Robust Econometric Inference with Mixed Integrated and Mildly Explosive Regressors

Peter C.B. Phillips * Ji Hyung Lee†

October, 2014

Abstract

This paper considers a convenient inference procedure for nonstationary variable regression that enables robust chi-square testing for a wide class of persistent and endogenous regressors. The approach uses the mechanism of self generated instruments called IVX instrumentation developed by Magdalinos and Phillips (2009b). We first show that these methods remain valid for regressors with locally and mildly explosive roots. It is further shown that Wald testing procedures remain robust for multivariate regressors with mixed degrees of persistence. These robustifications are useful in econometric inference, for example, when there are periods of mildly explosive trends in some or all of time series employed in the analysis but the exact knowledge on the regressor persistence is unavailable. Practical issues related to the choice of the IVX instruments are also addressed. The methods are straightforward to apply in practical work such as predictive regression applications in finance.

Keywords: Chi-square, Instrumentation, IVX methods, Local to unity, Mild integration, Mild explosiveness, Predictive regression, Robustness.

JEL classification: C22

1 Introduction

Many economic and financial time series exhibit characteristics that include temporary periods of explosive behavior. For macroeconomic series Stock (1991, Table 2) showed that 90% confidence intervals for the autoregressive (AR) roots of the Nelson-Plosser data set contain explosive parameter regions in all but one series (the unemployment rate). For financial series Campbell and Yogo (2006, Table 4) found that 95% confidence intervals for the AR coefficient of the S&P 500 dividend-price ratio and other series over long historical periods do not rule out explosive roots. In addition to these empirical findings, periodically occurring booms and episodes of financial exuberance support at least temporary explosive trends in economic and financial data. The idea that

*Phillips acknowledges support under NSF Grant No. SES 12-58258. Yale University, University of Auckland, Singapore Management University & University of Southampton. Email: peter.phillips@yale.edu

†University of Washington. Email: jihyung2@uw.edu

there are subperiods of explosive roots in economic and financial series is formally analyzed and empirically confirmed in Phillips et al. (2011, 2015) and Phillips and Yu (2011).

There has been growing interest in predictive regressions of the type used in the study by Campbell and Yogo. When the regressors in such models display some degree of persistence, inference procedures need to be robust to the value of the persistence parameter to ensure validity even asymptotically. The robustness requirements become more demanding in cases where there are many regressors with possibly different degrees of persistence. To address some of these complexities in inference in regressions with persistence regressors, Magdalinos and Phillips (2009b, hereafter MP) recently introduced a novel IV procedure (called IVX regression) that provided robust chi-square inference in a much wider vicinity of unity than existing studies that have typically only considered the near integrated (local to unity) single regressor case. In particular, MP showed that self-normalized IVX test statistics have an asymptotically pivotal chi-square distribution for multivariate regressors that may be integrated, near integrated, or mildly integrated and which thereby fall within a vector autoregressive framework (Lutkepohl, 2005) while allowing for more general time series inputs than martingale differences. The tests have been successfully used in applied work on predictive regressions (Kostakis et al., 2014; Gonzalo and Pitarakis, 2012). IVX methods have been also studied in long-horizon regression applications (Phillips and Lee, 2013) and in quantile regression (Lee, 2014) contexts.

The present paper extends the IVX methodology to include a wider range of potential regressors that includes locally explosive and mildly explosive roots, thereby covering periods of exuberance in economic and financial data. The limit theory involves some novel developments in the mildly explosive case, where the latent IVX instrument which depends on the true values of the localizing coefficients may no longer dominate the asymptotics. The chi-square limit theory of the same self-normalized test statistics of MP is shown to continue to be valid in this wider setting. We also confirm that the limit theory is robust under mixed degrees of persistence, allowing for the simultaneous presence of local to unity (or mildly integrated) roots and mildly explosive roots. As a result of these extensions, IVX regression provides a framework for unified test procedures covering a large class of persistent regressors whose individual characteristics may differ from each other. Empirical researchers may therefore use this framework without having to be specific about the particular properties of individual regressors.

The paper is organized as follows. Section 2 develops the limit theory for IVX regression under locally and mildly explosive roots and demonstrates robustness. Section 3 extends these robustness results to cases where the regressors have mixed degrees of persistence or explosive behavior. Section 4 discusses issues associated with the choice of the IVX tuning parameter and provides simulation results. Selected technical derivations, supporting lemmas, and proofs of the main results in the paper are contained in the Appendix.

2 IVX Regression with Explosive Roots

2.1 Framework

We follow the framework used in MP for the following system:

$$
\begin{aligned}
y_t &= Ax_t + u_{0t}, \qquad (2.1)\\
x_t &= R_n x_{t-1} + u_{xt},\\
R_n &= I_K + \frac{C}{n^{\alpha}}, \text{ for some } \alpha > 0,
\end{aligned}
$$

where A is an $m \times K$ coefficient matrix and $C = diag\,(c_1, c_2, ..., c_K)$ represents the localizing coefficients in the multivariate regressors.

The IVX approach of MP allowed the regressors x_t to be (I1) integrated ($C = 0$), (I2) near integrated ($C < 0, \alpha = 1$) and (I3) mildly integrated ($C < 0, \alpha \in (0, 1)$), and developed an inference procedure that is robust to the precise degree of integration. We will show these results are robust under the same framework but with (I4) locally explosive ($C > 0, \alpha = 1$) and (I5) mildly explosive roots ($C > 0, \alpha \in (0, 1)$), as well as possibly mixed versions (I6) of these roots.

For the structure of innovations, we follow the linear process set up of MP:

$$
u_t := \begin{bmatrix} u_{0t} \\ u_{xt} \end{bmatrix} = \sum_{j=0}^{\infty} F_j \varepsilon_{t-j},\ \varepsilon_t \sim iid\,(0, \Sigma)\,,\ \Sigma > 0,\ E\,\|\varepsilon_1\|^4 < \infty, \qquad (2.2)
$$

$$
F_0 = I_{m+K},\ \sum_{j=0}^{\infty} j\,\|F_j\| < \infty,\ F(z) = \sum_{j=0}^{\infty} F_j z^j \text{ and } F(1) = \sum_{j=0}^{\infty} F_j > 0.
$$

In the above, we use the spectral norm $\|M\| = \max_i \left\{\lambda_i^{1/2} : \lambda_i = \text{an eigenvalue of } M'M\right\}$. Other norms, such as the L_1 and L_2 norms, are specified in what follows as needed using the notation $\|\cdot\|_{L_i}$ $(i = 1, 2)$.

Under these conditions there exists a Beveridge-Nelson (BN) decomposition and the following component-wise expressions (Phillips and Solo, 1992)

$$
u_t = F(1)\varepsilon_t - \triangle\tilde{\varepsilon}_t,\ \tilde{\varepsilon}_t = \sum_{j=0}^{\infty} \tilde{F}_j \varepsilon_{t-j},\ \tilde{F}_j = \sum_{s=j+1}^{\infty} F_s, \qquad (2.3)
$$

$$
u_{0t} = F_0(1)\varepsilon_t - \triangle\tilde{\varepsilon}_{0t},\ u_{xt} = F_x(1)\varepsilon_t - \triangle\tilde{\varepsilon}_{xt}. \qquad (2.4)
$$

The long run covariance matrices associated with u_t are denoted as

$$
\Omega = \sum_{h=-\infty}^{\infty} E\left(u_t u_{t-h}'\right) = F(1)\Sigma F(1)',\ \Delta = \sum_{h=0}^{\infty} E\left(u_t u_{t-h}'\right), \qquad (2.5)
$$

$$
\Omega = \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix},\ \Delta = \begin{bmatrix} \Delta_{00} & \Delta_{0x} \\ \Delta_{x0} & \Delta_{xx} \end{bmatrix}.
$$

Under (2.2) we have the functional law (Phillips and Solo, 1992):

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns \rfloor} u_j := \frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns \rfloor}\begin{bmatrix} u_{0t} \\ u_{xt} \end{bmatrix} =: \begin{bmatrix} B_{0n}(s) \\ B_{xn}(s) \end{bmatrix} \Longrightarrow \begin{bmatrix} B_0(s) \\ B_x(s) \end{bmatrix} = BM\begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix}, \tag{2.6}$$

where $B = (B_0', B_x')'$ is vector Brownian motion (BM), and the following local to unity limit law for cases (I2)-(I4) also holds (Phillips, 1987):

$$\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \Longrightarrow J_x^c(r), \text{ where } J_x^c(r) = \int_0^r e^{(r-s)C} dB_x(s). \tag{2.7}$$

2.2 A Review of IVX Construction

MP showed that the limit theory of the IVX estimator of A is mixed normal and that suitably self-normalized test statistics have an asymptotic chi-square distribution. Since the limit theory is pivotal and therefore free of the nuisance parameter C, no information on the degree of regressor persistence is needed to execute tests as long as the regressors fall into the categories (I1), (I2) or (I3). The key step in IVX is the construction of an instrument using only the regressors $\{x_t\}$ - hence the terminology IVX:

$$\tilde{z}_t = \sum_{j=1}^{t} R_{nz}^{t-j} \triangle x_j, \quad \text{where } R_{nz} = I_K + \frac{C_z}{n^\beta}, \beta \in (0,1), C_z < 0. \tag{2.8}$$

Since $\triangle x_j = \frac{C}{n^\alpha} x_{j-1} + u_{xj}$, we have the decomposition $\tilde{z}_t = \sum_{j=1}^t R_{nz}^{t-j} u_{xj} + \frac{C}{n^\alpha}\sum_{j=1}^t R_{nz}^{t-j} x_{j-1}$ denoted as

$$\tilde{z}_t = z_t + \frac{C}{n^\alpha}\psi_{nt}. \tag{2.9}$$

Using conventional observation matrix notation, the bias corrected IVX estimator of A suggested by MP and its estimation error have the form

$$\tilde{A}_n = (Y'\tilde{Z} - n\widehat{\Delta}_{0x})(X'\tilde{Z})^{-1}, \tag{2.10}$$

$$\tilde{A}_n - A = (U_0'\tilde{Z} - n\widehat{\Delta}_{0x})(X'\tilde{Z})^{-1}, \tag{2.11}$$

where $\widehat{\Delta}_{0x}$ is some consistent estimate of Δ_{0x}. The estimator $\tilde{A}_n$ is a simple adjusted version of the conventional IV estimator $\hat{A} = (Y'\tilde{Z})(X'\tilde{Z})^{-1}$ using instruments $\tilde{z}_t$. In view of the decomposition in (2.9), z_t plays the role of a latent mildly integrated instrument and the remainder $\frac{C}{n^\alpha}\psi_{nt}$ is eliminated asymptotically due to its scaling coefficient $\frac{C}{n^\alpha}$. As a result we have nuisance parameter (C) free inference using $\tilde{A}_n$. Martingale limit theory applies to the numerator matrix $U_0'\tilde{Z} - n\widehat{\Delta}_{0x}$ in (2.11), and this leads to a mixed normal limit theory that is well suited to inference.

2.3 Limit Theory for Regressors with Locally Explosive Roots

It turns out that IVX regression limit theory may be extended to explosive cases. The following result holds for the IVX estimator (2.10) with (I4) locally explosive regressors. The limit theory under this (I4) case remains exactly the same as for the near integrated case (I2) or case N(ii) of MP.

Theorem 2.1 (Locally Explosive Regressor) *With* $\beta \in (1/2, 1)$,

$$vec\left[n^{\frac{1+\beta}{2}}\left(\tilde{A}_n - A\right)\right] \Longrightarrow MN\left(0, \left(\tilde{\Psi}_{xx}^{-1}\right)' C_z V_{zz}^x C_z \tilde{\Psi}_{xx}^{-1} \otimes \Omega_{00}\right),$$

where

$$\tilde{\Psi}_{xx} = \Omega_{xx} + \int_0^1 J_x^c(r) dJ_x^c(r)' \text{ and } V_{zz}^x = \int_0^\infty e^{pC_z}\Omega_{xx}e^{pC_z}dp.$$

2.4 Limit Theory for Regressors with Mildly Explosive Roots

The asymptotics in the (I5) case turn out to be more complicated because the remainder term $\frac{C}{n^\alpha}\psi_{nt}$ now dominates the latent instrument z_t in (2.9) and this has a substantial effect on the derivations, as discussed below. Moreover, it is not necessary to include the bias adjustment in this case and the limit theory for $\tilde{A}_n$ is identical to that of the unadjusted $\hat{A}$. Interestingly, the limit theory for $\hat{A}$ then coincides with that of the OLS estimator (see Theorem 4.1 in Magdalinos and Phillips (2009a)).

We consider case (I5) with mildly explosive regressors ($C > 0, \alpha \in (0,1)$) under the moment condition $E\|\varepsilon_1\|^q < \infty$, $q \geq 4$. The following lemmas provide limit theory for standardized versions of the regressor x_t that are needed for the asymptotic development. A rate condition on α ($\alpha > \frac{2}{q}$) is imposed in the second lemma to ensure a uniform strong approximation, ensuring that a wider range of α are admissible when higher moment conditions apply (see the discussion in the Proof of Lemma 2.2).

We start with the following lemma from Magdalinos and Phillips (2009a, lemma 4.1):

Lemma 2.1 *Define* $Y_{Cn} := \frac{1}{n^{\alpha/2}}\sum_{j=1}^{k_n} R_n^{-j}F_x(1)\varepsilon_j$ *for* $k_n \to \infty$ *such that* $\|R_n\|^{-k_n} \to 0$. *Then*

$$n^{-\alpha/2}R_n^{-n}x_n = \frac{1}{n^{\alpha/2}}\sum_{j=1}^{n} R_n^{-j}u_{xj} = \frac{1}{n^{\alpha/2}}\sum_{j=1}^{k_n} R_n^{-j}u_{xj} + o_p(1) =: Y_{Cn} + o_p(1),$$

and

$$Y_{Cn} \Longrightarrow Y_C \equiv N\left(0, \int_0^\infty e^{-pC}\Omega_{xx}e^{-pC}dp\right).$$

The requirement $\|R_n\|^{-k_n} \to 0$ in lemma 2.1 implies that $\frac{k_n}{n^\alpha} \to \infty$. As in MP (2009a), the quantities Y_{Cn} and Y_C play an important role in the asymptotic theory. In addition, the following uniform approximation holds, which helps to simplify proofs. A similar approximation was shown for the scalar case in Phillips and Magdalinos (2007b).

Lemma 2.2 (Uniform Approximation) *With the same k_n defined in lemma 2.1 and for a suitably expanded probability space*

$$\sup_{k_n \leq j-1 \leq n} \left\| \frac{1}{n^{\alpha/2}} R_n^{-(j-1)} x_{j-1} - \tilde{Y}_C \right\| = o_{a.s}(1),$$

where $\tilde{Y}_C$ is a distributionally equivalent copy of Y_C on the common probability space, so that $\tilde{Y}_C =^d Y_C$.

Remark 2.1 *The above uniformly strong approximation of the normalized process of a vector of mildly explosive regressors is particularly useful in developing limit theory in mildly explosive regression. But for the purpose of our proof here convergence in probability is enough and for that result somewhat weaker moment conditions may be used (see lemma 3.1 of Phillips, 2007).*

For the sequence k_n used in lemmas 2.1 and 2.2, we have $\frac{k_n}{n^\alpha} \to \infty$ and it follows that when $t < k_n$, $\left\| n^{-\alpha/2} R_n^{-t} x_t \right\|_{L_2}$ either degenerates to zero (when $t = o(n^\alpha)$) or is bounded (when $t = O(k_n)$). Combined with lemma 2.2, we have

$$\sup_{1 \leq t \leq n} \left\| n^{-\alpha/2} R_n^{-t} x_t \right\|_{L_2} < \infty. \tag{2.12}$$

A similar uniform approximation to that of lemma 2.2 holds for ψ_{nt} in (2.9), with a slight modification to the index range of t as shown in the following lemma.

