

Econometric Inference in the Vicinity of Unity.¹

Peter C. B. Phillips

Cowles Foundation for Research in Economics

Yale University

and

University of Auckland & University of York, UK

Tassos Magdalinos

University of Nottingham, UK

April 17, 2009

¹Magdalinos thanks CoFie and SKBI at Singapore Management University for hospitality in April 2009 when this paper was completed. Phillips acknowledges partial research support from the NSF under Grant Nos. SES 04-142254 and SES 06-47086.

Abstract

Present econometric methodology of inference in cointegrating regression is extended to mildly integrated time series of the type introduced by Magdalinos and Phillips (2007, 2009). It is well known that conventional approaches to estimating cointegrating regressions fail to produce even asymptotically valid inference procedures when the regressors are nearly integrated, and substantial size distortions can occur in econometric testing. The new framework developed here enables a general approach to inference that resolves this difficulty and is robust to the persistence characteristics of the regressors, making it suitable for general practical application. Mildly integrated instruments are employed, one using system regressors and internally generated instruments, the other using external instruments. These new IV techniques eliminate the endogeneity problems of conventional cointegration methods with near integrated regressors and robustify inference to uncertainty over the precise nature of the integration in the system. The use of mildly integrated instruments also provides a mechanism for linking the conventional treatment of endogeneity in simultaneous equations with the econometric methodology for cointegrated systems. The methods are easily implemented, widely applicable and help to alleviate practical concerns about the use of cointegration methodology when roots are in the vicinity of unity rather than precisely at unity.

Keywords: Central limit theory, Cointegration, Endogeneity bias, Instrumentation, Mild integration, Mixed normality, Robustness, Simultaneity.

JEL classification: C22

1. Introduction

For the last two decades, autoregressive models with roots near unity have played an important role in time series econometrics, resulting in a vast literature of theory and applications. Theoretical developments have engaged econometricians, probabilists, and statisticians, and the empirical applications extend well beyond economics and finance into other social and business sciences like mass media communications, political science, and marketing. The theoretical work makes extensive use of functional laws for partial sums to Brownian motion, functional laws of weighted partial sums to linear diffusions, mapping theorems for functionals of these processes, and weak convergence of discrete martingales to stochastic integrals and certain nonlinear functionals to Brownian local time. This theory provides a foundation for econometric estimation, testing, power curve evaluation and confidence interval construction for (unit root) nonstationary time series, certain nonlinear nonstationary models and nonstationary panels.

Much of the limit theory for nonstationary time series involves processes with autoregressive roots that are local to unity of the form $\rho = 1 + c/n$, where n is the sample size and c is a constant. Deviations from unity of this particular form are mathematically convenient and help in characterizing local asymptotics, power functions (Phillips, 1987, 1988a), point optimal asymptotic tests (Elliott, Rothenberg, Stock, 1996) and confidence interval construction. But the specific $O(n^{-1})$ rate of approach to unity has no intrinsic significance or economic meaning. To accommodate greater deviations from unity within this framework, we may allow the localizing parameter c to be large or even consider limits as $c \rightarrow \pm\infty$, as was done in Phillips (1987) and Chan and Wei (1988). Such analysis produces certain useful insights, but it does not resolve the difficulties of the discontinuity in the unit root asymptotics. In particular, it does not bridge the very different convergence rates of the stationary, unit root/local to unity, and explosive cases.

Phillips and Magdalinos (2007) and Magdalinos and Phillips (2009, hereafter MP) recently explored another approach, giving a limit theory for time series with an autoregressive root of the form $\rho_n = 1 + c/k_n$, where $(k_n)_{n \in \mathbb{N}}$ is a deterministic sequence increasing to infinity at a rate slower than n . Such roots represent moderate deviations from unity in the sense that they belong to larger asymptotic neighborhoods of one than conventional local to unity roots. But in practice, for finite n the differences need not be large: e.g, when $n = 100$ and $k_n = n/\log n$, $\rho = 0.95$ is captured by the value $c = -1.086$, whereas under the standard local to unity model the same value of ρ is captured with $c = -5$. When $c < 0$, time series generated with such roots may be regarded as mildly integrated, and when $c > 0$ they are mildly explosive. An interesting family of moderate deviations from unity roots occurs when we consider $k_n = n^\alpha$ or $\rho_n = 1 + c/n^\alpha$, where the exponent α lies on $(0, 1)$. The boundary value as $\alpha \rightarrow 1$ includes the conventional local to unity case, whereas the boundary value as $\alpha \rightarrow 0$ includes the stationary or explosive process, depending on the sign of c .

