1207.

Clark, T. E. (2009). Is the Great Moderation over? An empirical analysis. *Economic Review 4*, 5–42.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford university press.

Dolado, J. J., J. Gonzalo, and L. Mayoral (2002). A Fractional Dickey–Fuller Test for Unit Roots. *Econometrica 70*(5), 1963–2006.

Dolado, J. J. and H. Lütkepohl (1996). Making Wald Tests Work for Cointegrated VAR Systems. *Econometric Reviews 15*(4), 369–386.

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. In L. Le Cam and J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 59–82. Berkeley: University of California Press.

Elliott, G. and J. H. Stock (1994). Inference in Time Series Regression when the Order of Integration of a Regressor Is Unknown. *Econometric Theory 10*(3-4), 672–700.

Froot, K. and R. Thaler (1990). Anomalies: Foreign exchange. *Journal of Economic Perspectives 4*(1), 179–192.

Gorodnichenko, Y., A. Mikusheva, and S. Ng (2012). Estimators for persistent and possibly non-stationary data with classical properties. *Econometric Theory 28*(5), 1003–1036.

Hansen, B. (1992). Convergence to Stochastic Integrals for Dependent Heterogeneous Processes. *Econometric Theory 8*(4), 489–500.

Jansson, M. and M. J. Moreira (2006). Optimal Inference in Regression Models with Nearly Integrated Regressors. *Econometrica 74*(3), 681–714.

Kleptsyna, M. L. and A. Le Breton (2002). Statistical Analysis of the Fractional Ornstein–Uhlenbeck Type Process. *Statistical Inference for stochastic processes 5*(3), 229–248.

Lettau, M. and S. Van Nieuwerburgh (2008). Reconciling the return predictability evidence. *Review of Financial Studies 21*(4), 1607–1652.

Magdalinos, T. and P. C. B. Phillips (2009). Limit Theory for Cointegrated Systems with Moderately Integrated and Moderately Explosive Regressors. *Econometric Theory 25*(2), 482–526.

Marinucci, D. and P. M. Robinson (2000). Weak Convergence of Multivariate Fractional Processes. *Stochastic Processes and their Applications 86*(1), 103–120.

Maynard, A. and K. Shimotsu (2009). Covariance-based orthogonality tests for regressors with unknown persistence. *Econometric Theory 25*(1), 63–116.

Park, J. Y. (1990). Testing for unit roots and cointegration by variable addition. In T. B. Fomby and G. F. J. Rhodes (Eds.), *Advances in Econometrics, Volume 8*. JAI press Greenwich.

Phillips, P. C. B. (1987). Towards a Unified Asymptotic Theory for Autoregression. *Biometrika 74*(3), 535–547.

Phillips, P. C. B. (1998). New Tools for Understanding Spurious Regressions. *Econometrica 66*(6), 1299–1325.

Phillips, P. C. B. and S. N. Durlauf (1986). Multiple Time Series Regression with Integrated Processes. *The Review of Economic Studies 53*(4), 473–495.

Phillips, P. C. B. and T. Magdalinos (2007). Limit Theory for Moderate Deviations From a Unit Root. *Journal of Econometrics 136*(1), 115–130.

Phillips, P. C. B. and T. Magdalinos (2009). Econometric Inference in the Vicinity of Unity. Working paper, Singapore Management University.

Phillips, P. C. B. and V. Solo (1992). Asymptotics for Linear Processes. *The Annals of Statistics 20*(2), 971–1001.

Sensier, M. and D. van Dijk (2004). Testing for volatility changes in U.S. macroeconomic time series. *86*(3), 833–839.

Sims, C. A., J. H. Stock, and M. W. Watson (1990). Inference in Linear Time Series Models with Some Unit Roots. *Econometrica 58*(1), 113–144.

So, B. S. and D. W. Shin (1999). Cauchy Estimators for Autoregressive Processes with Applications to Unit Root Tests and Confidence Intervals. *Econometric Theory 15*(2), 165–176.

Stock, J. H. and M. W. Watson (2002). Has the business cycle changed and why? *NBER Macroeconomics Annual 17*(1), 159–218.

Toda, H. Y. and T. Yamamoto (1995). Statistical Inference in Vector Autoregressions with Possibly Integrated Processes. *Journal of Econometrics 66*(1-2), 225–250.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

White, H. (2001). *Asymptotic theory for econometricians.* Academic Press New York.