Lemma 2.3 *For k_n and k_n' satisfying the rate condition $\frac{n^{\alpha \vee \beta}}{k_n} + \frac{n^{\alpha \wedge \beta}}{k_n'} \to 0$ where $\alpha \vee \beta = \max(\alpha, \beta)$ and $\alpha \wedge \beta = \min(\alpha, \beta)$, we have*

$$\frac{1}{n^{\frac{\alpha}{2} + \alpha \wedge \beta}} R_n^{-t} \psi_{nt} = C_{z\alpha\beta} \tilde{Y}_C + o_p(1), \tag{2.13}$$

for all $t \in [k_n + k_n', n]$ where

$$C_{z\alpha\beta} := \begin{cases} -C_z^{-1}, & \text{if } \beta < \alpha \\ C^{-1}, & \text{if } \alpha < \beta \\ (C - C_z)^{-1}, & \text{if } \alpha = \beta \end{cases}.$$

Remark 2.2 *As the proof of the lemma in the Appendix reveals, the index set $[k_n + k_n',\ n]$ for t ensures the negligibility of certain frontal sums involving the standardized components of ψ_{nt}. As will become clear, the condition is used only in the proofs of the intermediate lemmas and does not appear in the main results because the full sums are dominated by the tail summation. Note that even for moderate sample sizes n, the set is well defined. For example, when $n = 30$, $\alpha \vee \beta = 0.8$ and $\alpha \wedge \beta = 0.6$, we have $n - n^{0.85} - n^{0.65} > 0$ so the index set is not empty and evidently has $O(n)$ observations as $n \to \infty$.*

Remark 2.3 *In (2.13) the standardization by $n^{\frac{\alpha}{2}+\alpha\wedge\beta}$ involves (potentially) both localizing coefficients α and β (depending on their respective magnitudes). The intuition for this standardization is that the quantity $\psi_{nt} = \sum_{j=1}^{t} R_{nz}^{t-j} x_{j-1} = \sum_{j=1}^{t} \left\{ R_{nz}^{t-j} R_n^{(j-1)} \right\} R_n^{-(j-1)} x_{j-1}$ involves the weighting matrices R_{nz}^{t-j} which downweight the components x_{j-1} in the sum, whereas these components themselves, being mildly explosive, are weighted by $R_n^{-(j-1)}$, which leads to an interacting weighting system involving the matrices $R_{nz}^{t-j} R_n^{(j-1)}$. Upon summation these weights may be dominated by the near stationary components that involve R_{nz} (and β) or the mildly explosive components that involve R_n (and α).*

Remark 2.4 *In consequence, both localizing coefficients α and β appear in the component limit theory for the numerator and denominator matrices of the IVX estimator (see lemma 2.4 below). Intuitively, if β is smaller than α, then the instruments z_t are near stationary and these near-stationary instruments tend to attenuate the mildly explosive behavior of x_t in the IV regression, thereby leading to the factor $n^{\alpha\wedge\beta}$ in the normalization and slowing down the rate of convergences in the components.*

As in (2.12), when $t < k_n + k_n'$, it can be shown that $\left\| n^{-(\alpha/2+(\alpha\wedge\beta))} R_n^{-t} \psi_{nt} \right\|_{L_2}$ is either degenerate or bounded. So we have

$$\sup_{1\le t\le n} \left\| n^{-(\alpha/2+(\alpha\wedge\beta))} R_n^{-t} \psi_{nt} \right\|_{L_2} < \infty. \tag{2.14}$$

With these results in hand, the limit theory follows for the numerator and denominator of the IVX estimator under (I5) mildly explosive regressors. The stronger signal in the remainder terms results in a new limit theory that involves the nuisance parameters C and $C_{z\alpha\beta}$.

Lemma 2.4 *(IVX Numerator and Denominator) Define $\tilde{\Psi}_{yy} := \int_0^\infty e^{-pC} Y_C Y_C' e^{-pC} dp$.*

(1) $vec\left(\frac{1}{n^{(\alpha\wedge\beta)}} \sum_{t=1}^{n} u_{0t} \tilde{z}_t' R_n^{-n} \right) \Longrightarrow (C C_{z\alpha\beta} \otimes I_m) \times MN\left(0, \tilde{\Psi}_{yy} \otimes \Omega_{00}\right).$

(2) $\frac{1}{n^{\alpha+(\alpha\wedge\beta)}} \sum R_n^{-n} x_t \tilde{z}_t' R_n^{-n} \Longrightarrow \tilde{\Psi}_{yy} \times (C C_{z\alpha\beta}).$

Remark 2.5 *The limit theory of the numerator given in lemma 2.4-(1) shows that $\sum_{t=1}^{n} u_{0t} z_t'$ becomes asymptotically negligible in relation to the term $\sum_{t=1}^{n} u_{0t} \psi_{nt}'$, which is usually a (negligible) remainder, and it is this term that dominates the asymptotics. Moreover, unlike the other cases (I1)-(I4), we no longer need the built-in serial correlation bias correction associated with the $O_p(n)$ sample covariance term $\sum_{t=1}^{n} u_{0t} z_t'$ and an estimate of the corresponding one sided long run covariance matrix. A similar property arises in the case of OLS estimation under mildly explosive regressors (Magdalinos and Phillips, 2009a).*

The limit theory of the IVX estimator (2.10) under (I5) is therefore equivalent to that of the OLS estimator in Magdalinos and Phillips (2009a, theorem 4.1).

Theorem 2.2 (Mildly Explosive Regressor) *With* $\alpha \in (2/q, 1)$ *and* $\beta \in (1/2, 1)$,

$$vec\left[n^{\alpha}\left(\tilde{A}_n - A\right) R_n^n\right] \Longrightarrow MN\left(0, \tilde{\Psi}_{yy}^{-1} \otimes \Omega_{00}\right),$$

where

$$\tilde{\Psi}_{yy} = \int_0^{\infty} e^{-pC} Y_C Y_C' e^{-pC} dp, \text{ and } Y_C \equiv N\left(0, \int_0^{\infty} e^{-pC} \Omega_{xx} e^{-pC} dp\right).$$

The results of the limit theory given in theorem 2.1 and theorem 2.2 for the locally explosive and mildly explosive cases depend on the unknown localizing coefficient matrix C and the unknown rate parameter α. So these results are not feasible for practical work as they stand. However,self normalized versions of the statistics have a chi-square limit theory in both the (I4) and (I5) cases, providing a convenient basis for inference.

We follow the usual (linear) hypothesis testing framework in which H is a known $r \times mK$ matrix of rank r and h is a known vector. Obvious extensions hold for general analytic restrictions.

Theorem 2.3 (Locally or Mildly Explosive Regressor) *Under* $H_0 : Hvec(A) = h$ *with* $\alpha \in (2/q, 1)$ *and* $\beta \in (1/2, 1)$,

$$\left(Hvec\left(\tilde{A}_n\right) - h\right)' \left[H\left\{\left(X'P_{\tilde{Z}}X\right)^{-1} \otimes \hat{\Omega}_{00}\right\} H'\right]^{-1} \left(Hvec\left(\tilde{A}_n\right) - h\right) \Longrightarrow \chi^2(r),$$

where

$$\left(X'P_{\tilde{Z}}X\right)^{-1} = \left\{\left(\sum_{t=1}^{n} x_t \tilde{z}_t'\right)\left(\sum_{t=1}^{n} \tilde{z}_t \tilde{z}_t'\right)^{-1}\left(\sum_{t=1}^{n} x_t \tilde{z}_t'\right)'\right\}^{-1}.$$

Although the limit theory of the IVX estimator under (I5) mildly explosive regressors differs from the other (I1)-(I4) cases, the usual chi-square limit theory for the self-normalized test statistic still holds. Theorem 2.3, taken together with theorem 3.8 of MP, therefore shows that IVX regression leads to a single inference procedure in all cases (I1)-(I5). The unified limit theory is helpful in empirical work where there is inevitable uncertainty about the degree of persistence in the regressors.

3 IVX Regression with Mixed Roots

The results above combined with those of MP show that the IVX approach is applicable in a wide vicinity of unity including all (I1) - (I5) cases. We might also expect the same method to be valid when there are mixed degrees of persistence in the regressors, which is likely in some empirical work (e.g. Campbell and Yogo, 2006). This section confirms that conjecture, showing the same limit theory given in theorem 2.2 holds when there are multiple regressors with mixed roots in the vicinity of unity.

For exposition, it is convenient to set $m = 1$ and $K = 2$ in (2.1) and consider the case where y_t is scalar and x_t is a bivariate AR(1) process with mixed roots. The simplified system has the form

$$y_t = a'x_t + u_{0t}, \; a' = \begin{bmatrix} a_1 & a_2 \end{bmatrix}, \; x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}, \tag{3.1}$$

$$x_t = R_n x_{t-1} + u_{xt}, \; R_n = \begin{bmatrix} \rho_n & 0 \\ 0 & \theta_n \end{bmatrix}, \; u_{xt} = \begin{bmatrix} u_{x1t} \\ u_{x2t} \end{bmatrix}. \tag{3.2}$$

We impose different degrees of persistence in the regressors with the following specification:

$$\begin{aligned} \rho_n &= 1 + \frac{c_1}{n^{\alpha_1}}, \text{ where } c_1 \in \begin{cases} (-\infty, 0), & \text{if } \alpha_1 \in (0,1) \\ (-\infty, \infty), & \text{if } \alpha_1 = 1 \end{cases}, \\ \theta_n &= 1 + \frac{c_2}{n^{\alpha_2}}, \text{ where } c_2 \in (0, \infty) \text{ and } \alpha_2 \in (0,1). \end{aligned}$$

Accordingly, x_{1t} falls under one of the specifications (I1)-(I4), while x_{2t} is a mildly explosive regressor corresponding to (I5). Dual manifestations of nonstationarity with different roots of this type have been analyzed in a different context by Phillips and Lee (2014) who considered inference about the roots in a vector autoregression of the type (3.2). Here we demonstrate the robustness of IVX estimation for the system parameter a in (3.1) in a mixed root regressor environment.

From the decomposition (2.3) we may express u_t in component form as

$$u_t = \begin{bmatrix} u_{0t} \\ u_{x1t} \\ u_{x2t} \end{bmatrix} = \begin{bmatrix} F_0(1)_{1\times 3} \\ F_{x1}(1)_{1\times 3} \\ F_{x2}(1)_{1\times 3} \end{bmatrix} \begin{bmatrix} \varepsilon_{0t} \\ \varepsilon_{x1t} \\ \varepsilon_{x2t} \end{bmatrix} - \triangle \begin{bmatrix} \tilde{\varepsilon}_{0t} \\ \tilde{\varepsilon}_{x1t} \\ \tilde{\varepsilon}_{x2t} \end{bmatrix},$$

under the same assumptions as (2.2). The long run variance matrices and the limit theory are the same as in (2.5), (2.6) and (2.7) except that the subscripts $0, 1$ and 2 now signify u_0, u_{x1} and u_{x2}, respectively. The IVX instrument is constructed in the same way as (2.8) with $C_z = diag(c_{z1}, c_{z2}) < 0$ and $\beta \in (1/2, 1)$.

The bias corrected IVX estimator of a has estimation error

$$\begin{aligned} \tilde{a} - a &= \begin{bmatrix} \tilde{a}_1 - a_1 \\ \tilde{a}_2 - a_2 \end{bmatrix} = \left(\tilde{Z}'X\right)^{-1} \left(\tilde{Z}'U_0 - n\hat{\Delta}_{x0}\right) \\ &= \begin{bmatrix} \sum_{t=1}^n \tilde{z}_{1t} x_{1t} & \sum_{t=1}^n \tilde{z}_{1t} x_{2t} \\ \sum_{t=1}^n \tilde{z}_{2t} x_{1t} & \sum_{t=1}^n \tilde{z}_{2t} x_{2t} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^n \tilde{z}_{1t} u_{0t} - n\hat{\Delta}_{01} \\ \sum_{t=1}^n \tilde{z}_{2t} u_{0t} - n\hat{\Delta}_{02} \end{bmatrix}. \end{aligned}$$

We employ the following normalizing matrices to accommodate the different orders of magnitude

$$D_n = \begin{bmatrix} n^{\frac{1+(\alpha_1 \wedge \beta)}{2}} & 0 \\ 0 & \theta_n^n n^{\alpha_2 \wedge \beta} \end{bmatrix}, \; C_n = \begin{bmatrix} n^{\frac{1+(\alpha_1 \wedge \beta)}{2}} & 0 \\ 0 & \theta_n^n n^{\alpha_2} \end{bmatrix}.$$

Then

$$C_n(\tilde{a}-a) = \left(D_n^{-1}\tilde{Z}'XC_n^{-1}\right)^{-1} D_n^{-1}\left(\tilde{Z}'U_0 - n\hat{\Delta}_{x0}\right).$$

To show the joint convergence of $n^{\frac{1+(\alpha_1\wedge\beta)}{2}}(\tilde{a}_1 - a_1)$ and $\theta_n^n n^{\alpha_2}(\tilde{a}_2 - a_2)$, we first show the asymptotic independence of these two components. As in Magdalinos and Phillips (2009a), the asymptotic behavior of $\theta_n^n n^{\alpha_2}(\tilde{a}_2 - a_2)$ is determined by two independent asymptotic Gaussian processes

$$Y_n = \begin{bmatrix} Y_{c_0 n} \\ Y_{c_2 n} \end{bmatrix} := \begin{bmatrix} \frac{1}{n^{\frac{\alpha_2}{2}}}\sum_{t=k_n+1}^{n}\theta_n^{-(n-t)}F_0(1)\varepsilon_t \\ \frac{1}{n^{\frac{\alpha_2}{2}}}\sum_{j=1}^{k_n}\theta_n^{-j}F_{x2}(1)\varepsilon_j \end{bmatrix} \Longrightarrow \begin{bmatrix} Y_{c_0} \\ Y_{c_2} \end{bmatrix} \equiv N\left(0_{2\times 1}, \begin{bmatrix} \frac{\Omega_{00}}{2c_2} & 0 \\ 0 & \frac{\Omega_{22}}{2c_2} \end{bmatrix}\right).$$

From MP, the vector martingale $\sum_{t=1}^{n}\xi_{nt}$ with

$$\xi_{nt} = \begin{bmatrix} \xi_{n1t} \\ \xi_{n2t} \end{bmatrix} := \begin{cases} \begin{bmatrix} \frac{1}{n^{\frac{1+\beta}{2}}} z_{1t-1}F_0(1)\varepsilon_t \\ \frac{1}{\sqrt{n}}F_{x1}(1)\varepsilon_t \end{bmatrix}, & \text{when } (\alpha_1\wedge\beta)=\beta, \\ & \text{i.e., } x_{1t} \text{ is I(1), I(2), I(4) or I(3) with } \beta<\alpha_1, \\ \begin{bmatrix} \frac{1}{n^{\frac{1+\alpha_1}{2}}} x_{1t-1}F_0(1)\varepsilon_t \\ \frac{1}{\sqrt{n}}F_{x1}(1)\varepsilon_t \end{bmatrix}, & \text{when } (\alpha_1\wedge\beta)=\alpha_1, \text{ i.e., } x_{1t} \text{ is I(3) with } \alpha_1<\beta, \end{cases}$$

determines the Gaussian limit theory of $n^{\frac{1+(\alpha_1\wedge\beta)}{2}}(\tilde{a}_1 - a_1)$. When $(\alpha_1\wedge\beta)=\beta$, the asymptotic independence between $\sum_{t=1}^{n}\xi_{n1t}$ and $\sum_{t=1}^{n}\xi_{n2t}$ is shown in proposition A1 in MP. The same proof also holds for the $(\alpha_1\wedge\beta)=\alpha_1$ case. The joint convergence of $n^{\frac{1+(\alpha_1\wedge\beta)}{2}}(\tilde{a}_1 - a_1)$ and $\theta_n^n n^{\alpha_2}(\tilde{a}_2 - a_2)$ is therefore achieved by showing the asymptotic independence between Y_n and $\sum_{t=1}^{n}\xi_{nt}$, which is done in the following lemma.

Lemma 3.1 *Y_n is asymptotically independent of the vector martingale $\sum_{t=1}^{n}\xi_{nt}$.*

The normalized denominator matrix is asymptotically diagonal under some mild rate conditions on β and α_i $(i=1,2)$, which enhances the development of the joint limit theory.

Lemma 3.2 *Under the rate condition $\alpha_1+\alpha_2<1+\beta$, we have*

$$D_n^{-1}\tilde{Z}'XC_n^{-1} \Longrightarrow \begin{bmatrix} \tilde{\Psi}_{11} & 0 \\ 0 & c_{xz2}(\alpha,\beta)c_2\tilde{\Psi}_{22} \end{bmatrix},$$

where $\tilde{\Psi}_{22} = \frac{Y_{c_2}^2}{2c_2}$, and

$$\tilde{\Psi}_{11} = \begin{cases} \frac{1}{-c_{z1}}\left\{\Omega_{11} + \int_0^1 B_1 dB_1\right\} & \textit{if } x_{1t} \textit{ is unit root : I(1),} \\ \frac{1}{-c_{z1}}\left\{\Omega_{11} + \int_0^1 J_x^{c1} dJ_x^{c1}\right\} & \textit{if } x_{1t} \textit{ is local to unity : (I2) or (I4),} \\ \frac{1}{-2c_{z1}}\Omega_{11} & \textit{if } x_{1t} \textit{ is mildly integrated - (I3) with } \beta<\alpha_1 \\ \frac{1}{-2(c_1+c_{z1})}\Omega_{11} & \textit{if } x_{1t} \textit{ is mildly integrated - (I3) with } \beta=\alpha_1 \\ \frac{1}{-2c_1}\Omega_{11} & \textit{if } x_{1t} \textit{ is mildly integrated - (I3) with } \alpha_1<\beta \end{cases}.$$

Remark 3.1 *Note that $\alpha_1 + \alpha_2 < 1 + \beta$ may not hold in some cases. In particular, if we have a local-to-unity regressor x_{1t} ($\alpha_1 = 1$) and a mildly explosive regressor x_{2t} whose root is very close to being local to unity (i.e., α_2 is close to unity), then $\alpha_2 < \beta$ may not hold and we may not have a diagonal limit for the moment matrix $\tilde{Z}'X$ upon standardization. For example, if $\alpha_1 = 1$ and $\beta < \alpha_2$, then $\alpha_1 + \alpha_2 < 1 + \beta$ fails. Intuitively, even though x_{1t} and x_{2t} have different orders of magnitude in this case, their asymptotic behavior is not distinct enough to ensure negligibility of the off diagonal elements when IVX leads to an instrument $\tilde{z}_{2t}$ that is close to stationarity (β close to 0.5). In such cases, the range of β for which $\tilde{Z}'X$ is diagonal asymptotically is restricted to the smaller region $\beta \in (\alpha_2, 1)$.*

The limit theory of $C_n(\tilde{a} - a)$ is therefore obtained from the independent marginal convergence of the two components and the mixed roots affect each of these components separately in the limit.