These linkages provided a mechanism for bridging the discontinuities between stationary, local to unity and explosive asymptotics in autoregressions, thereby helping to complete the passage of the limit theory as the autoregressive coefficient moves through unity. Mildly explosive asymptotics extend the range of central limit arguments and thereby inference to the explosive domain. They have recently been found useful in the empirical study of financial bubble phenomena (Phillips, Wang, Yu, 2008). Mildly integrated series may also occur in multivariate systems providing a framework for generalizing standard cointegrating regressions and for linking these regressions to simultaneous equations models with stationary regressors. As Elliott (1998) showed, conventional approaches to estimating cointegrating regressions do not produce valid asymptotic inference in cases where the regressors have autoregressive roots that are local to unity, leading to what can be substantial size distortion in econometric testing. An analogous situation arises when the regressors are mildly integrated. In such cases, recognizing that roots at unity are a special (if important) case in practical applications, there is a need to develop more robust approaches to estimation and inference that do not rely upon knowledge of the precise form of regressor persistence.

The present paper contributes by tackling this problem. We present results that provide a framework of limit theory that can be used to validate inference in cointegrating models with regressors whose time series characteristics fall into the very general class of processes having roots in arbitrary neighbourhoods of unity. We consider cases where the regressors are (i) exactly integrated, (ii) local to unity (in the usual sense of $O(n^{-1})$ departures) and (iii) mildly integrated. Instrumental variables (IV) procedures are developed that address the need for asymptotically valid inference procedures in cointegrated systems where the regressors have roots in the vicinity of unity, but their precise integration properties are unknown.

Two IV approaches are presented. The first uses instrumental variables that are constructed by direct filtering of the (endogenous) regressor variable (x_t), so that no external information (such as the existence of an exogenous instrumental variable) is used in this procedure. We call this approach IVX estimation since the instrumental variable relies directly on the regressor (x_t). This approach makes use of the (possibly mild) nonstationarity in x_t in the construction of the instrument and the IVX regression estimator takes the form of a simple bias adjusted IV regression. It is shown that, by virtue of its construction, the IVX procedure satisfies relevance and orthogonality conditions that lead to a mixed normal limit theory and simple inferential procedures that are robust to the precise form of integration in the system. The second approach uses instrumental variables that are external to the original system and satisfy certain relevance and orthogonality conditions, just as in the case of simultaneous equations estimation. This approach also validates and robustifies inference in cointegrated systems with mildly integrated regressors. It has the particular advantage of relating closely to the classical IV procedure in conventional simultaneous equations theory.

Both approaches involve simple linear estimation methods and are straightforward

to implement in practical work. It is hoped that the resulting theory will resolve many outstanding issues of inference in cointegrated models with roots near unity and provide useful new practical tools for time series econometric work that are as easy to implement as conventional methods for fitting cointegrated systems with unit root regressors.

The paper is organized as follows. Section 2 defines a general cointegrated system with possibly mildly integrated regressors and instruments, and lays out basic regularity conditions that facilitate the asymptotic development. Section 3 proposes the IVX estimation approach, showing how instrumentation with a mildly integrated process removes the usual endogeneity problem in cointegrated regression arising from local to unity regressors. This section also discusses second order bias correction procedures, estimation of systems with mildly integrated regressors, and develops limit theory for all of these cases, establishing the robustness of the methods to the nature of the integration. Section 4 develops the alternative IV approach using external instruments and shows how these methods may also be used for inference in cointegrated systems when there is uncertainty about the precise integration orders of the regressors and external instruments are available. Section 5 concludes. Technical material and proofs are given in Section 6.

2. Cointegration with arbitrary persistence

As is often emphasized in empirical work, economic time series seem to have autoregressive roots in the general neighborhood of unity and so insistence that roots be at unity in cointegrating regressions is likely to be too harsh a requirement in practice. Matrix cases, where the long-run autoregressive coefficient matrix has the form $R_n = I + C/n$ were considered in Phillips (1988a, 1988b) to address this issue and the resulting theory has been useful in developing power functions for testing problems in cointegrated regressions (Phillips, 1988a; Johansen, 1995) and in the analysis of cointegrating regressions for near integrated time series (Elliot, 1998). The present paper allows the regressors to have a much wider range of persistence, covering unit root processes, local to unity processes and processes with roots that lie close to the boundary with stationarity. Accommodating this wider range of possibilities is important in building a connection between cointegrated systems and stationary models of joint dependence such as the simultaneous equations model.

A general modeling framework that is convenient for this purpose is the following multivariate ‘cointegrated’ system with time series regressors in the general vicinity of unity

$$y_t = Ax_t + u_{0t}, \tag{1}$$

$$x_t = R_n x_{t-1} + u_{xt}, \tag{2}$$

for each $t = 1, \dots, n$, for some $m \times K$ coefficient matrix A and diagonal autoregressive matrix R_n whose roots $|\lambda_i(R_n)| \leq 1$ and which satisfies $R_n \rightarrow I_K$ as $n \rightarrow \infty$.

The innovations u_{0t} and u_{xt} are correlated linear processes defined in Assumption LP below and the system is initialized at some x_0 that could be any constant or a random process $x_0(n) = o_p(n^{(\alpha \wedge 1)/2})$ with α specified by Assumption N below. The effect of more general assumptions on the initial condition are discussed in other recent work (Phillips and Magdalinos, 2009) and will not be pursued here to avoid unnecessary complications. Deterministic components may also be included in (1) - (2) and such extensions are easily accommodated, following Park and Phillips (1988).

When $R_n \rightarrow I_K$ very slowly as $n \rightarrow \infty$, the system (1) - (2) has characteristics that are similar to those of a stationary simultaneous equations model. For example, suitably standardized sample moments converge to constant matrices rather than random matrices as $n \rightarrow \infty$ (c.f., MP, and in the univariate case, Phillips and Magdalinos, 2007), in which case the traditional effects of (first order) simultaneous equations bias begin to manifest in ordinary least squares estimation. On the other hand, when $R_n \rightarrow I_K$ very quickly as $n \rightarrow \infty$, the signal of the cointegrating regressors dominates and a second order bias effect, which arises from the long run endogeneity in the system, manifests in least squares limit theory (Phillips and Durlauf, 1986; Stock, 1987).

In consequence, asymptotic inference about the matrix A of cointegrating coefficients depends on the degree of persistence of the nearly integrated regressor x_t , i.e. the rate at which the autoregressive matrix R_n converges to the identity matrix. The effect of changes in persistence on cointegration methods can be categorised more precisely by distinguishing among three classes of neighbourhoods of unity. These classes characterise the asymptotic behavior of the regressor in (2) and are listed in the following (neighborhood of unity) assumption.

Assumption N. *The autoregressive matrix in (2) satisfies the following condition:*

$$R_n = I_K + \frac{C}{n^\alpha}, \quad \text{for some } \alpha > 0 \quad (3)$$

and some matrix $C = \text{diag}(c_1, \dots, c_K)$, with $c_i \leq 0$ for all $i \in \{1, \dots, K\}$. The regressor x_t in (2) belongs to one of the following classes:

- (i) *Integrated regressors, if $C = 0$ or $\alpha > 1$ in (3).*
- (ii) *Local to unity regressors, if $C < 0$ and $\alpha = 1$ in (3).*
- (iii) *Mildly integrated regressors, if $C < 0$ and $\alpha \in (0, 1)$ in (3).*

Some aspects of the limit theory and inferential methods in each of the above cases are well documented: see Park and Phillips (1988), Phillips and Hansen (1990), Johansen (1991), Saikkonen (1991) and Stock and Watson (1993) for (i), Phillips (1988a,b) and Elliott (1998) for (ii) and MP for (iii). The problem is that the validity

of all of the inferential machinery that has been developed in this work is conditional on correct specification of the degree of persistence of the regressor, i.e. on *a priori* knowledge that the regressor x_t belongs to class (i), (ii) or (iii). Without such prior knowledge the effects can be dramatic. Elliott (1998), for instance, showed that the presence of local to unity regressors induces a size distortion in testing that can be serious when conventional cointegration estimators are used. Similar and potentially more serious distortions arise when conventional cointegration methods are applied in the presence of mildly integrated regressors or stationary regressors.