Theorem 3.1 *Under the rate conditions $\alpha_1 \in (1/3, 1)$ and $\beta \in ((\alpha_2 \vee 2/3), 1)$,*

$$C_n(\tilde{a} - a) \Longrightarrow MN\left(0_{2\times 1}, \begin{bmatrix} \tilde{\Phi}_{11}^{-1}\Omega_{00} & 0 \\ 0 & \tilde{\Psi}_{22}^{-1}\Omega_{00} \end{bmatrix}\right),$$

where $\tilde{\Psi}_{22} = \frac{Y_{c2}^2}{2c_2}$, $Y_{c2} \equiv N(0, \frac{\Omega_{22}}{2c_2})$ as in theorem 2.2, and $\tilde{\Psi}_{11}$ is given as

$$\tilde{\Phi}_{11}^{-1} = \begin{cases} \left\{\frac{1}{-c_{z1}}\left(\Omega_{11} + \int_0^1 B_1 dB_1\right)\right\}^{-2}\left(\frac{1}{-2c_{z1}}\Omega_{11}\right) & \text{if } x_{1t} \text{ is unit root - I(1)} \\ \left\{\frac{1}{-c_{z1}}\left\{\Omega_{11} + \int_0^1 J_x^{c1} dJ_x^{c1}\right\}\right\}^{-2}\left(\frac{1}{-2c_{z1}}\Omega_{11}\right) & \text{if } x_{1t} \text{ is local to unity - (I2) or (I4)} \\ \left(\frac{1}{-2c_{z1}}\Omega_{11}\right)^{-1} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \beta < \alpha_1 \\ \frac{(c_1 + c_{z1})^2}{c_1^2}\left(\frac{1}{-2c_1}\Omega_{11}\right)^{-1} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \beta = \alpha_1 \\ \left(\frac{1}{-2c_1}\Omega_{11}\right)^{-1} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \alpha_1 < \beta \end{cases}$$

.

Remark 3.2 *The rate condition $\alpha_1 \in (1/3, 1)$ ensures that the numerator of IVX estimator of the first component $\sum_{t=1}^n \tilde{z}_{1t} u_{0t} - n\hat{\Delta}_{01}$ has an asymptotic normal distribution. If we replace the weakly dependent structure of u_{0t} with an iid or mds structure for u_{0t}, as is common in predictive regressions, a wider region of α_1 is possible (see Kostakis et al., 2014). The condition that $\beta \in ((\alpha_2 \vee 2/3), 1)$ is explained as follows. First, the requirement that the choice parameter β exceeds $2/3$ accommodates consistent estimation of the long run covariance using $\hat{\Delta}_{x0}$ (see Lemma A0 in MP). The condition that β exceeds α_2 ensures asymptotic diagonality of the denominator matrix of IVX (the rate condition in Lemma 3.2). Section 4 shows that reliable choices of β are contained in this region.*

Remark 3.3 *Since $\tilde{\Psi}_{22} = \frac{Y_{c_2}^2}{2c_2}$ and $Y_{c_0} \equiv N\left(0, \frac{\Omega_{00}}{2c_2}\right)$ is independent of Y_{c_2}, we have*

$$\theta_n^n n^{\alpha_2}(\tilde{a}_2 - a_2) \Longrightarrow 2c_2 \frac{Y_{c_0}}{Y_{c_2}} \equiv 2c_2\left(\frac{\Omega_{00}}{\Omega_{22}}\right)^{1/2}\mathcal{C}$$

where $\mathcal{C}$ is a standard Cauchy variate, giving the same result as that of Magdalinos and Phillips (2009a, remark 4.1a). This result again confirms the asymptotic equivalence of IVX and OLS estimation with mildly explosive regressors.

As anticipated, the same self-normalized test statistic continues to deliver an asymptotic chi-square test free of any nuisance parameters.

Theorem 3.2 *Under $H_0 : a = a_0$ and the rate condition in Theorem 3.2,*

$$(\tilde{a} - a_0)' \left[\left(X' P_{\widetilde{Z}} X \right)^{-1} \hat{\Omega}_{00} \right]^{-1} (\tilde{a} - a_0) \Longrightarrow \chi^2 (2) \, ,$$

where $\left(X' P_{\widetilde{Z}} X \right)^{-1}$ is defined in Theorem 2.3.

This result demonstrates the robustness of the IVX approach to mixed degrees of persistence in the regressors, thereby providing a single valid procedure for inference that allows for a large class of persistent, but differently behaved regressors and widening the ambit of empirical research covered by this procedure. More general cases that allow for multivariate regressors with multiple mixed roots are treated in the same way. For example, the process $x_t = R_n x_{t-1} + u_{xt}$ may have a coefficient matrix of the form

$$R_n = \begin{bmatrix} (R_{1n})_{K_1 \times 1} & 0 \\ 0 & (R_{2n})_{K_2 \times 1} \end{bmatrix},$$

where R_{1n} has roots in the (I1)-(I4) class and R_{2n} involves mildly explosive coefficients of the form (I5). Analogous arguments to those in this section lead to asymptotic independence between suitably standardized versions of the processes for each group of regressors and the sample moment matrices $X' \tilde{Z}$ and $\tilde{Z}' \tilde{Z}$ will be asymptotically block diagonal after similar normalizations. Hence, in the same way as Theorem 3.2 we end up with a chi-square test that applies for a very general class of mixed regressors.

4 On the Choice of IVX Tuning Parameter

Implementation of IVX estimation requires choice of the tuning parameter β that is involved in the generation of the IVX instruments via (2.8). Evidently, larger values of β generally produce higher rates of convergence (c.f. theorem 2.1) and more efficient test procedures may therefore be expected. On the other hand, the central idea of IVX instrumentation – filtering a persistent regressor to generate an instrument of less persistence and ensure the validity of chi–squared test limit theory – suggests that we need to impose an upper bound for β that is less than unity.

To fix ideas in the following discussion simple, we use the predictive regression setting that has been widely adopted in empirical finance, whereby the one period ahead dependent variable y_{t+1} is used instead of y_t in (2.1) and the regression error u_{0t+1} is assumed to be a martingale difference. These modifications do not change the limit theory presented earlier and produce the standard environment of existing studies, such as Campbell and Yogo (2006) and Jansson and Moreira (2006). The martingale difference structure on u_{0t+1} implies that there is no predictability of y_{t+1} under the null hypothesis $H_0 : a = 0$. With this structure, the following DGP is imposed

$$\begin{aligned} y_{t+1} &= a' x_t + u_{0t+1},\ u_{0t+1} \sim mds\ \left(0, \Sigma_{00}\right), \\ x_{t+1} &= \left(I_K + \frac{C}{n}\right) x_t + u_{xt+1}, \end{aligned} \tag{4.1}$$

where u_{xt+1} is a linear process generated as in (2.2).

4.1 Inapplicability of MSE Criteria

Since larger values of β improve convergence rates whereas the IVX limit theory fails when $\beta = 1$, it might be expected that conventional asymptotic mean squared error (MSE) criteria (and associated cross validation approaches) might lead to suitable empirical choice criteria for β. As we now show, the asymptotic MSE criterion monotonically decreases as β increases, encouraging an upper bound unity choice for β.

To fix ideas let $m = 1$ and $K = 1$ in (4.1). Then

$$y_{t+1} = a x_t + u_{0t+1}, \tag{4.2}$$

$$x_{t+1} = \rho x_t + u_{xt+1},\ \rho = 1 + \frac{c}{n},\ c \in (-\infty, \infty) \tag{4.3}$$

Our analysis focuses on the (I2)-(I4) cases. Based on the IVX construction (2.8) with $\rho_z = 1 + \frac{c_z}{n^\beta}$, the generated AR(1) IVX series is

$$\tilde{z}_{t+1} = \rho_z \tilde{z}_t + \triangle x_t \sim \rho_z \tilde{z}_t + u_{xt+1},$$

and it is straightforward to show the decomposition

$$(\hat{a}_{IVX} - a) = \frac{\sum_{t=1}^n \tilde{z}_t u_{0.xt+1}}{\sum_{t=1}^n \tilde{z}_t x_t} + \left(\frac{\Sigma_{0x}}{\Omega_{xx}}\right) \frac{\sum_{t=1}^n \tilde{z}_t u_{xt+1}}{\sum_{t=1}^n \tilde{z}_t x_t},$$

where $u_{0.xt+1} = u_{0t+1} - \frac{\Sigma_{0x}}{\Omega_{xx}} u_{xt+1}$ and Ω_{xx} and Σ_{0x} are the corresponding (long-run) covariances. This decomposition is frequently used in predictive regression literature and is adapted here to the IVX regression framework to investigate possible choices of β.

By IVX limit theory we may use

$$\frac{\sum_{t=1}^n \tilde{z}_t u_{xt+1}}{\sum_{t=1}^n \tilde{z}_t x_t} = \frac{\sum_{t=1}^n \tilde{z}_t u_{xt+1}}{\sum_{t=1}^n \tilde{z}_t^2} + o_p(1) = (\hat{\rho}_z - \rho_z) + o_p(1)$$

since the difference between $\sum_{t=1}^n \tilde{z}_t x_t$ and $\sum_{t=1}^n \tilde{z}_t^2$ is negligible for the (I2)-(I4) cases, as shown in

MP. Then

$$
\begin{aligned}
(\hat{a}_{IVX} - a) &\sim \frac{\sum_{t=1}^{n} \tilde{z}_t u_{0.xt+1}}{\sum_{t=1}^{n} \tilde{z}_t x_t} + \left(\frac{\Sigma_{0x}}{\Omega_{xx}}\right)(\hat{\rho}_z - \rho_z) \\
&= \frac{\sum_{t=1}^{n} \tilde{z}_t u_{0.xt+1}}{\sum_{t=1}^{n} \tilde{z}_t x_t} + \left(\frac{\Sigma_{0x}}{\Omega_{xx}}\right)\left(\hat{\rho}_z - \rho_z - \frac{1}{n^{\beta}} Bias(\hat{\rho}_z)\right) + \frac{1}{n^{\beta}}\left(\frac{\Sigma_{0x}}{\Omega_{xx}}\right) Bias(\hat{\rho}_z)
\end{aligned}
$$

where the expression for the bias, $Bias(\hat{\rho}_z)$, can be found in Theorem 4.2 in Phillips and Magdalinos (2007b). Next observe that

$$
n^{\frac{1+\beta}{2}}(\hat{a}_{IVX} - a) = \frac{n^{-\frac{1+\beta}{2}} \sum_{t=1}^{n} \tilde{z}_t u_{0.xt+1}}{n^{-(1+\beta)} \sum_{t=1}^{n} \tilde{z}_t x_t} + \left(\frac{\Sigma_{0x}}{\Omega_{xx}}\right) n^{\frac{1+\beta}{2}} \left(\hat{\rho}_z - \rho_z - \frac{1}{n^{\beta}} Bias(\hat{\rho}_z)\right) + n^{\frac{1-\beta}{2}} \left(\frac{\Sigma_{0x}}{\Omega_{xx}}\right) Bias(\hat{\rho}_z)
$$

from which we deduce that

$$
AMSE\left(\hat{a}_{IVX}\right) = \frac{1}{n^{1+\beta}} V + \frac{1}{n^{2\beta}} B^2,
$$

where $V = var\left(\hat{a}_{IVX}\right)$ and $B = bias\left(\hat{a}_{IVX}\right)$ symbolically. Since $AMSE\left(\hat{a}_{IVX}\right)$ is strictly decreasing in $\beta \in (0,1)$ this criterion suggests that β be chosen as close to unity as possible. The approach therefore provides no informative guidance on an upper bound $\bar{\beta} < 1$ for use in practical work. Simulations with cross-validation methods (not reported here) show that these methods encounter the same difficulty.

4.2 Simulation-based guidance

This section reports simulations performed to assess the performance of IVX and provide some practical guidance on suitable IVX persistence (β or R_{nz}) choices in finite sample sizes. We follow the same DGP as in (4.1) with innovations

$$
u_t = \begin{pmatrix} u_{0t} \\ u_{xt} \end{pmatrix} \sim iid\ N(0_{(K+1)\times 1}, \Sigma_{(K+1)\times(K+1)}). \tag{4.4}
$$

The IVX instruments are constructed as in (2.8) and inference is based on the bias corrected IVX estimator (2.10). We set $C_z = -5$ and vary β to explore size and power properties according to different degrees of IVX persistence. In finite samples of size $n = 100, 250$, the value $C_z = -5$ and choices of $\beta \in (1/2, 1)$ (or $\beta \in (\alpha_2, 1)$) deliver a suitably wide range of autoregressive coefficients R_{nz} for the generation of the IVX instruments for investigation.

4.2.1 Single regressor cases

We run simulations with a single local to unity (rate parameter $\alpha = 1$) regressor. Although some of the main results in the paper relate to locally and mildly explosive regressors, we also include results for stationary-side local to unity and unit root cases. Accordingly localization parameters for the regressor are selected from $c \in \{-20, -2, 0, 2, 20\}$. The first three cases $\{-20, -2, 0\}$ have been commonly studied in earlier works. The mildly explosive choices $\{2, 20\}$ are new. The variance matrix of the innovations in (4.4) is parameterized as

$$\Sigma = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}$$

with $\delta = -0.95$. This value reflects realistic error correlation in predictive regressions and is commonly used in simulations reported in the literature.

Table 1 summarizes the size performance of predictability tests using the IVX estimator with various choices of β (or R_{nz}). The empirical size is calculated from the rejection frequency of one-sided standard normal test of $H_0 : a = 0$ based on the test statistic in Theorem 2.3. The nominal (asymptotic) test size is 0.05, the sample size (n) is 100 and the number of replications is $5,000$.

Table 1: Empirical size with a single local to unity regressor ($n = 100$)

		β	0.89	0.79	0.74	0.69	0.64	0.59	0.54
c	R_n	R_{nz}	0.92	0.87	0.83	0.79	0.74	0.67	0.58
-20	0.8		0.0601	0.0548	0.0575	0.0537	0.0589	0.0524	0.0483
-2	0.98		0.0645	0.0631	0.0633	0.0648	0.0577	0.0618	0.0565
0	1		0.0669	0.0671	0.0633	0.0606	0.0582	0.0545	0.0551
2	1.02		0.0718	0.0712	0.0668	0.0702	0.0588	0.0634	0.0556
20	1.2		0.0491	0.0478	0.0494	0.0479	0.0509	0.0515	0.0525

Evidently, size is well controlled and is robust across choices of $\beta \in (1/2, 1)$, with generated IVX instrument using $R_{nz} = 1 - \frac{5}{n^\beta}$. There is mild over-rejection in a few cases but the size rarely approaches 7% and generally lies between 4-7%. Size improves further for the larger sample size $n = 250$ as shown in Table 2.

Table 2: Empirical size with a single local to unity regressor ($n = 250$)

		β	0.93	0.75	0.69	0.63	0.57	0.51
c	R_n	R_{nz}	0.97	0.92	0.89	0.85	0.79	0.70
-20	0.92		0.0582	0.0561	0.0548	0.056	0.0589	0.0532
0	1		0.0684	0.0645	0.0645	0.0591	0.0587	0.0535
20	1.08		0.0521	0.0523	0.0477	0.0488	0.0475	0.0499

To investigate power performance, we used a sequence of local alternatives with $H_{a_n} : a_n = \frac{a}{n}$ for integer values of $a \in [0, 25]$ and various choices of $\beta \in (1/2, 1)$. Figures 1-2 show the power functions approach unity in all cases with more rapid convergence occurring for larger values of β, as expected. For the case $c = -20$ with $n = 100$, the convergence is slower and this scenario corresponds to a regressor that is stationary with autoregressive coefficient ($R_n = 0.8$) some distance from unity. In the unit root regressor case with $n = 100$, the power curves rise quickly to unity 1 for all choices of β, as shown in Figure 2 below.

Figure 1: Local Power, n=100, C= -20

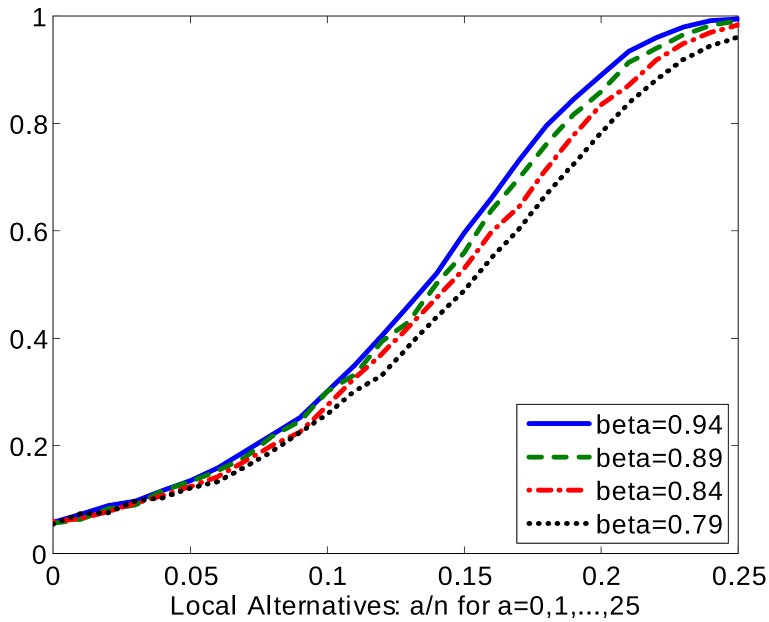
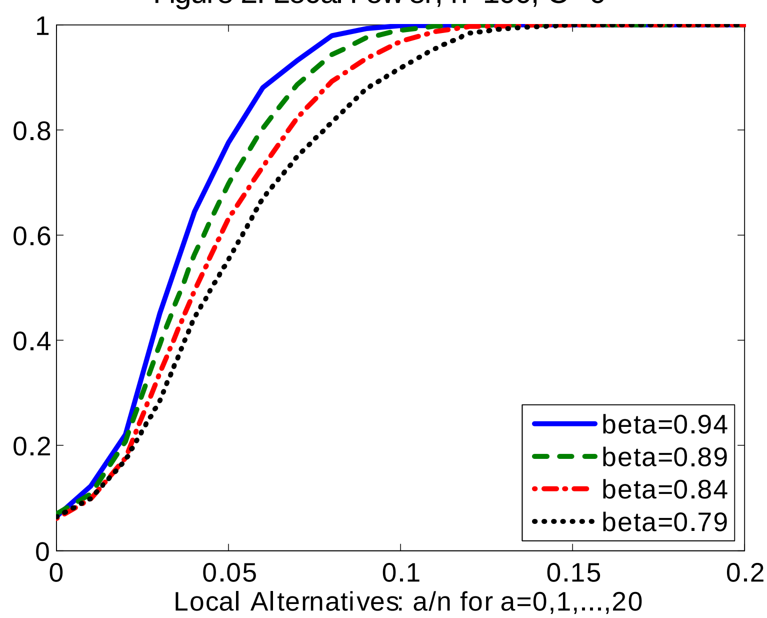


Figure 2: Local Power, n=100, C= 0



For locally explosive and mildly explosive cases, the local power reaches unity rapidly in all cases considered. In fact, the first non-zero alternative $H_{a_n} : a_n = 1/n$ already has unit power, so the results are not reported for this case. This outcome is anticipated since an explosive regressor is expected to have strong signal and predictive capability.