This paper addresses the large question of how to conduct inference in such ‘cointegrated’ systems in a context that is of sufficient generality to be useful in practical work. The goal is valid inference without knowledge of the precise degree of integration of the regressors. To reach this goal, the paper develops an estimating methodology and associated Wald tests on the cointegrating matrix A which are robust to the type of persistence of the regressor in (2) and where the limit theory is mixed normal, normal and standard chi-squared.

The key step to robustifying inference is the development of an instrumental variables procedure based on mildly integrated instruments. The intuition is as follows: as we move substantially towards stationarity ($C < 0$ and $0 < \alpha < 1$ in (2)) valid instrumentation is needed to assist in identifying and estimating the system, just as in a stationary simultaneous equations model; on the other hand, when the regressors have greater persistence and therefore carry a stronger signal, it is possible to utilize this information constructively to eliminate endogeneities in the limit theory, even for case (ii) where the regressors have roots that are local to unity.

Given a K_z -vector of mildly integrated instruments

$$z_t = R_{nz}z_{t-1} + u_{zt} \quad (4)$$

with $K_z \geq K$ and a *known* autoregressive matrix

$$R_{nz} = I_{K_z} + \frac{C_z}{n^\beta}, \quad \beta \in (0, 1), \quad C_z = \text{diag}(c_{z,1}, \dots, c_{z,K_z}), \quad c_{z,i} < 0 \quad (5)$$

it is possible to remove the long run endogeneity that is present in conventional cointegration theory even in the local to unity regressor case. Such instruments may be constructed directly from the regressors (as shown in equation (10) of Section 3 below). So the method is feasible given only the information contained in (1) and (2). On the other hand, if extra information is available and the instruments in (4) are constructed exogenously (that is, so that orthogonality and relevance conditions hold in relation to the original system (1)-(2)), it is possible to deal with systems where the regressors are close to the boundary of stationarity and thereby establish a connection between the present methods and formulation and those of the conventional simultaneous equations system.

In order to accommodate these various possibilities, it is notationally convenient to consider the equations (1), (2) and (4) jointly in one system and defer the construction

of the instruments in (4) until the next section. The correlation structure of the innovations of the system defined by (1), (2) and (4) is provided by the following general purpose assumption, where $\|\cdot\|$ denotes the spectral norm.

Assumption LP. *Let $u_t = (u'_{0t}, u'_{xt}, u'_{zt})'$. For each $t \in \mathbb{N}$, u_t has Wold representation*

$$u_t = F(L)\varepsilon_t = \sum_{j=0}^{\infty} F_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j \|F_j\| < \infty,$$

where $F(z) = \sum_{j=0}^{\infty} F_j z^j$, $F_0 = I_{m+K+K_z}$, $F(1)$ has full rank and $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a sequence of independent and identically distributed $(0, \Sigma)$ random vectors satisfying $\Sigma > 0$ and the moment condition $E \|\varepsilon_1\|^4 < \infty$.

Under LP, the partial sums of u_t satisfy a functional central theorem (cf. Phillips and Solo, 1992)

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} u_t \Rightarrow B(s) \quad \text{on } D_{\mathbb{R}^{m+K+K_z}} [0, 1], \quad (6)$$

where B denotes a Brownian motion with variance $\Omega = F(1)\Sigma F(1)' > 0$ and $D_{\mathbb{R}^p} [0, 1]$ denotes the Skorokhod space of \mathbb{R}^p -valued cadlag functions on $[0, 1]$. The matrices F_j , $F(1)$ and Ω may be partitioned conformably with $u_t = (u'_{0t}, u'_{xt}, u'_{zt})'$ as

$$F_j = \begin{bmatrix} F_{0j} \\ F_{xj} \\ F_{zj} \end{bmatrix}, \quad F(1) = \begin{bmatrix} F_0(1) \\ F_x(1) \\ F_z(1) \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Omega_{00} & \Omega_{0x} & \Omega_{0z} \\ \Omega_{x0} & \Omega_{xx} & \Omega_{xz} \\ \Omega_{z0} & \Omega_{zx} & \Omega_{zz} \end{bmatrix}$$

so that $\Omega_{ij} = F_i(1)\Sigma F_j(1)'$ for each $i, j \in \{0, x, z\}$. The one sided long run covariance matrices $\Lambda = \sum_{h=1}^{\infty} E(u_t u'_{t-h})$ and $\Delta = \sum_{h=0}^{\infty} E(u_t u'_{t-h})$ may be partitioned in a similar way to Ω . The Brownian motion in (6) can also be partitioned conformably to u_t as follows:

$$B(s) = [B_0(s)', B_x(s)', B_z(s)']' \quad (7)$$

where $B_i(s)$ is a Brownian motion with variance Ω_{ii} for each $i \in \{0, x, z\}$. Finally, following the notation of MP we define the covariance matrices

$$V_{xx} = \int_0^{\infty} e^{rC} \Omega_{xx} e^{rC} dr, \quad (8)$$

$$V_{zz} = \int_0^{\infty} e^{rC_z} \Omega_{zz} e^{rC_z} dr \quad \text{and} \quad V_{xz} = \int_0^{\infty} e^{rC} \Omega_{xz} e^{rC_z} dr, \quad (9)$$

corresponding to limiting sample moments of mildly integrated processes.

Semiparametric inference in cointegrated systems typically requires the estimation of the long run covariance matrices Ω and Δ . For example, fully modified least squares

(Phillips and Hansen, 1990) makes substantial use of estimates of these quantities in designing adjustments for long run endogeneity and serial correlation bias. It is well known that (bandwidth/lag truncation) rate conditions on such estimates are required to ensure suitable asymptotic properties in estimation and in particular rates of convergence in the estimated quantities. This is particularly important, for example, in unrestricted FM-VAR estimation where there are unknown numbers of unit roots and cointegrated vectors (Phillips, 1995). Similarly, in the present case, when Δ is estimated by $\hat{\Delta}_n$ using sample autocovariances of (consistent estimates) \hat{u}_t of u_t , it is convenient to impose a lag truncation expansion rate under which $\hat{\Delta}_n$ converges fast enough. The Appendix provides a result of this kind (Lemma A0) which is used for consistent estimation of Δ in the development that follows.

3. IVX estimation: filtered regressors as instruments

This section introduces a new IV procedure for the estimation of the coefficient matrix A in the cointegrated system (1) - (2). The idea is to construct instruments for x_t in (1) by means of a suitable filtering of the regressors themselves, so that there is no external data (such as exogenous instruments) employed in the procedure. Accordingly, we call the approach “IVX estimation” because the instruments depend only on the regressors x_t . In this respect the approach bears all the usual hallmarks of cointegrated system estimation. The intuition underlying the construction is to produce (from the differences of x_t) an instrumental variable whose degree of persistence is explicitly controlled so that the process is mildly integrated. The IVX instruments take the form of mildly integrated processes analogous to (4).

It turns out that this approach eliminates the long run endogeneity that is present in conventional cointegration methods with unit root and local to unity regressors (see Lemma 3.2 below) without specific knowledge of whether the time series have roots at or local to unity. The limit distribution of the IVX estimator is mixed normal and the resulting inferential methods follow standard chi-squared asymptotics and are robust to the degree of persistence in the regressors.

Given the cointegrated system in (1) and (2), we construct instruments \tilde{z}_t from (2) as follows:

$$\tilde{z}_t = \sum_{j=1}^t R_{nz}^{t-j} \Delta x_j \quad (10)$$

where R_{nz} is defined as in (5) for some given C_z , β and $K_z = K$. It suffices to let $C_z = -I_K$ and choose β according to broadly defined criteria that are laid out below. The construction (10) is organized so that only information present in the series x_t is used – that is, no outside time series data (such as an exogenous instrumental variable) is used in the construction. In fact, the weighted partial sum process (10)

produces a mildly integrated instrument that is correlated with the regressor x_t . As shown below in Lemmas 3.1 and 3.2, this instrument satisfies a suitable asymptotic relevance condition for the regressor x_t in (2), while at the same time serving as a valid instrument in terms of inducing a suitable asymptotic orthogonality between its sample covariance with the equation error in (1) and the limit process corresponding to the regressor x_t . Importantly, as we will see, these two properties hold irrespective of whether x_t is an integrated, near integrated process or mildly integrated process. So, the approach has wide generality for application.