These results confirm the limit theory that the IVX procedure is robust for various tuning parameter choices $\beta \in (0.75, 0.95)$ and copes well with a range of empirically relevant persistent regressors in the I(1)-I(5) class.

4.2.2 Multiple regressor cases

We next consider predictive regressions with multiple regressors, which is relevant in much empirical practice. We take the bivariate case $K = 2$ in (4.1) to illustrate and consider two examples: (i) the stationary local to unity and unit root cases (with $c_1 = -20$ and $c_2 = 0$); and (ii) unit root and mildly explosive root ($c_1 = 0$ and $c_2 = 5$ with $\alpha_2 = 0.75$). These scenarios might be regarded as stock return predictive regressions using the T-bill rate and D/P ratio as regressors. Case (i) would then represent normal periods and case (ii) describes expansionary or boom periods. The innovation structure follows (4.4) and is given the covariance matrix

$$\Sigma = \begin{pmatrix} 1 & -0.9 & 0.1 \\ -0.9 & 1 & 0 \\ 0.1 & 0 & 1 \end{pmatrix}.$$

Table 3 reports size performance in testing $H_0 : a_1 = a_2 = 0$ using the χ^2 test of Theorem 3.2. The IVX persistence parameter is selected from $\beta \in (0.75, 1)$ and the localizing coefficients are set to $C_z = \text{diag}(-5, -5)$. The nominal (asymptotic) test size is 0.05, the sample size $n = 100$, and the number of replications is 5,000. Again, finite sample test size is evidently well controlled, with a few cases showing under-rejection.

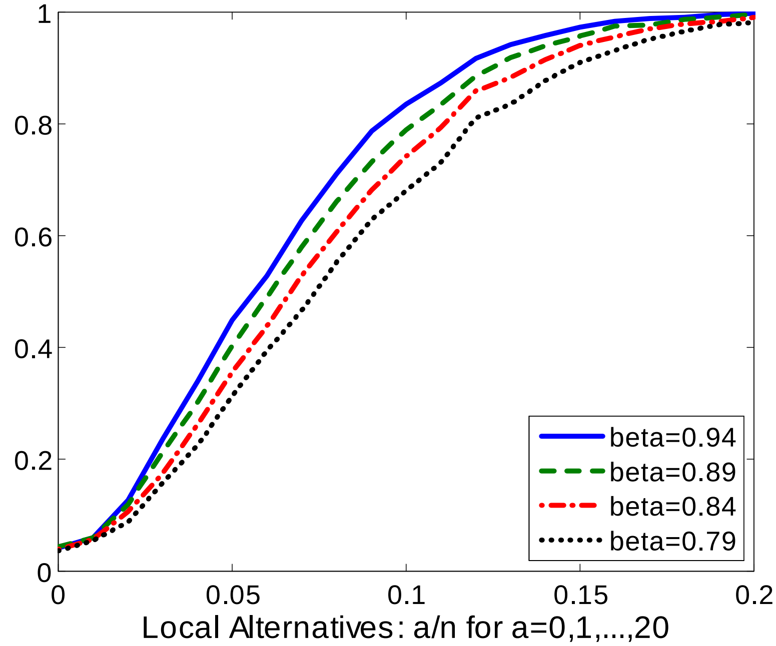
Table 3: Empirical size with multiple nonstationary regressors ($n = 100$)

	β	0.99	0.94	0.89	0.84	0.79
	R_{nz}	0.95	0.93	0.92	0.90	0.87
(I) $c_1 = -20$ ($\rho_n = 0.8$), $c_2 = 0$ ($\theta_n = 1$)		0.0466	0.0442	0.0462	0.0402	0.0388
(II) $c_1 = 0$ ($\rho_n = 1$), $c_2 = 5$ ($\theta_n = 1.16$), $\alpha_2 = 0.75$		0.0422	0.0393	0.037	0.0325	0.0328

Local alternatives were generated as before and the power functions showed evidence of fairly rapid convergence to unity, analogous to the single regressor case. Figure 3 illustrates with the results for case (i). For case (ii), the power curve again reaches unity at the first non-zero alternative

$a_n = \left(\frac{1}{n}, \frac{1}{n}\right)$ so these results are not reported.

Figure 3: Local Power, n=100, c1=-20, c2=0



These results reveal that finite sample size and power properties of IVX predictability tests seem reliable across a range of different cases. Flexible choices of $\beta \in (0.75, 0.95)$ with $C_z = -5$ seem to work well for finite sample sizes as low as $n = 100$, for single and multiple regressor cases, and for regressors in classes (I1)-(I6). These robustness findings corroborate recent simulation evidence for IVX testing reported in Kostakis et al (2014).

5 Conclusion

This paper shows that the IVX method of Magdalinos and Phillips (2009b) is robust under locally and mildly explosive regressors as well as mixed integrated regressors. The framework is sufficiently general that the regressors may have mixed degrees of persistence while still preserving the pivotal chi-square limit theory in testing. These results help econometric practice when, as is often the case, there are multiple regressors each of which manifests somewhat different forms of nonstationarity. Combined with the results of MP and those of Kostakis et al (2014), this limit theory for cointegrated systems gives a very general theory of regression that allows for a wide autoregressive parameter space in the vicinity of unity among the regressors. Unlike existing methods, there is no need for pretesting or simulation methods to cope with the unknown localizing coefficients. These advantages offer substantial convenience and robustness to empirical researchers working with co-moving systems of nonstationary data and predictive regressions involving data whose autoregressive roots are in a wide vicinity of unity.

A limitation of the IVX approach is that the localizing parameter β used in the construction of the instruments must be chosen by the empirical investigator. Asymptotic theory justifies a wide range of flexibility in the choice of β provided some general restrictions on this parameters are observed. But since the localizing rate α that controls the degree of persistence in the regressors is unknown these restrictions are imperfectly known. A convenient approach for much practical work is for investigators to simply assume that the regressors have unit roots, or are local to unity, or mildly explosive. This framework covers most practical situations including those that are commonly used in predictive regression. For this general setting, theory indicates that the procedure offers a wide degree of flexibility in the choice of β and the construction of the IVX instruments. Simulations confirm that this flexibility continues to hold in finite samples and that good size and power properties hold for all choices of $\beta \in (0, 75, 0.95)$.

6 Technical Appendix

This Appendix provides some useful preliminary lemmas and their proofs as well as proofs of the main theorems in the paper. The locally explosive and mildly explosive cases are investigated in the following sections and proofs for the case of mixed roots follow.

6.1 Locally Explosive Regressors: (I4)

We consider the (I4) locally explosive case ($C > 0, \alpha = 1$) with $\frac{1}{2} < \beta < 1$. Here, the same limit theory as MP continues to hold.

Lemma 6.1 *(Lemma 3.1 in PM with $C > 0$ and $\alpha = 1$)*

1. $n^{-\frac{1+\beta}{2}} \sum_{t=1}^{n} u_{0t} \tilde{z}_t' = n^{-\frac{1+\beta}{2}} \sum_{t=1}^{n} u_{0t} z_t' + o_p(1)$
2. $n^{-(1+\beta)} \sum_{t=1}^{n} x_t \tilde{z}_t' = n^{-(1+\beta)} \sum_{t=1}^{n} x_t z_t' - n^{-2} \sum_{t=1}^{n} x_{t-1} x_{t-1}' C C_z^{-1} + o_p(1)$
3. $n^{-(1+\beta)} \sum_{t=1}^{n} \tilde{z}_t \tilde{z}_t' = n^{-(1+\beta)} \sum_{t=1}^{n} z_t z_t' + o_p(1)$

Proof. 1. Propositions A1 and A2 in MP hold with $C > 0$ and $\alpha = 1$ without any substantial change in the proofs. To get the uniform bound for $E \left\| \psi_{nt} \right\|^2$, instead of using $\left\| R_n \right\|^{i-l} \leq 1$ for $l \leq i$ when $C < 0$, we can use when $C > 0$,

$$\left\| R_n \right\|^{i-l} \leq \exp(C) + o(1), \text{ for } i - l \leq n,$$

and we still have the same order of magnitude $\sup_{1 \leq t \leq n} E \left\| \psi_{nt} \right\|^2 = O(n^{1+2\beta})$ implying

$$n^{-\left(\frac{1+\beta}{2}+1\right)} \sum_{t=1}^{n} u_{0t} \psi_{nt}' = o_p(1), \text{ and } n^{-\frac{1+\beta}{2}} \sum_{t=1}^{n} u_{0t} \tilde{z}_t' = n^{-\frac{1+\beta}{2}} \sum_{t=1}^{n} u_{0t} z_t' + o_p(1).$$

Consequently, the limit theory of the numerator $n^{-\frac{1+\beta}{2}}\sum_{t=1}^{n}u_{0t}\tilde{z}_t'$ does not involve C at all. The proofs of 2 and 3 do not depend on the sign of C and use the distributional limit result

$$n^{-1/2}x_{\lfloor ns\rfloor}\Longrightarrow J_x^c(r)$$

for $C>0$ from Phillips (1987). ■

The limit theory of the IVX estimator (2.10) is therefore the same as the case under (I2), or N(ii) in MP, proving Theorem 3.1-1 and 3.2 for (I4) locally explosive regressors.

6.2 Mildly Explosive Regressors: (I5)

We collect proofs for the (I5) mildly explosive regressor case here. As in MP, we consider two possible cases $\frac{2}{q}<\beta<\alpha<1$ and $\frac{2}{q}<\alpha<\beta<1$. In both cases, IVX asymptotics with mildly explosive regressors is developed and the pivotal chi-square limit theory is shown to be valid.

Proof of Lemma 2.2. By back substitution and from the BN decomposition

$$\frac{1}{n^{\alpha/2}}R_n^{-(j-1)}x_{j-1}=\frac{1}{n^{\alpha/2}}\sum_{i=1}^{j-1}R_n^{-i}u_{xi}=\frac{1}{n^{\alpha/2}}\sum_{i=1}^{j-1}R_n^{-i}F_x(1)\varepsilon_i-\frac{1}{n^{\alpha/2}}\sum_{i=1}^{j-1}R_n^{-i}\triangle\tilde{\varepsilon}_{xi}.$$

Consider the partial sum process $\eta_n(r)=\frac{1}{\sqrt{n}}S_{\lfloor nr\rfloor}$ with $S_j=\sum_{i=1}^{j}\varepsilon_i$. Using a strong approximation, we can enlarge the original probability space and construct a vector Brownian motion $\omega=BM\left(\Sigma\right)$ on this space with the property that for $\varepsilon>0$, $E\left[\|\varepsilon_i\|^q\right]<\infty$,

$$\sup_{0\le r\le 1}\|\eta_n(r)-\omega\left(r\right)\|=o_{a.s.}\left(n^{-1/2+1/q+\varepsilon}\right).$$

Define $\frac{e_i}{\sqrt{n}}=\omega\left(\frac{i}{n}\right)-\omega\left(\frac{i-1}{n}\right)$ whose distribution is iid normal with $E\left(e_ie_i'\right)=\Sigma$. We generate another mildly explosive AR(1) process,

$$z_i=R_nz_{i-1}+v_{xi},$$

where $v_{xi}=F_x(1)e_i-\triangle\widetilde{e}_{xi}$ is a linear process defined in a similar way to u_{xi} but using e_i instead of ε_i. Then

$$\frac{1}{n^{\alpha/2}}R_n^{-(j-1)}z_{j-1}=\frac{1}{n^{\alpha/2}}\sum_{i=1}^{j-1}R_n^{-i}F_x(1)e_i-\frac{1}{n^{\alpha/2}}\sum_{i=1}^{j-1}R_n^{-i}\triangle\widetilde{e}_{xi},$$

and

$$\frac{1}{n^{\alpha/2}}\sum_{i=1}^{j-1}R_n^{-i}F_x(1)e_i:=\tilde{Y}_{Cn}\equiv N\left(0,\frac{1}{n^{\alpha}}\sum_{i=1}^{j-1}R_n^{-i}\Omega_{xx}R_n^{-i}\right),$$

by construction. Define the limit process $\lim_n\tilde{Y}_{Cn}=\tilde{Y}_C\equiv N\left(0,\int_0^\infty e^{-pC}\Omega_{xx}e^{-pC}dp\right)$. The strong approximation between these two processes will be shown below in (iii). Observe that

$$\left\| \frac{1}{n^{\alpha/2}} R_n^{-(j-1)} x_{j-1} - \tilde{Y}_C \right\| \le \left\| \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \varepsilon_i - \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} F_x(1) e_i \right\| \tag{6.1}$$

$$+ \left\| \tilde{Y}_{Cn} - \tilde{Y}_C \right\| + \left\| \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} \triangle \tilde{\varepsilon}_{xi} \right\| + \left\| \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} \triangle \widetilde{e}_{xi} \right\|.$$

We now show that each term in (6.1) is $o_{a.s.}(1)$ uniformly over $j-1 \in [k_n, n]$.

(i) $\sup_{k_n \le j-1 \le n} \left\| \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \varepsilon_i - \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} F_x(1) e_i \right\| = o_{a.s}(1)$.

$$\begin{aligned}
\frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \varepsilon_i &= \frac{n^{1/2}}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \left(\eta_n \left(\frac{i}{n} \right) - \eta_n \left(\frac{i-1}{n} \right) \right) \\
&= \frac{n^{1/2}}{n^{\alpha/2}} \left\{ \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \eta_n \left(\frac{i}{n} \right) - \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \eta_n \left(\frac{i-1}{n} \right) \right\} \\
&= \frac{n^{1/2}}{n^{\alpha/2}} \left\{ \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \eta_n \left(\frac{i}{n} \right) - \sum_{s=0}^{j-2} R_n^{-s-1} F_x(1) \eta_n \left(\frac{s}{n} \right) \right\} \\
&= \frac{n^{1/2}}{n^{\alpha/2}} \left\{ \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \eta_n \left(\frac{i}{n} \right) - \sum_{i=1}^{j-1} R_n^{-i-1} F_x(1) \eta_n \left(\frac{i}{n} \right) + R_n^{-j-2} F_x(1) \eta_n \left(\frac{j-1}{n} \right) \right\} \\
&= \frac{n^{1/2}}{n^{\alpha/2}} \left\{ \left(1 - R_n^{-1}\right) \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \eta_n \left(\frac{i}{n} \right) + R_n^{-j-2} F_x(1) \eta_n \left(\frac{j-1}{n} \right) \right\}.
\end{aligned}$$

Similarly,

$$\frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} F_x(1) e_i = \frac{n^{1/2}}{n^{\alpha/2}} \left\{ \left(1 - R_n^{-1}\right) \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \omega \left(\frac{i}{n} \right) + R_n^{-j-2} F_x(1) \omega \left(\frac{j-1}{n} \right) \right\}.$$

Therefore,

$$\begin{aligned}
&\left\| \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \varepsilon_i - \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} F_x(1) e_i \right\| \\
\le\ & \frac{n^{1/2}}{n^{\alpha/2}} \left\| F_x(1) \right\| \left(\sup_{0 \le r \le 1} \left\| \eta_n(r) - \omega(r) \right\| \right) \left\| 1 - R_n^{-1} \right\| \sum_{i=1}^{j-1} \left\| R_n \right\|^{-i} \\
&+ \frac{n^{1/2}}{n^{\alpha/2}} \left\| R_n \right\|^{-j-2} \left\| F_x(1) \right\| \left(\sup_{0 \le r \le 1} \left\| \eta_n(r) - \omega(r) \right\| \right) \\
=\ & \frac{n^{1/2}}{n^{\alpha/2}} \left(\sup_{0 \le r \le 1} \left\| \eta_n(r) - \omega(r) \right\| \right) \left\| F_x(1) \right\| \left(\left\| 1 - R_n^{-1} \right\| \sum_{i=1}^{j-1} \left\| R_n \right\|^{-i} + \left\| R_n \right\|^{-j-2} \right),
\end{aligned}$$

and for $k_n \le j \le n$, $\left(\left\| 1 - R_n^{-1} \right\| \sum_{i=1}^{j-1} \left\| R_n \right\|^{-i} + \left\| R_n \right\|^{-j-2} \right)$ is $O(1)$ since $\left\| R_n \right\|^{-j} \le \exp(-\min(c_i) \frac{j}{n^{\alpha}}) +$

$o(1) = o(1)$ with $\frac{n^\alpha}{k_n} + \frac{n^\alpha}{n} \to 0$.

Therefore,

$$\sup_{k_n \le j-1 \le n} \left\| \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} F_x(1) \varepsilon_i - \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} F_x(1) e_i \right\| = \frac{n^{1/2}}{n^{\alpha/2}} o_{a.s.} \left(n^{-1/2+1/q+\varepsilon} \right) = o_{a.s.} \left(n^{\frac{1}{q} - \frac{\alpha}{2} + \varepsilon} \right),$$

and the exponent $\frac{1}{q} - \frac{\alpha}{2} + \varepsilon < 0$ for small enough ε because $\alpha > \frac{2}{q}$ and $q \geq 4$. The strong approximation is therefore sharper under higher moment conditions and when the signal strength of the mildly explosive regressors is closer to the local to unity region. Conversely, if the signal strength is closer to the purely explosive case ($\alpha \leq \frac{1}{2}$) higher moment conditions are needed to ensure that $\frac{1}{q} - \frac{\alpha}{2} + \varepsilon < 0$. The approximation is not possible in the pure explosive case ($\alpha = 0$).

(ii) $\sup_{k_n \le j-1 \le n} \left\| \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} \triangle \tilde{\varepsilon}_{xi} \right\| = o_{a.s}(1) = \sup_{k_n \le j-1 \le n} \left\| \frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} \triangle \widetilde{e}_{xi} \right\|$.

Using summation by parts

$$\frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-1} R_n^{-i} \triangle \tilde{\varepsilon}_{xi} = -\frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-2} \triangle R_n^{-i-1} \tilde{\varepsilon}_{xi} - \frac{R_n^{-1}}{n^{\alpha/2}} \tilde{\varepsilon}_{x0} + \frac{R_n^{-(j-1)}}{n^{\alpha/2}} \tilde{\varepsilon}_{xj-1},$$

and from the given moment condition

$$\begin{aligned} P\left(\left\| \frac{1}{n^{\alpha/2}} \tilde{\varepsilon}_{xn} \right\| > \epsilon \right) &\leq \frac{E \left\| \tilde{\varepsilon}_{xn} \right\|^q}{n^{\frac{q\alpha}{2}} \epsilon^q} = O\left(\frac{1}{n^{\frac{q\alpha}{2}}} \right), \\ \sum_{n=1}^{\infty} P\left(\left\| \frac{1}{n^{\alpha/2}} \tilde{\varepsilon}_{xn} \right\| \right) &\leq O\left(\sum_{n=1}^{\infty} \frac{1}{n^{\frac{q\alpha}{2}}} \right) < \infty, \end{aligned}$$

as long as $q\alpha > 2$ which is satisfied by the condition $\alpha > \frac{2}{q}$. Thus, $n^{-\alpha/2} R_n^{-1} \tilde{\varepsilon}_{x0} = o_{a.s.}(1)$ by the Borel-Cantelli lemma. Using the fact that with $\|R_n\|^{-j} = o(1)$ when $j-1 \in [k_n, n]$, we also have $n^{-\alpha/2} R_n^{-(j-1)} \tilde{\varepsilon}_{xj-1} = o_{a.s.}(1)$. Note further that

$$\frac{1}{n^{\alpha/2}} \sum_{i=1}^{j-2} \triangle R_n^{-i-1} \tilde{\varepsilon}_{xi} = \frac{R_n^{-1} C}{n^{\frac{3\alpha}{2}}} \sum_{i=1}^{j-2} R_n^{-i} \tilde{\varepsilon}_{xi},$$

and $R_n^{-1} C = O(1)$. For $k_n \leq j \leq n$

$$E\left(\left\| \frac{1}{n^{\frac{3\alpha}{2}}} \sum_{i=1}^{j-2} R_n^{-i} \tilde{\varepsilon}_{xi} \right\|^q \right)^{\frac{1}{q}} = \left\| \frac{1}{n^{\frac{3\alpha}{2}}} \sum_{i=1}^{j-2} R_n^{-i} \tilde{\varepsilon}_{xi} \right\|_{L_q} \leq \frac{\|\tilde{\varepsilon}_{xi}\|_{L_q}}{n^{\frac{3\alpha}{2}}} \sum_{i=1}^{j-2} \|R_n\|^{-i} = O\left(\frac{1}{n^{\frac{\alpha}{2}}} \right),$$

since $\tilde{\varepsilon}_{xi}$ is stationary and $n^{-\alpha} \sum_{i=1}^{j-2} \|R_n\|^{-i} = O(1)$ for $k_n \leq j \leq n$.Hence

$$E\left(\left\| \frac{1}{n^{\frac{3\alpha}{2}}} \sum_{i=1}^{j-2} R_n^{-i} \tilde{\varepsilon}_{xi} \right\|^q \right) = O\left(\frac{1}{n^{\frac{q\alpha}{2}}} \right),$$

and

$$P\left(\left\|\frac{1}{n^{\frac{3\alpha}{2}}}\sum_{i=1}^{j-2}R_n^{-i}\tilde{\varepsilon}_{xi}\right\|>\epsilon\right)\leq\frac{E\left(\left\|\frac{1}{n^{\frac{3\alpha}{2}}}\sum_{i=1}^{j-2}R_n^{-i}\tilde{\varepsilon}_{xi}\right\|^q\right)}{\epsilon^{\frac{q}{2}}}=O\left(\frac{1}{n^{\frac{q\alpha}{2}}}\right).$$

We then have $n^{-3\alpha/2}R_n^{-1}C\sum_{i=1}^{j-2}R_n^{-i}\tilde{\varepsilon}_{xi}=o_{a.s}(1)$ under the given moment conditions just as before. The proof for $n^{-\alpha/2}\sum_{i=1}^{j-1}R_n^{-i}\triangle\widetilde{e}_{xi}$ is exactly same.

(iii) $\left\|\tilde{Y}_{Cn}-\tilde{Y}_C\right\|=o_{a.s.}(1)$

For clarity, we denote $n^{-\alpha/2}\sum_{i=1}^{j-1}R_n^{-i}F_x(1)e_i=\tilde{Y}_{Cn}:=\tilde{Y}_{Cn,j-1}$, and define the martingale array $\left\{\mathcal{F}_{n,j-1},\tilde{Y}_{Cn,j-1}:j-1\geq k_n\right\}$ with natural filtration $\mathcal{F}_{n,j-1}$. We have $E_{\mathcal{F}_{n,j-1}}\left[\tilde{Y}_{Cn,j}\right]=\tilde{Y}_{Cn,j-1}$ and

$$E\left[\tilde{Y}_{Cn,j}\tilde{Y}_{Cn,j}'\right]=\frac{1}{n^\alpha}\sum_{i=1}^{j-1}R_n^{-i}\Omega_{xx}R_n^{-i}\to\int_0^\infty e^{-pC}\Omega_{xx}e^{-pC}dp<\infty\text{ for }j-1\geq k_n.$$

By the martingale convergence theorem for L_2-bounded martingales (e.g., Hall and Heyde, 1980),

$$\tilde{Y}_{Cn,j-1}\to_{a.s.}\tilde{Y}_C.$$

and $\tilde{Y}_C\equiv N\left(0,\int_0^\infty e^{-pC}\Omega_{xx}e^{-pC}dp\right)$, which is a distributionally equivalent copy of Y_C. Combining (i)-(iii) gives the required result. ■

Proof of Lemma 2.3. Since k_n satisfies $\frac{n^{\alpha\vee\beta}}{k_n}\to 0$, we have both $\|R_n\|^{-k_n}\to 0$ and $\|R_{nz}\|^{k_n}\to 0$, so the condition for k_n in lemma 2.2 holds. In addition $\|R_{nz}\|^t\to 0$ and $\|R_n\|^{-t}\to 0$ since $t>k_n$. We have

$$\frac{1}{n^{\frac{\alpha}{2}+\alpha\wedge\beta}}R_n^{-t}\psi_{nt}=\frac{1}{n^{\frac{\alpha}{2}+\alpha\wedge\beta}}R_n^{-t}\sum_{j=1}^{t}R_{nz}^{t-j}x_{j-1}=\frac{1}{n^{\frac{\alpha}{2}+\alpha\wedge\beta}}R_n^{-t}\sum_{j=k_n+1}^{t}R_{nz}^{t-j}x_{j-1}+o_p(1),$$

because the frontal summation over $1\leq j\leq k_n$ is negligible as we now show. In particular, using (2.12), we have

$$\begin{aligned}\left\|\frac{1}{n^{\frac{\alpha}{2}+\alpha\wedge\beta}}R_n^{-t}\sum_{j=1}^{k_n}R_{nz}^{t-j}x_{j-1}\right\|_{L_1}&=\left\|\frac{1}{n^{\alpha\wedge\beta}}R_n^{-t}\sum_{j=1}^{k_n}R_{nz}^{t-j}R_n^{(j-1)}\left(\frac{R_n^{-(j-1)}x_{j-1}}{n^{\alpha/2}}\right)\right\|_{L_1}\\&\leq\sup_{j-1}\left\|\frac{R_n^{-(j-1)}x_{j-1}}{n^{\alpha/2}}\right\|_{L_2}\frac{1}{n^{\alpha\wedge\beta}}\|R_n\|^{-t}\|R_{nz}\|^t\sum_{j=1}^{k_n}\|R_{nz}\|^{-j}\|R_n\|^{j-1}\\&=O\left(\frac{1}{n^{\alpha\wedge\beta}}\|R_{nz}\|^{t-1}\|R_n\|^{-t}\frac{\|R_{nz}\|^{-k_n}\|R_n\|^{k_n}-1}{\|R_{nz}\|^{-1}\|R_n\|-1}\right)\\&=O\left(\frac{\|R_{nz}\|^{t-k_n-1}\|R_n\|^{k_n-t}-\|R_n\|^{-t}\|R_{nz}\|^{t-1}}{n^{\alpha\wedge\beta}\left(\|R_{nz}\|^{-1}\|R_n\|-1\right)}\right)=o(1),\end{aligned}$$

since

$$n^{\alpha\wedge\beta}\left(\left\|R_{nz}\right\|^{-1}\left\|R_n\right\|-1\right) = n^{\alpha\wedge\beta}\left(\frac{1+\frac{\max(c_i)}{n^\alpha}}{1+\frac{\max(c_{zi})}{n^\beta}}-1\right)=n^{\alpha\wedge\beta}\left(\frac{\frac{\max(c_i)}{n^\alpha}-\frac{\max(c_{zi})}{n^\beta}}{1+\frac{\max(c_{zi})}{n^\beta}}\right)$$

$$\to \left\{\begin{array}{cc} -\max\left(c_{zi}\right), & \text{if } \beta<\alpha \\ \max\left(c_i\right), & \text{if } \alpha<\beta \\ \max\left(c_i\right)-\max\left(c_{zi}\right), & \text{if } \alpha=\beta \end{array}\right. ,$$

which is non-zero and finite in all cases (recall $c_{zi}<0$, $c_i>0$ for all i). Also $\left\|R_n\right\|^{-t}\left\|R_{nz}\right\|^t=o(1)$ and $\left\|R_{nz}\right\|^{t-k_n}\left\|R_n\right\|^{k_n-t}=o\left(1\right)$ where the second equality holds because

$$\left\|R_{nz}\right\|^{t-k_n}\left\|R_n\right\|^{k_n-t}=O\left(e^{\max(c_{zi})\frac{t-k_n}{n^\beta}}e^{-\max(c_i)\frac{t-k_n}{n^\alpha}}\right),$$

and $\frac{k_n'}{n^{\alpha\wedge\beta}}\leq\frac{t-k_n}{n^{\alpha\wedge\beta}}\to\infty$ by the given rate condition. Hence $n^{-\left(\frac{\alpha}{2}+\alpha\wedge\beta\right)}R_n^{-t}\sum_{j=1}^{k_n}R_{nz}^{t-j}x_{j-1}=o_p(1)$.

Now we have

$$\begin{aligned}\frac{1}{n^{\frac{\alpha}{2}+\alpha\wedge\beta}}R_n^{-t}\psi_{nt} &= \frac{1}{n^{\frac{\alpha}{2}+\alpha\wedge\beta}}R_n^{-t}\sum_{j=k_n+1}^{t}R_{nz}^{t-j}x_{j-1}+o_p(1)\\ &= \frac{1}{n^{\alpha\wedge\beta}}R_n^{-t}\sum_{j=k_n+1}^{t}R_{nz}^{t-j}R_n^{(j-1)}\left(R_n^{-(j-1)}\frac{x_{j-1}}{n^{\alpha/2}}\right)+o_p(1)\end{aligned}$$

$$\begin{aligned}&= \left(\frac{1}{n^{\alpha\wedge\beta}}R_n^{-t}\sum_{j=k_n+1}^{t}R_{nz}^{t-j}R_n^{(j-1)}\right)\left(\tilde{Y}_C+o_p\left(1\right)\right)\ \text{(using lemma 2.2)}\\ &= \left(\frac{1}{n^{\alpha\wedge\beta}}R_n^{-t}\sum_{j=1}^{t}R_{nz}^{t-j}R_n^{(j-1)}\right)\left(\tilde{Y}_C+o_p\left(1\right)\right),\ \text{(putting back the negligible front sum).}\end{aligned}$$

Note that

$$\begin{aligned}&\frac{1}{n^{(\alpha\wedge\beta)}}\sum_{j=1}^{t}R_{nz}^{t-j}R_n^{j-1-t}\\ =& \frac{1}{n^{(\alpha\wedge\beta)}}diag\left(\sum_{j=1}^{t}\left(1+\frac{c_{z1}}{n^\beta}\right)^{t-j}\left(1+\frac{c_1}{n^\alpha}\right)^{j-1-t},...,\sum_{j=1}^{n}\left(1+\frac{c_{zK}}{n^\beta}\right)^{t-j}\left(1+\frac{c_K}{n^\alpha}\right)^{j-1-t}\right).\end{aligned}$$

For all i

$$\frac{1}{n^{(\alpha\wedge\beta)}}\sum_{j=1}^{t}\left(1+\frac{c_{zi}}{n^{\beta}}\right)^{t-j}\left(1+\frac{c_i}{n^{\alpha}}\right)^{j-1-t}$$

$$= \frac{1}{n^{(\alpha\wedge\beta)}}\left(1+\frac{c_{zi}}{n^{\beta}}\right)^{t-1}\left(1+\frac{c_i}{n^{\alpha}}\right)^{-t}\frac{\left\{\left(1+\frac{c_{zi}}{n^{\beta}}\right)^{-t}\left(1+\frac{c_i}{n^{\alpha}}\right)^{t}-1\right\}}{\left(1+\frac{c_{zi}}{n^{\beta}}\right)^{-1}\left(1+\frac{c_i}{n^{\alpha}}\right)-1}$$

$$= \frac{1}{n^{(\alpha\wedge\beta)}}\frac{\left\{\left(1+\frac{c_{zi}}{n^{\beta}}\right)^{-1}-\left(1+\frac{c_{zi}}{n^{\beta}}\right)^{t-1}\left(1+\frac{c_i}{n^{\alpha}}\right)^{-t}\right\}}{\left(1+\frac{c_{zi}}{n^{\beta}}\right)^{-1}\left(1+\frac{c_i}{n^{\alpha}}\right)-1} = \frac{1+o(1)}{n^{(\alpha\wedge\beta)}\left(\frac{c_i}{n^{\alpha}}-\frac{c_{zi}}{n^{\beta}}-\frac{cc_{zi}}{n^{\alpha+\beta}}\right)}$$

$$\to -\frac{1}{c_{zi}} \text{ if } \beta<\alpha,\ \frac{1}{c_i} \text{ if } \alpha<\beta \text{ and } \frac{1}{c_i-c_{zi}} \text{ if } \alpha=\beta.$$

Hence,

$$\frac{1}{n^{(\alpha\wedge\beta)}}R_n^{-t}\sum_{j=1}^{t}R_{nz}^{t-j}R_n^{(j-1)} \to \left\{\begin{array}{cc} -C_z^{-1}, & \text{if } \beta<\alpha \\ C^{-1}, & \text{if } \alpha<\beta \\ (C-C_z)^{-1}, & \text{if } \alpha=\beta \end{array}\right\} =: C_{z\alpha\beta}, \tag{6.2}$$

giving the required result. ■

Proof of Lemma 2.4. Part (1). From (2.9), we have

$$\begin{aligned}\frac{1}{n^{(\alpha\wedge\beta)}}\sum_{t=1}^{n}u_{0t}\tilde{z}_t'R_n^{-n} &= \frac{1}{n^{(\alpha\wedge\beta)}}\sum_{t=1}^{n}u_{0t}z_t'R_n^{-n}+\frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum_{t=1}^{n}u_{0t}\psi_{nt}'R_n^{-n}C \\ &= \frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum_{t=1}^{n}u_{0t}\psi_{nt}'R_n^{-n}C+o_p(1),\end{aligned}$$

where the last equality holds because R_n^{-n} dominates the order of magnitude of $\sum_{t=1}^{n}u_{0t}z_t' = O_p(n)$ (c.f., Magdalinos and Phillips, 2009a, equation (10)).