In order to develop the limit theory, we separate out two cases of central importance. We begin by discussing the case where the instrument is less persistent than the regressor, i.e. the instrument autoregressive matrix R_{nz} satisfies the restriction

$$\beta < \min(\alpha, 1).$$

Since $\Delta x_t = u_{xt} + \frac{C}{n^\alpha} x_{t-1}$ and R_{nz} and C commute by virtue of being diagonal matrices, the process \tilde{z}_t may be decomposed in terms of a mildly integrated instrument $z_t = \sum_{j=1}^t R_{nz}^{t-j} u_{xj}$ that satisfies

$$z_t = R_{nz} z_{t-1} + u_{xt}, \quad t \in \{1, \dots, n\} \quad z_0 = 0, \quad (11)$$

and a remainder term

$$\psi_{nt} = \sum_{j=1}^t R_{nz}^{t-j} x_{j-1} \quad (12)$$

that arises as a result of quasi-differencing the regressor in (2) as follows:

$$\tilde{z}_t = z_t + \frac{C}{n^\alpha} \psi_{nt}. \quad (13)$$

The following lemma shows that the effects of quasi-differencing on the IVX regression sample moments are manifest only in the signal (relevance) matrix $\sum_{t=1}^n x_t \tilde{z}_t'$.

3.1 Lemma. *Consider the model (1) - (2) satisfying Assumptions N and LP and instruments \tilde{z}_t defined by (10) with $1/2 < \beta < \min(\alpha, 1)$. The following approximations are valid as $n \rightarrow \infty$:*

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

where z_t is defined in (11).

Lemma 3.1 reveals the asymptotic behavior of the key sample moments in (i) and (iii). Since z_t is a mildly integrated process satisfying (11) with $R_{nz} = I_K + C_z/n^\beta$,

equations (7) and (9) of MP yield the following limit matrix for the sample second moments

$$\frac{1}{n^{1+\beta}} \sum_{t=1}^n z_t z_t' \rightarrow_p V_{zz}^x := \int_0^\infty e^{rC_z} \Omega_{xx} e^{rC_z} dr, \quad (14)$$

as $n \rightarrow \infty$; and, for any $\beta \in (1/3, 1)$, the centred and scaled sample covariance has the following martingale form

$$\frac{1}{n^{\frac{1+\beta}{2}}} \sum_{t=1}^n \text{vec}(u_{0t} z_t' - \Delta_{0x}) = U_n(1) + o_p(1), \quad (15)$$

where $U_n(\cdot)$ is the martingale array

$$U_n(s) = \frac{1}{n^{\frac{1+\beta}{2}}} \sum_{t=1}^{\lfloor ns \rfloor} [z_{t-1} \otimes F_0(1) \varepsilon_t] \quad (16)$$

on the Skorokhod space $D_{\mathbb{R}^{mK}}[0, 1]$.

The restriction $\beta > 1/3$ is important and ensures that the system instruments are not so close to the stationary boundary that the bias effects (manifested in Δ_{0x}) cannot be easily eliminated. The potential for bias elimination is one of the important features of cointegrated systems and this feature extends to near-cointegrated models where the regressors are mildly integrated. By contrast, in a fully stationary system, the bias effects have the same order as the limit and in that case the (simultaneous) system (1) is unidentified without further information (such as the observability of an exogenous instrument). This issue is of great importance in distinguishing simultaneous equations systems from cointegrated models and has a substantial effect on estimation methodology. The condition $\beta > 1/3$ in this section of the paper signals this distinction. When the condition is relaxed, the required econometric methodology moves closer to that of traditional simultaneous equations IV theory.

The asymptotic behavior of the sample moment $\sum_{t=1}^n x_t z_t'$ in part (ii) of Lemma 3.1 is determined as follows. Using the recursive property of (2) and (11) we can write

$$x_t z_t' = R_n x_{t-1} z_{t-1}' R_{zn} + R_n x_{t-1} u_{xt}' + u_{xt} z_{t-1}' R_{zn} + u_{xt} u_{xt}'.$$

Vectorising and summing over $t \in \{1, \dots, n\}$ we obtain

$$\begin{aligned} & (I_{K^2} - R_{zn} \otimes R_n) \frac{1}{n} \sum_{t=1}^n \text{vec}(x_{t-1} z_{t-1}') \\ &= [I_K + o_p(1)] \text{vec} \left[\frac{1}{n} \sum_{t=1}^n x_{t-1} u_{xt}' + \frac{1}{n} \sum_{t=1}^n u_{xt} z_{t-1}' + \frac{1}{n} \sum_{t=1}^n u_{xt} u_{xt}' \right] \\ &= \text{vec} \left[\frac{1}{n} \sum_{t=1}^n x_{t-1} u_{xt}' + \Lambda'_{xx} + E(u_{x1} u_{x1}') \right] + o_p(1), \end{aligned} \quad (17)$$

by Lemma 3.1(d) of MP and the ergodic theorem. The asymptotic behavior of $n^{-1} \sum_{t=1}^n x_{t-1} u'_{xt}$ depends on the order of persistence of x_t . A standard application of the Phillips and Solo (1992) method yields

$$\frac{1}{n} \sum_{t=1}^n x_{t-1} u'_{xt} = \frac{1}{n} \sum_{t=1}^n x_{t-1} \varepsilon'_t F_x(1)' + \Lambda_{xx} + o_p(1) \quad \text{as } n \rightarrow \infty. \quad (18)$$