To use lemma 2.3 we first show that $n^{-(\alpha+(\alpha\wedge\beta))}\sum_{t=1}^{k_n+k_n'-1}u_{0t}\psi_{nt}'R_n^{-n} = o_p(1)$ for k_n and k_n' satisfying the conditions of lemma 2.3. Note that

$$\left\|\frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum_{t=1}^{k_n+k_n'-1}u_{0t}\psi_{nt}'R_n^{-n}\right\|_{L_1} = \left\|\frac{1}{n^{\frac{\alpha}{2}}}\sum_{t=1}^{k_n+k_n'-1}u_{0t}\left(\frac{1}{n^{\frac{\alpha}{2}+(\alpha\wedge\beta)}}R_n^{-t}\psi_{nt}\right)'R_n^{t-n}\right\|_{L_1}$$

$$\leq \left(\sup_t\|u_{0t}\|_{L_2}\right)\left(\sup_t\left\|\frac{1}{n^{\frac{\alpha}{2}+(\alpha\wedge\beta)}}R_n^{-t}\psi_{nt}\right\|_{L_2}\right)\frac{\|R_n\|^{-n}}{n^{\frac{\alpha}{2}}}\sum_{t=1}^{k_n+k_n'-1}\|R_n\|^{t} \text{ and using (2.14),}$$

$$= O\left(\frac{n^{\frac{\alpha}{2}}\|R_n\|^{k_n+k_n'-n}}{n^{\alpha}(1-\|R_n\|)}\right) = O\left(n^{\frac{\alpha}{2}}\|R_n\|^{k_n+k_n'-n}\right) = o(1),$$

because of the exponentially fast convergence of $\|R_n\|^{k_n+k_n'-n}\to 0$.

Hence, we have $n^{-(\alpha+\alpha\wedge\beta)}\sum_{t=1}^{k_n+k_n'-1}u_{0t}\psi_{nt}'R_n^{-n} = o_p(1)$.

Using the same sum splitting argument as in lemma 2.3,

$$vec\left(\frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum_{t=1}^{n}u_{0t}\psi_{nt}'R_n^{-n}\right)=vec\left(\frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum_{t=k_n+k_n'}^{n}u_{0t}\psi_{nt}'R_n^{-n}\right)+o_p(1)$$

$$\begin{aligned}
&= vec\left(\frac{1}{n^{\frac{\alpha}{2}}}\sum_{t=k_n+k_n'}^{n}u_{0t}\left(\frac{1}{n^{\frac{\alpha}{2}+(\alpha\wedge\beta)}}R_n^{-t}\psi_{nt}\right)'\left(R_n^{t-n}\right)'\right)+o_p(1)\\
&= vec\left(\frac{1}{n^{\frac{\alpha}{2}}}\sum_{t=k_n+k_n'}^{n}u_{0t}\left(C_{z\alpha\beta}\tilde{Y}_C+o_p(1)\right)'\left(R_n^{t-n}\right)'\right)+o_p(1)\quad \text{(using lemma 2.3)}\\
&= vec\left(\frac{1}{n^{\frac{\alpha}{2}}}\sum_{t=k_n+k_n'}^{n}u_{0t}\left(\tilde{Y}_C\right)'\left(R_n^{t-n}C_{z\alpha\beta}\right)'\right)+o_p(1)
\end{aligned}$$

$$\begin{aligned}
&= \left(\frac{1}{n^{\frac{\alpha}{2}}}\sum_{t=k_n+k_n'}^{n}R_n^{-(n-t)}C_{z\alpha\beta}\otimes u_{0t}\right)\left(\tilde{Y}_C\right)+o_p(1)\\
&= \left(C_{z\alpha\beta}\otimes I_m\right)\left(\frac{1}{n^{\frac{\alpha}{2}}}\sum_{t=1}^{n}R_n^{-(n-t)}\otimes u_{0t}\right)\left(\tilde{Y}_C\right)+o_p(1)\quad \text{(putting back the front sum)}\\
&\Longrightarrow \left(C_{z\alpha\beta}\otimes I_m\right)\times MN\left(0,\int_0^\infty e^{-pC}\tilde{Y}_C\tilde{Y}_C'e^{-pC}dp\otimes\Omega_{00}\right),
\end{aligned}$$

where the last step comes from the same procedure as in equations (22)-(26) of Magdalinos and Phillips (2009a). Finally,

$$\begin{aligned}
vec\left(\frac{1}{n^{(\alpha\wedge\beta)}}\sum_{t=1}^{n}u_{0t}\tilde{z}_t'R_n^{-n}\right) &= \left(C\otimes I_m\right)vec\left(\frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum_{t=1}^{n}u_{0t}\psi_{nt}'R_n^{-n}\right)+o_p(1)\\
&\Longrightarrow \left(CC_{z\alpha\beta}\otimes I_m\right)\times MN\left(0,\int_0^\infty e^{-pC}Y_CY_C'e^{-pC}dp\otimes\Omega_{00}\right),
\end{aligned}$$

where we have used Y_C instead of its distributional copy $\tilde{Y}_C$ since we are concerned with weak convergence in the original probability space from this point onwards.

Part (2). From $\sum x_t\tilde{z}_t'=\sum x_tz_t'+n^{-\alpha}\sum_{t=1}^n x_t\psi_{nt}'C$, we have

$$\begin{aligned}
\frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum R_n^{-n}x_t\tilde{z}_t'R_n^{-n} &= \frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum R_n^{-n}x_tz_t'R_n^{-n}+\frac{1}{n^{2\alpha+(\alpha\wedge\beta)}}\sum_{t=1}^{n}R_n^{-n}x_t\psi_{nt}'R_n^{-n}C\\
&= \frac{1}{n^{2\alpha+(\alpha\wedge\beta)}}\sum_{t=1}^{n}R_n^{-n}x_t\psi_{nt}'R_n^{-n}C+o_p(1),
\end{aligned}$$

as in Part (1) and using the same sum splitting argument again with lemma 2.2 and 2.3,

$$\begin{aligned}
\frac{1}{n^{2\alpha+(\alpha\wedge\beta)}}\sum_{t=1}^{n} R_n^{-n}x_t\psi_{nt}'R_n^{-n} &= \frac{1}{n^{2\alpha+(\alpha\wedge\beta)}}\sum_{t=k_n+k_n'}^{n} R_n^{-n}x_t\psi_{nt}'R_n^{-n} + o_p(1) \\
&= \frac{1}{n^{\alpha}}\sum_{t=k_n+k_n'}^{n} R_n^{-(n-t)}\left(\frac{R_n^{-t}}{n^{\frac{\alpha}{2}}}x_t\right)\left(\frac{R_n^{-t}}{n^{\frac{\alpha}{2}+(\alpha\wedge\beta)}}\psi_{nt}\right)' R_n^{-(n-t)} + o_p(1) \\
&= \frac{1}{n^{\alpha}}\sum_{t=k_n+k_n'}^{n} R_n^{-(n-t)}\left(\tilde{Y}_C\right)\left(\tilde{Y}_C\right)' R_n^{-(n-t)}C_{z\alpha\beta} + o_p(1)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^{\alpha}}\sum_{t=1}^{n} R_n^{-(n-t)}\left(\tilde{Y}_C\right)\left(\tilde{Y}_C\right)' R_n^{-(n-t)}C_{z\alpha\beta} + o_p(1) \\
&\Longrightarrow \int_0^{\infty} e^{-pC}Y_CY_C'e^{-pC}dpC_{z\alpha\beta}.
\end{aligned}$$

It follows that

$$\frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum R_n^{-n}x_t\tilde{z}_t'R_n^{-n} \Longrightarrow \int_0^{\infty} e^{-pC}Y_CY_C'e^{-pC}dpCC_{z\alpha\beta}.$$

■

Proof of Theorem 2.2. We have the following limit theory for the IVX estimator $\hat{A}_n = \left(Y'\tilde{Z}\right)\left(X'\tilde{Z}\right)^{-1}$ without the bias correction shown in (2.11). Joint convergence of the numerator and denominator in lemma 2.4 parts (1) and (2) is established as in Magdalinos and Phillips (2009a, Proof of Theorem 4.1). Using these results with $\hat{A}_n - A = \left(U_0'\tilde{Z}\right)\left(X'\tilde{Z}\right)^{-1}$ we have

$$\begin{aligned}
vec\left[n^{\alpha}\left(\hat{A}_n - A\right)R_n^n\right] &= \left[\left\{\left(\frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum R_n^{-n}x_t\tilde{z}_t'R_n^{-n}\right)'\right\}^{-1}\otimes I_m\right] vec\left(\frac{1}{n^{(\alpha\wedge\beta)}}\sum_{t=1}^{n} u_{0t}\tilde{z}_t'R_n^{-n}\right) \\
&\Longrightarrow \left[\left(\left\{\tilde{\Psi}_{yy}\times\left(CC_{z\alpha\beta}\right)\right\}'\right)^{-1}\otimes I_m\right] \\
&\quad \times\left(CC_{z\alpha\beta}\otimes I_m\right)\times MN\left(0,\tilde{\Psi}_{yy}\otimes\Omega_{00}\right) \\
&\equiv MN\left(0,\left(\tilde{\Psi}_{yy}\right)^{-1}\otimes\Omega_{00}\right).
\end{aligned}$$

The bias corrected IVX estimator given in (2.11) is asymptotically equivalent to the uncorrected estimator $\hat{A}_n$ due to the signal strength of the $X'\tilde{Z}$ matrix, i.e.,

$$\hat{A}_n - \tilde{A}_n = n\widehat{\Delta}_{0x}(X'\tilde{Z})^{-1} = O_p\left(\frac{n}{n^{\alpha+(\alpha\wedge\beta)}\left\|R_n\right\|^{2n}}\right) = o_p(1),$$

and

$$
\begin{aligned}
n^{\alpha}\left(\tilde{A}_n - A\right) R_n^n &= n^{\alpha}\left(\hat{A}_n - A\right) R_n^n + O_p\left(\frac{n^{1-(\alpha\wedge\beta)}}{\|R_n\|^n}\right) \\
&= n^{\alpha}\left(\hat{A}_n - A\right) R_n^n + o_p(1).
\end{aligned}
$$

As a result, the same limit theory holds regardless of the bias correction, proving theorem 2.2. ■

The following lemma helps in characterizing the variance matrix asymptotics for (2.10).

Lemma 6.2 $\frac{1}{n^{2(\alpha\wedge\beta)}}\sum R_n^{-n}\tilde{z}_t\tilde{z}_t' R_n^{-n} \Longrightarrow CC_{z\alpha\beta}\tilde{\Psi}_{yy}CC_{z\alpha\beta}$.

Proof. We have the decomposition of the sample moment matrix

$$
\begin{aligned}
\sum \tilde{z}_t\tilde{z}_t' &= \sum\left(z_t + \frac{C}{n^{\alpha}}\psi_{nt}\right)\left(z_t + \frac{C}{n^{\alpha}}\psi_{nt}\right)' \\
&= \sum z_t z_t' + \frac{C}{n^{\alpha}}\sum \psi_{nt}z_t' + \frac{1}{n^{\alpha}}\sum z_t\psi_{nt}'C + \frac{1}{n^{2\alpha}}\sum C\psi_{nt}\psi_{nt}'C'.
\end{aligned}
$$

By the same methods used in lemma 2.3 and 2.4, it is straightforward to show that

$$
\begin{aligned}
&\frac{1}{n^{2\alpha+2(\alpha\wedge\beta)}}\sum_{t=1}^{n} R_n^{-n}\psi_{nt}\psi_{nt}'R_n^{-n} \\
=\ & \frac{1}{n^{\alpha}}\sum_{t=1}^{n} R_n^{-(n-t)}\left(\frac{1}{n^{\frac{\alpha}{2}+(\alpha\wedge\beta)}}R_n^{-t}\psi_{nt}\right)\left(\frac{1}{n^{\frac{\alpha}{2}+(\alpha\wedge\beta)}}R_n^{-t}\psi_{nt}\right)' R_n^{-(n-t)} \\
=\ & \frac{1}{n^{\alpha}}\sum_{t=k_n+k_n'}^{n} R_n^{-(n-t)}\left(\frac{1}{n^{\frac{\alpha}{2}+(\alpha\wedge\beta)}}R_n^{-t}\psi_{nt}\right)\left(\frac{1}{n^{\frac{\alpha}{2}+(\alpha\wedge\beta)}}R_n^{-t}\psi_{nt}\right)' R_n^{-(n-t)} + o_p(1) \\
=\ & \frac{1}{n^{\alpha}}\sum_{t=k_n+k_n'}^{n} R_n^{-(n-t)}\left(C_{z\alpha\beta}\tilde{Y}_C\right)\left(C_{z\alpha\beta}\tilde{Y}_C\right)' R_n^{-(n-t)} + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
&= (C_{z\alpha\beta})\left(\frac{1}{n^{\alpha}}\sum_{t=k_n+k_n'}^{n} R_n^{-(n-t)}\left(\tilde{Y}_C\right)\left(\tilde{Y}_C\right)' R_n^{-(n-t)}\right)(C_{z\alpha\beta}) + o_p(1) \\
&= (C_{z\alpha\beta})\left(\frac{1}{n^{\alpha}}\sum_{t=1}^{n} R_n^{-(n-t)}\left(\tilde{Y}_C\right)\left(\tilde{Y}_C\right)' R_n^{-(n-t)}\right)(C_{z\alpha\beta}) + o_p(1) \\
&\Longrightarrow (C_{z\alpha\beta})\int_0^{\infty} e^{-pC}Y_C Y_C' e^{-pC}dp\,(C_{z\alpha\beta}).
\end{aligned}
$$

This term clearly dominates all the other terms since z_t is mildly integrated. Hence

$$\begin{aligned}
\frac{1}{n^{2(\alpha\wedge\beta)}}\sum R_n^{-n}\tilde{z}_t\tilde{z}_t' R_n^{-n} &= C\left(\frac{1}{n^{2\alpha+2(\alpha\wedge\beta)}}\sum_{t=1}^{n} R_n^{-n}\psi_{nt}\psi_{nt}' R_n^{-n}\right)C + o_p(1) \\
&\Longrightarrow (CC_{z\alpha\beta})\int_0^\infty e^{-pC}Y_C Y_C' e^{-pC}dp\,(CC_{z\alpha\beta}) \\
&\equiv CC_{z\alpha\beta}\tilde{\Psi}_{yy}CC_{z\alpha\beta}.
\end{aligned}$$

■

The robust chi-square limit theory of the self-normalized IVX estimator follows and is given in the next proof.

Proof of Theorem 2.3. We have the following limit theory for the variance estimator,

$$\begin{aligned}
& n^{2\alpha}\left(R_n^n\otimes I_m\right)\left[\left(X'P_{\widetilde{Z}}X\right)^{-1}\otimes\hat{\Omega}_{00}\right]\left(R_n^n\otimes I_m\right) \\
=& \left[\frac{1}{n^{2\alpha}}R_n^{-n}\left\{X'\tilde{Z}\left(\tilde{Z}'\tilde{Z}\right)^{-1}\tilde{Z}'X\right\}R_n^{-n}\right]^{-1}\otimes\hat{\Omega}_{00} \\
=& \left[\left(\frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum R_n^{-n}x_t\tilde{z}_t' R_n^{-n}\right)\left(\frac{1}{n^{2(\alpha\wedge\beta)}}\sum R_n^{-n}\tilde{z}_t\tilde{z}_t' R_n^{-n}\right)^{-1}\left(\frac{1}{n^{\alpha+(\alpha\wedge\beta)}}\sum R_n^{-n}x_t\tilde{z}_t' R_n^{-n}\right)'\right]^{-1}\otimes\hat{\Omega}_{00} \\
\Longrightarrow& \left[\left(\tilde{\Psi}_{yy}\left(CC_{z\alpha\beta}\right)\right)\left(\left(CC_{z\alpha\beta}\right)\tilde{\Psi}_{yy}\left(CC_{z\alpha\beta}\right)\right)^{-1}\left(\tilde{\Psi}_{yy}\left(CC_{z\alpha\beta}\right)\right)'\right]^{-1}\otimes\Omega_{00} \\
\equiv& \tilde{\Psi}_{yy}^{-1}\otimes\Omega_{00},
\end{aligned}$$

and again the weak convergence is joint with that of the estimator components. Hence

$$\begin{aligned}
& vec\left(\tilde{A}_n - A\right)'\left[\left(X'P_{\widetilde{Z}}X\right)^{-1}\otimes\hat{\Omega}_{00}\right]^{-1}vec\left(\tilde{A}_n - A\right) \\
=& vec\left(\tilde{A}_n - A\right)'\left(R_n^n\otimes I_m\right)\left(R_n^n\otimes I_m\right)^{-1}\left[\left(X'P_{\widetilde{Z}}X\right)^{-1}\otimes\hat{\Omega}_{00}\right]^{-1}\left(R_n^n\otimes I_m\right)^{-1}\left(R_n^n\otimes I_m\right)vec\left(\tilde{A}_n - A\right) \\
=& \left\{vec\left[n^\alpha\left(\tilde{A}_n - A\right)R_n^n\right]\right\}'\left[n^{2\alpha}\left(R_n^n\otimes I_m\right)\left[\left(X'P_{\widetilde{Z}}X\right)^{-1}\otimes\hat{\Omega}_{00}\right]\left(R_n^n\otimes I_m\right)\right]^{-1}vec\left[n^\alpha\left(\tilde{A}_n - A\right)R_n^n\right] \\
\Longrightarrow& \chi^2\left(mK\right),
\end{aligned}$$

proving theorem 2.3. ■

6.3 The Case of Mixed Roots: (I6)

We collect together the proofs for mixed roots case - I(6).

Proof of Lemma 3.1. We only provide the proof for the case $(\alpha_1\wedge\beta)=\beta$. The other case $(\alpha_1\wedge\beta)=\alpha_1$ is proved in the exactly same way.

Note that $\theta_n^{-k_n} = o(1)$ (from lemma 2.1) and $n^{\alpha_2}(\theta_n - 1) = c_2$, so that

$$
\begin{aligned}
E\left[Y_{c_2n}\left(\sum_{t=1}^{n}\xi_{n1t}\right)\right] &= E\left[\left(\frac{1}{n^{\frac{\alpha_2}{2}}}\sum_{j=1}^{k_n}\theta_n^{-j}F_{x2}(1)\varepsilon_j\right)\left(\sum_{t=1}^{n}\frac{1}{n^{\frac{1+\beta}{2}}}z_{1t-1}\varepsilon_t'F_0(1)'\right)\right] \\
&= F_{x2}(1)\Sigma F_0(1)'\frac{1}{n^{\frac{1+\beta+\alpha_2}{2}}}\sum_{j=1}^{k_n}\theta_n^{-j}E\left[z_{1j-1}\right] \leq \sup_{1\leq t\leq n}\left\|\frac{z_{1t}}{n^{\beta/2}}\right\|_{L_2}\Omega_{20}\frac{1}{n^{\frac{1}{2}+\frac{\alpha_2}{2}}}\frac{1-\theta_n^{-k_n}}{1-\theta_n^{-1}}\theta_n^{-1} \\
&= O\left(\frac{n^{\frac{\alpha_2}{2}}}{n^{\frac{1}{2}}}\frac{1}{n^{\alpha_2}(\theta_n-1)}\right) = o(1),
\end{aligned}
$$

using the law of iterated expectation for the second equality and the fact that

$$
\sup_{1\leq t\leq n}\left\|\frac{z_{1t}}{n^{\beta/2}}\right\|_{L_2} < \infty, \tag{6.3}
$$

(from MP) for the third inequality.

Similarly,

$$
\begin{aligned}
E\left[Y_{c_2n}\left(\sum_{t=1}^{n}\xi_{n2t}\right)\right] &= E\left[\left(\frac{1}{n^{\frac{\alpha_2}{2}}}\sum_{j=1}^{k_n}\theta_n^{-j}F_{x2}(1)\varepsilon_j\right)\left(\sum_{t=1}^{n}\frac{1}{\sqrt{n}}\varepsilon_t'F_{x1}(1)'\right)\right] = \Omega_{21}\frac{1}{n^{\frac{1+\alpha_2}{2}}}\sum_{j=1}^{k_n}\theta_n^{-j} \\
&= O\left(\frac{n^{\frac{\alpha_2}{2}}}{n^{\frac{1}{2}}}\frac{1}{n^{\alpha_2}(\theta_n-1)}\right) = o(1).
\end{aligned}
$$

The covariance $E\left[Y_{c_0n}\left(\sum_{t=1}^{n}\xi_{nt}\right)\right] \to 0_{2\times 1}$ can be shown in the exactly same way, thereby confirming asymptotic independence since limit distributions are all Gaussian. ■

Proof of Lemma 3.2. We first show that

$$
\begin{aligned}
\theta_n^{-n}n^{-(1+(\alpha_1\wedge\beta))/2-\alpha_2}\sum_{t=1}^{n}\tilde{z}_{1t}x_{2t} &= o_p(1), &(6.4)\\
\theta_n^{-n}n^{-(1+(\alpha_1\wedge\beta))/2-(\alpha_2\wedge\beta)}\sum_{t=1}^{n}\tilde{z}_{2t}x_{1t} &= o_p(1), &(6.5)
\end{aligned}
$$

and so the off-diagonal entries are asymptotically negligible. To prove (6.4), we consider (i) $\beta < \alpha_1$ and (ii) $\alpha_1 \leq \beta$ seperately.

(i) $\beta < \alpha_1$: in this case $(\alpha_1 \wedge \beta) = \beta$ and we have

$$
\frac{1}{n^{\frac{1+\beta}{2}+\alpha_2}\theta_n^n}\sum_{t=1}^{n}\tilde{z}_{1t}x_{2t} = \frac{1}{n^{\frac{1+\beta}{2}+\alpha_2}\theta_n^n}\sum_{t=1}^{n}z_{1t}x_{2t} + \frac{c_1}{n^{\frac{1+\beta}{2}}\theta_n^n n^{\alpha_1+\alpha_2}}\sum_{t=1}^{n}\psi_{1nt}x_{2t}, \tag{6.6}
$$

and then

$$\begin{aligned}\left\|\frac{1}{n^{\frac{1+\beta}{2}+\alpha_2}\theta_n^n}\sum_{t=1}^n z_{1t}x_{2t}\right\|_{L_1} &= \left\|\frac{1}{n^{\frac{1}{2}+\frac{\alpha_2}{2}}\theta_n^n}\sum_{t=1}^n\left(\frac{z_{1t}}{n^{\beta/2}}\right)\left(\frac{x_{2t}}{n^{\frac{\alpha_2}{2}}\theta_n^t}\right)\theta_n^t\right\|_{L_1}\\ &\leq \sup_{1\leq t\leq n}\left\|\frac{z_{1t}}{n^{\beta/2}}\right\|_{L_2}\sup_{1\leq t\leq n}\left\|\frac{x_{2t}}{n^{\frac{\alpha_2}{2}}\theta_n^t}\right\|_{L_2}\frac{1}{n^{\frac{1}{2}+\frac{\alpha_2}{2}}\theta_n^n}\sum_{t=1}^n\theta_n^t \text{ and using (2.12) and (6.3),}\\ &= O\left(\frac{\theta_n^n-1}{\theta_n^n}\frac{1}{n^{\alpha_2}(\theta_n-1)}\frac{n^{\frac{\alpha_2}{2}}}{n^{\frac{1}{2}}}\right)=o(1),\end{aligned}$$

leading to $\theta_n^{-n}n^{-(1+\beta)/2-\alpha_2}\sum_{t=1}^n z_{1t}x_{2t}=o_p(1)$. Using $\sup_{1\leq t\leq n}\left\|\frac{\psi_{1nt}}{n^{\frac{\alpha_1}{2}+\beta}}\right\|=O_p(1)$ from MP, the negligibility of the second component of (6.6),viz., $\frac{c_1}{n^{\frac{1+\beta}{2}}\theta_n^n n^{\alpha_1+\alpha_2}}\sum_{t=1}^n\psi_{1nt}x_{2t}=o_p(1)$, can be shown in a similar way.

(ii) $\alpha_1\leq\beta$: in this case $(\alpha_1\wedge\beta)=\alpha_1$. From equation (23) from MP, we use

$$\tilde{z}_{1t}=x_{1t}+\frac{c_{z1}}{\beta}\psi_{1nt}$$

so

$$\frac{1}{n^{\frac{1+\alpha_1}{2}+\alpha_2}\theta_n^n}\sum_{t=1}^n\tilde{z}_{1t}x_{2t}=\frac{1}{n^{\frac{1+\alpha_1}{2}+\alpha_2}\theta_n^n}\sum_{t=1}^n x_{1t}x_{2t}+\frac{c_1}{n^{\frac{1+\alpha_1}{2}}\theta_n^n n^{\beta+\alpha_2}}\sum_{t=1}^n\psi_{1nt}x_{2t}$$

and then

$$\begin{aligned}\left\|\frac{1}{n^{\frac{1+\alpha_1}{2}+\alpha_2}\theta_n^n}\sum_{t=1}^n x_{1t}x_{2t}\right\|_{L_1} &= \left\|\frac{1}{n^{\frac{1}{2}+\frac{\alpha_2}{2}}\theta_n^n}\sum_{t=1}^n\left(\frac{x_{1t}}{n^{\alpha_1/2}}\right)\left(\frac{x_{2t}}{n^{\frac{\alpha_2}{2}}\theta_n^t}\right)\theta_n^t\right\|_{L_1}\\ &\leq \sup_{1\leq t\leq n}\left\|\frac{x_{1t}}{n^{\alpha_1/2}}\right\|_{L_2}\sup_{1\leq t\leq n}\left\|\frac{x_{2t}}{n^{\frac{\alpha_2}{2}}\theta_n^t}\right\|_{L_2}\frac{1}{n^{\frac{1}{2}+\frac{\alpha_2}{2}}\theta_n^n}\sum_{t=1}^n\theta_n^t\\ &= O\left(\frac{\theta_n^n-1}{\theta_n^n}\frac{1}{n^{\alpha_2}(\theta_n-1)}\frac{n^{\frac{\alpha_2}{2}}}{n^{\frac{1}{2}}}\right)=o(1).\end{aligned}$$

Therefore, $\frac{1}{n^{\frac{1+(\alpha_1\wedge\beta)}{2}+\alpha_2}\theta_n^n}\sum_{t=1}^n\tilde{z}_{1t}x_{2t}=o_p(1)$, as required

For (6.5), we also consider (i) $\beta<\alpha_1$ and (ii) $\alpha_1\leq\beta$ seperately.

(i) $\beta<\alpha_1$: in this case $(\alpha_1\wedge\beta)=\beta$ and start by noting that

$$\frac{1}{n^{\frac{1+\beta}{2}+(\alpha_2\wedge\beta)}\theta_n^n}\sum_{t=1}^n\tilde{z}_{2t}x_{1t}=\frac{1}{n^{\frac{1+\beta}{2}+(\alpha_2\wedge\beta)}\theta_n^n}\sum_{t=1}^n z_{2t}x_{1t}+\frac{c_2}{n^{\frac{1+\beta}{2}+(\alpha_2\wedge\beta)}\theta_n^n n^{\alpha_2}}\sum_{t=1}^n\psi_{2nt}x_{1t}. \qquad (6.7)$$

The first component of (6.7) $\frac{1}{n^{\frac{1+3\beta}{2}}\theta_n^n}\sum_{t=1}^n z_{2t}x_{1t}=o_p(1)$ because for mildly integrated z_{2t} and at most local to unity x_{1t} the sum does not require the additional θ_n^n normalization and hence is dominated by the exponential growth of $\theta_n^n \sim \exp\left(c_2 n^{1-\alpha_2}\right)$, as earlier. It follows that

$$\frac{1}{n^{\frac{1+\beta}{2}+(\alpha_2\wedge\beta)}\theta_n^n}\sum_{t=1}^n \tilde{z}_{2t}x_{1t} = \frac{c_2}{n^{\frac{1+\beta}{2}+(\alpha_2\wedge\beta)+\alpha_2}\theta_n^n}\sum_{t=1}^n \psi_{2nt}x_{1t} + o_p(1).$$

To obtain a bound for $\sum_{t=1}^n \psi_{2nt}x_{1t}$, note that

$$\begin{aligned}\left\|\sum_{t=1}^n \psi_{2nt}x_{1t}\right\|_{L_1} &= n^{\frac{\alpha_1}{2}+\frac{\alpha_2}{2}+(\alpha_2\wedge\beta)}\left\|\sum_{t=1}^n \theta_n^t\left(\frac{\psi_{2nt}}{n^{\frac{\alpha_2}{2}+(\alpha_2\wedge\beta)}\theta_n^t}\right)\left(\frac{x_{1t}}{n^{\frac{\alpha_1}{2}}}\right)\right\|_{L_1} \\ &\leq \left(\sup_t\left\|\frac{\psi_{2nt}}{n^{\frac{\alpha_2}{2}+(\alpha_2\wedge\beta)}\theta_n^t}\right\|_{L_2}\right)\left(\sup_t\left\|\frac{x_{1t}}{n^{\frac{\alpha_1}{2}}}\right\|_{L_2}\right)n^{\frac{\alpha_1}{2}+\frac{\alpha_2}{2}+(\alpha_2\wedge\beta)}\sum_{t=1}^n\theta_n^t\end{aligned}$$

which is $O\left(n^{\frac{\alpha_1}{2}+\frac{3\alpha_2}{2}+(\alpha_2\wedge\beta)}\theta_n^n\right)$. Therefore,

$$\left\|\frac{1}{n^{\frac{1+\beta}{2}+(\alpha_2\wedge\beta)+\alpha_2}\theta_n^n}\sum_{t=1}^n \psi_{2nt}x_{1t}\right\|_{L_1} \leq O\left(n^{\frac{(\alpha_1-1)}{2}+\frac{(\alpha_2-\beta)}{2}}\right). \tag{6.8}$$

(ii) $\alpha_1 \leq \beta$: in this case $(\alpha_1\wedge\beta) = \alpha_1$ and

$$\frac{1}{n^{\frac{1+\alpha_1}{2}+(\alpha_2\wedge\beta)}\theta_n^n}\sum_{t=1}^n \tilde{z}_{2t}x_{1t} = \frac{1}{n^{\frac{1+\alpha_1}{2}+(\alpha_2\wedge\beta)}\theta_n^n}\sum_{t=1}^n z_{2t}x_{1t} + \frac{c_2}{n^{\frac{1+\alpha_1}{2}+(\alpha_2\wedge\beta)}\theta_n^n n^{\alpha_2}}\sum_{t=1}^n \psi_{2nt}x_{1t}$$

then by the same procedure

$$\left\|\frac{1}{n^{\frac{1+\alpha_1}{2}+(\alpha_2\wedge\beta)+\alpha_2}\theta_n^n}\sum_{t=1}^n \psi_{2nt}x_{1t}\right\|_{L_1} \leq O\left(n^{\frac{(\alpha_2-1)}{2}}\right) = o(1)$$

since $\alpha_2 - 1 < 0$.

There are a range of rate conditions that will ensure the negligibility of the term (6.8), thereby producing a diagonal limit of the moment matrix $\tilde{Z}'X$ upon standardization. In general, we need the overall condition

$$\alpha_1 + \alpha_2 < 1 + \beta, \tag{6.9}$$

but since $\alpha_1 \in (0,1]$ it is sufficient for (6.9) that

$$\alpha_2 < \beta, \tag{6.10}$$

which requires that β be large enough in relation to α_2.

Using (6.4) and (6.5) we have

$$\begin{aligned}
D_n^{-1}\tilde{Z}'XC_n^{-1} &= D_n^{-1}\begin{bmatrix} \sum_{t=1}^n \tilde{z}_{1t}x_{1t} & \sum_{t=1}^n \tilde{z}_{1t}x_{2t} \\ \sum_{t=1}^n \tilde{z}_{2t}x_{1t} & \sum_{t=1}^n \tilde{z}_{2t}x_{2t} \end{bmatrix} C_n^{-1} \\
&= \begin{bmatrix} \frac{1}{n^{1+(\alpha_1\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{1t}x_{1t} & \frac{1}{\theta_n^n n^{\frac{1+(\alpha_1\wedge\beta)}{2}+\alpha_2}}\sum_{t=1}^n \tilde{z}_{1t}x_{2t} \\ \frac{1}{\theta_n^n n^{\frac{1+(\alpha_1\wedge\beta)}{2}} n^{(\alpha_2\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{2t}x_{1t} & \frac{1}{\theta_n^{2n} n^{\alpha_2+(\alpha_2\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{2t}x_{2t} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n^{1+(\alpha_1\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{1t}x_{1t} & o_p(1) \\ o_p(1) & \frac{1}{\theta_n^{2n} n^{\alpha_2+(\alpha_2\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{2t}x_{2t} \end{bmatrix}.
\end{aligned}$$

For $n^{-1-(\alpha_1\wedge\beta)}\sum_{t=1}^n \tilde{z}_{1t}x_{1t}$ we consider (i) $\beta < \alpha_1$ and (ii) $\alpha_1 \leq \beta$ seperately.

(i) $\beta < \alpha_1$: in this case $(\alpha_1 \wedge \beta) = \beta$ and from Theorem 3.4 of MP,

$$\frac{1}{n^{1+\beta}}\sum_{t=1}^n \tilde{z}_{1t}x_{1t} \Longrightarrow \begin{cases} \frac{1}{-c_{z1}}\left\{\Omega_{11} + \int_0^1 B_1 dB_1\right\} & \text{if } x_{1t} \text{ is unit root - I(1)} \\ \frac{1}{-c_{z1}}\left\{\Omega_{11} + \int_0^1 J_x^{c1} dJ_x^{c1}\right\} & \text{if } x_{1t} \text{ is local to unity - (I2) or (I4)} \\ \frac{1}{-2c_{z1}}\Omega_{11} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \beta < \alpha_1 \end{cases}$$

(ii) $\alpha_1 \leq \beta$: in this case $(\alpha_1 \wedge \beta) = \alpha_1$ and from Lemma 3.5 and 3.6 from MP,

$$\frac{1}{n^{1+\alpha_1}}\sum_{t=1}^n \tilde{z}_{1t}x_{1t} \longrightarrow_p \begin{cases} \frac{1}{-2(c_1+c_{z1})}\Omega_{11} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \beta = \alpha_1 \\ -\frac{1}{2c_1}\Omega_{11} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \alpha_1 < \beta \end{cases}$$

and by defining

$$\tilde{\Psi}_{11} = \begin{cases} \frac{1}{-c_{z1}}\left\{\Omega_{11} + \int_0^1 B_1 dB_1\right\} & \text{if } x_{1t} \text{ is unit root : I(1),} \\ \frac{1}{-c_{z1}}\left\{\Omega_{11} + \int_0^1 J_x^{c1} dJ_x^{c1}\right\} & \text{if } x_{1t} \text{ is local to unity : (I2) or (I4),} \\ \frac{1}{-2c_{z1}}\Omega_{11} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \beta < \alpha_1 \\ \frac{1}{-2(c_1+c_{z1})}\Omega_{11} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \beta = \alpha_1 \\ \frac{1}{-2c_1}\Omega_{11} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \alpha_1 < \beta \end{cases}. \tag{6.11}$$

we have

$$\frac{1}{n^{1+(\alpha_1\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{1t}x_{1t} \Longrightarrow \tilde{\Psi}_{11}$$

where the weak convergence to a constant (when x_{1t} is mildly integrated) is equivalent to the convergence in probability. Together with $\frac{1}{\theta_n^{2n} n^{\alpha_2+(\alpha_2\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{2t}x_{2t} \Longrightarrow c_{z_2\alpha\beta}c_2\tilde{\Psi}_{22}$ from lemma 2.4-(2), the stated result is proved. ■

Proof of Theorem 3.1. We write the scaled estimation error as

$$C_n(\tilde{a} - a) = \left(D_n^{-1}\begin{bmatrix} \sum_{t=1}^n \tilde{z}_{1t}x_{1t} & \sum_{t=1}^n \tilde{z}_{1t}x_{2t} \\ \sum_{t=1}^n \tilde{z}_{2t}x_{1t} & \sum_{t=1}^n \tilde{z}_{2t}x_{2t} \end{bmatrix} C_n^{-1}\right)^{-1} D_n^{-1}\begin{bmatrix} \sum_{t=1}^n \tilde{z}_{1t}u_{0t} - n\hat{\Delta}_{0x1} \\ \sum_{t=1}^n \tilde{z}_{2t}u_{0t} - n\hat{\Delta}_{0x2} \end{bmatrix},$$

and using Lemma 3.2,

$$\begin{bmatrix} n^{\frac{1+(\alpha_1\wedge\beta)}{2}}(\tilde{a}_1-a_1) \\ \theta_n^n n^{\alpha_2}(\tilde{a}_2-a_2) \end{bmatrix} = \begin{bmatrix} \left(\frac{1}{n^{1+(\alpha_1\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{1t}x_{1t}\right)^{-1} \frac{1}{n^{\frac{1+(\alpha_1\wedge\beta)}{2}}}\sum_{t=1}^n\left(\tilde{z}_{1t}u_{0t}-\hat{\Delta}_{0x1}\right) \\ \left(\frac{1}{\theta_n^{2n}n^{\alpha_2+(\alpha_2\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{2t}x_{2t}\right)^{-1}\frac{1}{\theta_n^n n^{(\alpha_2\wedge\beta)}}\sum_{t=1}^n\left(\tilde{z}_{2t}u_{0t}-\hat{\Delta}_{0x2}\right)\end{bmatrix}+o_p(1)$$

The limit theory of $n^{\frac{1+(\alpha_1\wedge\beta)}{2}}(\tilde{a}_1-a_1)$ is exactly a scalar version of Theorem 3.4 and 3.7 from MP so that

$$n^{\frac{1+(\alpha_1\wedge\beta)}{2}}(\tilde{a}_1-a_1)$$
$$\Longrightarrow \begin{cases} MN(0,\left\{\frac{1}{-c_{z1}}\left(\Omega_{11}+\int_0^1 B_1dB_1\right)\right\}^{-2}\left(\frac{1}{-2c_{z1}}\Omega_{11}\right)\Omega_{00})) & \text{if } x_{1t} \text{ is unit root - I(1)} \\ MN(0,\left\{\frac{1}{-c_{z1}}\left\{\Omega_{11}+\int_0^1 J_x^{c1}dJ_x^{c1}\right\}\right\}^{-2}\left(\frac{1}{-2c_{z1}}\Omega_{11}\right)\Omega_{00})) & \text{if } x_{1t} \text{ is local to unity - (I2) or (I4)} \\ N(0,\left(\frac{1}{-2c_{z1}}\Omega_{11}\right)^{-1}\Omega_{00}) & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \beta<\alpha_1 \\ N(0,\frac{(c_1+c_{z1})^2}{c_1^2}\left(\frac{1}{-2c_1}\Omega_{11}\right)^{-1}\Omega_{00}) & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \beta=\alpha_1 \\ N(0,\left(\frac{1}{-2c_1}\Omega_{11}\right)^{-1}\Omega_{00}) & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \alpha_1<\beta \end{cases}$$

By defining

$$\tilde{\Phi}_{11}^{-1} = \begin{cases} \left\{\frac{1}{-c_{z1}}\left(\Omega_{11}+\int_0^1 B_1dB_1\right)\right\}^{-2}\left(\frac{1}{-2c_{z1}}\Omega_{11}\right) & \text{if } x_{1t} \text{ is unit root - I(1)} \\ \left\{\frac{1}{-c_{z1}}\left\{\Omega_{11}+\int_0^1 J_x^{c1}dJ_x^{c1}\right\}\right\}^{-2}\left(\frac{1}{-2c_{z1}}\Omega_{11}\right) & \text{if } x_{1t} \text{ is local to unity - (I2) or (I4)} \\ \left(\frac{1}{-2c_{z1}}\Omega_{11}\right)^{-1} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \beta<\alpha_1 \\ \frac{(c_1+c_{z1})^2}{c_1^2}\left(\frac{1}{-2c_1}\Omega_{11}\right)^{-1} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \beta=\alpha_1 \\ \left(\frac{1}{-2c_1}\Omega_{11}\right)^{-1} & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \alpha_1<\beta \end{cases} \quad (6.12)$$

we have

$$n^{\frac{1+(\alpha_1\wedge\beta)}{2}}(\tilde{a}_1-a_1) \Longrightarrow MN(0,\tilde{\Phi}_{11}^{-1}\Omega_{00}))\ .$$

From lemma 2.4-(1), we have

$$\frac{1}{\theta_n^n n^{(\alpha_2\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{2t}u_{0t} \Longrightarrow c_{z_2\alpha\beta}c_2 MN\left(0,\tilde{\Psi}_{22}\Omega_{00}\right).$$

and together with lemma 3.2 this leads to

$$\theta_n^n n^{\alpha_2}(\tilde{a}_2-a_2) \Longrightarrow MN\left(0,\tilde{\Psi}_{22}^{-1}\Omega_{00}\right),$$

which is a special case of theorem 2.2. Joint convergence and asymptotic independence follow from lemma 3.1, thereby completing the proof. ■

The same mechanism for variance estimation, as shown in the following lemma, now leads to nuisance parameter free inference in the corresponding self-normalized test statistics.

Lemma 6.3 $\left[C_n\left(X'P_{\tilde{Z}}X\right)^{-1}C_n\hat{\Omega}_{00}\right]^{-1} \Longrightarrow \left[\left[\begin{array}{cc} \tilde{\Phi}_{11}\Omega_{00}^{-1} & 0 \\ 0 & \tilde{\Psi}_{22}\Omega_{00}^{-1} \end{array}\right]\right].$

Proof. We show the limit matrix is diagonal, as in the earlier development. Note that

$$D_n^{-1}\tilde{Z}'\tilde{Z}D_n^{-1} = \left[\begin{array}{cc} \frac{1}{n^{1+(\alpha_1\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{1t}^2 & \frac{1}{n^{\frac{1+(\alpha_1\wedge\beta)}{2}}\theta_n^n n^{(\alpha_2\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{1t}\tilde{z}_{2t} \\ \frac{1}{n^{\frac{1+(\alpha_1\wedge\beta)}{2}}\theta_n^n n^{(\alpha_2\wedge\beta)}}\sum_{t=1}^n\sum_{t=1}^n \tilde{z}_{1t}\tilde{z}_{2t} & \frac{1}{\theta_n^{2n}n^{2(\alpha_2\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{2t}^2 \end{array}\right],$$

and the exponentially fast normalizer θ_n^n gives

$$\frac{1}{n^{\frac{1+(\alpha_1\wedge\beta)}{2}}\theta_n^n n^{(\alpha_2\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{1t}\tilde{z}_{2t} = \frac{c_2}{n^{\frac{1+(\alpha_1\wedge\beta)}{2}}\theta_n^n n^{(\alpha_2\wedge\beta)+\alpha_2}}\sum_{t=1}^n \tilde{z}_{1t}\psi_{n2t} + o_p(1)$$
$$= \frac{c_2}{n^{\frac{1+(\alpha_1\wedge\beta)}{2}}\theta_n^n n^{(\alpha_2\wedge\beta)+\alpha_2}}\sum_{t=1}^n z_{1t}\psi_{n2t} + \frac{c_1c_2}{n^{\frac{1+(\alpha_1\wedge\beta)}{2}}\theta_n^n n^{(\alpha_2\wedge\beta)+\alpha_1+\alpha_2}}\sum_{t=1}^n \psi_{n1t}\psi_{n2t} + o_p(1).$$

Using the same earlier argument, we have

$$\frac{c_2}{n^{\frac{1}{2}}\theta_n^n n^{\frac{\alpha_2}{2}}}\sum_{t=1}^n\left(\frac{z_{1t}}{n^{\frac{(\alpha_1\wedge\beta)}{2}}}\right)\left(\frac{\psi_{n2t}}{n^{\frac{\alpha_2}{2}+(\alpha_2\wedge\beta)}\theta_n^t}\right)\theta_n^t = O_p\left(\frac{n^{\frac{\alpha_2}{2}}}{n^{\frac{1}{2}}}\right) = o_p(1),$$

and similarly

$$\frac{c_1c_2}{n^{\frac{1+(\alpha_1\wedge\beta)}{2}}\theta_n^n n^{(\alpha_2\wedge\beta)+\alpha_1+\alpha_2}}\sum_{t=1}^n \psi_{n1t}\psi_{n2t} = o_p(1).$$

It is easy to show, using Lemma 3.1, 3.5 and 3.6 from MP, that

$$\frac{1}{n^{1+(\alpha_1\wedge\beta)}}\sum_{t=1}^n \tilde{z}_{1t}^2 \longrightarrow_p \tilde{\Xi}_{11} := \left\{\begin{array}{ll} \frac{1}{-2c_{z1}}\Omega_{11}, & \text{if } x_{1t} \text{ is (I1), (I2), (I4) or (I3) with } \beta<\alpha_1 \\ \frac{1}{-2(c_1+c_{z1})}\Omega_{11}, & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \beta=\alpha_1 \\ \frac{1}{-2c_1}\Omega_{11}, & \text{if } x_{1t} \text{ is mildly integrated - (I3) with } \alpha_1<\beta \end{array}\right. \tag{6.13}$$

hence, from lemma 6.2

$$D_n^{-1}\tilde{Z}'\tilde{Z}D_n^{-1} \Longrightarrow \left[\begin{array}{cc} \tilde{\Xi}_{11} & 0 \\ 0 & \left(c_{z_2\alpha\beta}c_2\right)^2\tilde{\Psi}_{22} \end{array}\right].$$

Combining with the result of lemma 3.2

$$D_n^{-1}\tilde{Z}'XC_n^{-1} \Longrightarrow \left[\begin{array}{cc} \tilde{\Psi}_{11} & 0 \\ 0 & c_{xz2}(\alpha,\beta)c_2\tilde{\Psi}_{22} \end{array}\right],$$

we have

$$
\begin{aligned}
& \left[C_n \left(X' P_{\tilde{Z}} X\right)^{-1} C_n \hat{\Omega}_{00}\right] \\
= & \left[\left\{C_n^{-1} X' \tilde{Z} \left(\tilde{Z}' \tilde{Z}\right)^{-1} \tilde{Z}' X C_n^{-1}\right\}\right]^{-1} \hat{\Omega}_{00} \\
= & \left[\left\{C_n^{-1} X' \tilde{Z} D_n^{-1} \left(D_n^{-1} \tilde{Z}' \tilde{Z} D_n^{-1}\right)^{-1} D_n^{-1} \tilde{Z}' X C_n^{-1}\right\}\right]^{-1} \hat{\Omega}_{00} \\
\Longrightarrow & \left[\left\{\begin{bmatrix} \tilde{\Psi}_{11} & 0 \\ 0 & c_{z_2 \alpha\beta} c_2 \tilde{\Psi}_{22} \end{bmatrix} \begin{bmatrix} \tilde{\Xi}_{11} & 0 \\ 0 & \left(c_{z_2 \alpha\beta} c_2\right)^2 \tilde{\Psi}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\Psi}_{11} & 0 \\ 0 & c_{z_2 \alpha\beta} c_2 \tilde{\Psi}_{22} \end{bmatrix}\right\}\right]^{-1} \Omega_{00} \\
\equiv & \begin{bmatrix} \tilde{\Psi}_{11}^2 \tilde{\Xi}_{11}^{-1} & 0 \\ 0 & \tilde{\Psi}_{22} \end{bmatrix}^{-1} \Omega_{00}.
\end{aligned}
$$

Using the definition of $\tilde{\Psi}_{11}$, $\tilde{\Phi}_{11}$ and $\tilde{\Xi}_{11}$ from (6.11), (6.12) and (6.13), we can easily check

$$\tilde{\Psi}_{11}^2 \tilde{\Xi}_{11}^{-1} = \tilde{\Phi}_{11}.$$

Thus

$$\left[C_n \left(X' P_{\tilde{Z}} X\right)^{-1} C_n \hat{\Omega}_{00}\right] \Longrightarrow \begin{bmatrix} \tilde{\Phi}_{11} & 0 \\ 0 & \tilde{\Psi}_{22} \end{bmatrix}^{-1} \Omega_{00}$$

and then

$$\left[C_n \left(X' P_{\tilde{Z}} X\right)^{-1} C_n \hat{\Omega}_{00}\right]^{-1} \Longrightarrow \begin{bmatrix} \tilde{\Phi}_{11} \Omega_{00}^{-1} & 0 \\ 0 & \tilde{\Psi}_{22} \Omega_{00}^{-1} \end{bmatrix},$$

as stated. ■

The final result on robust pivotal chi-square limit theory now follows.

Proof of Theorem 3.2. Using Lemma 3.2, Theorem 3.1 and lemma 6.3, we have directly by continuous mapping

$$
\begin{aligned}
& \left(\tilde{a} - a_0\right)' \left[\left(X' P_{\tilde{Z}} X\right)^{-1} \hat{\Omega}_{00}\right]^{-1} \left(\tilde{a} - a_0\right) \\
= & \left(\tilde{a} - a_0\right)' C_n \left[C_n \left(X' P_{\tilde{Z}} X\right)^{-1} C_n \hat{\Omega}_{00}\right]^{-1} C_n \left(\tilde{a} - a\right) \\
\Longrightarrow & \chi^2(2).
\end{aligned}
$$

■

7 References

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

Gonzalo, J. and J-Y Pitarakis (2012). "Regime specific predictability in predictive regressions." *Journal of Business and Economic Statistics*, 30(2), 229-241.

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

Homm, U., and Breitung, J., (2012), Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics*, 10(1):198-231.

Jansson, M and M. Moreira (2006). "Optimal inference in regression models with nearly integrated regressors," *Econometrica*, 74(3), 681-715.

Kostakis, A., A. Magdalinos and M. Stamatogiannis (2014). "Robust econometric inference for stock return predictability," *Unpublished Manuscript,* University of Southampton.

Lee, J.H. (2014). "Predictive quantile regression with persistent covariates," *Unpublished Manuscript,* University of Washington.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* New York: Springer.

Magdalinos, T. and P. C. B Phillips (2009a). "Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors". *Econometric Theory*, 25, 482-526.

Magdalinos, T. and P. C. B Phillips (2009b). "Econometric inference in the vicinity of unity". *CoFie Working Paper (7),* Singapore Management University.

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

Phillips, P. C. B. (2007). "Unit root periodogram regression," *Journal of Econometrics* 138, 104–124.

Phillips, P. C. B. and J.H. Lee (2014). "Limit Theory VARs with Mixed Roots Near Unity," *Econometric Reviews,* forthcoming.

Phillips, P. C. B, and J.H. Lee (2013). "Predictive regression under various degrees of persistence and robust long-horizon regression". *Journal of Econometrics*, 177(2), 250-264.

Phillips, P. C. B., T. Magdalinos (2007a), "Limit Theory for Moderate Deviations from a Unit Root," *Journal of Econometrics* 136, 115-130.

Phillips, P. C. B., T. Magdalinos (2007b), "Limit Theory for Moderate Deviations from a Unit Root Under Weak Dependence," in G. D. A. Phillips and E. Tzavalis (Eds.) *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis.* Cambridge: Cambridge University Press, pp.123-162.

Phillips, P. C. B., T. Magdalinos (2013), "Inconsistent VAR Regression with Common Explosive Roots," *Econometric Theory*, 29(04), 808-837.

Phillips, P.C.B., S. Shi, and Y. Yu (2015).“Testing for Multiple Bubbles: Limit theory of Real Time Detectors”, *International Economic Review*, 2015 (forthcoming).

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

Phillips P. C. B., Y. Wu and J. Yu (2011). “Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?”, *International Economic Review,* 52, pp. 201-226.

Phillips P. C. B. and J. Yu (2011). “Dating the Timeline of Financial Bubbles during the Subprime Crisis,” *Quantitative Economics,* 2, pp. 455-491.

Stock, J. (1991). "Confidence intervals for the largest autoregressive root in US macroeconomic time series," *Journal of Monetary Economics*, 28(3), 435-459.