If x_t is mildly integrated in the sense of Assumption N(iii), the right side of (18) reduces asymptotically to the constant matrix Λ_{xx} by Lemma 3.3 of MP. If x_t is local to unity, the martingale array on the right side of (18) converges weakly to the matrix stochastic integral $\int_0^1 J_C(s) dB_x(s)'$, where $B_x(s)$ is the Brownian motion with variance Ω_{xx} defined in (7) and

$$J_C(s) = \int_0^s e^{C(s-r)} dB_x(r) \quad (19)$$

is the associated Ornstein-Uhlenbeck process. Since, when $\alpha > \beta$,

$$I_{K^2} - R_{zn} \otimes R_n = -\frac{1}{n^\beta} (C_z \otimes I_K) \left[I_K + O_p\left(\frac{1}{n^{\alpha-\beta}}\right) \right] \quad \text{as } n \rightarrow \infty$$

(17) and (18) imply that

$$\frac{1}{n^{1+\beta}} \sum_{t=1}^n x_{t-1} z'_{t-1} \Rightarrow \begin{cases} -\left(\int_0^1 B_x dB'_x + \Omega_{xx}\right) C_z^{-1} & \text{under N(i)} \\ -\left(\int_0^1 J_C dB'_x + \Omega_{xx}\right) C_z^{-1} & \text{under N(ii)} \\ -\Omega_{xx} C_z^{-1} & \text{under N(iii)} \end{cases} \quad (20)$$

as $n \rightarrow \infty$, the result for N(iii) applying under the additional condition that $\alpha > \beta$. The limit (20) shows that the standardized moment $n^{-1-\beta} \sum_{t=1}^n x_{t-1} z'_{t-1}$ takes various forms, including both random and nonrandom, depending on the nature of the nonstationarity in the regressor and instrument.

The results of Lemma 3.1, the representation (15) and the limits in (20) determine the individual asymptotic behavior of relevant sample moments in the estimation limit theory. In particular, they reveal the potential for an orthogonality condition via (15) and a suitable relevance condition via (20). However, individually, these results offer no information on the presence or lack of endogeneity in the limit, which is the critical issue underlying robust inference in nearly cointegrated systems. The next lemma addresses this issue by establishing asymptotic independence between the two central components $n^{-\frac{1+\beta}{2}} \sum_{t=1}^n (u_{0t} z'_t - \Delta_{0x})$ and $n^{-(1+\beta)} \sum_{t=1}^n x_t z'_t$, and showing that joint convergence applies.

3.2 Lemma. *Consider the model (1) - (2) satisfying Assumptions N and LP and the mildly integrated process z_t defined in (11). The martingale array $U_n(s)$ in (16) satisfies*

$$U_n(s) \Rightarrow U(s) \quad \text{on } D_{\mathbb{R}^{mK}} [0, 1]$$

as $n \rightarrow \infty$, where U is a Brownian motion, independent of B_x , with variance V_{zz}^x defined in (14). Joint convergence in distribution of $U_n(1)$, $n^{-1} \sum_{t=1}^n x_{t-1} \varepsilon_t'$ and $n^{-1-\alpha} \sum_{t=1}^n x_{t-1} x'_{t-1}$ also applies as $n \rightarrow \infty$.

3.3 Remarks.

- (i) Lemma 3.2 is a consequence of Proposition A1 in the Appendix. It is important to note that, unlike conventional methods for removing endogeneity, the asymptotic independence between the martingale part of the sample covariance matrix $\sum_{t=1}^n u_{0t} z_t'$ and the signal matrix $\sum_{t=1}^n x_t \tilde{z}_t'$ holds by virtue of the reduced order of magnitude of the instrument z_t and, importantly, does *not* rely on a separate orthogonality correction or condition on the instrument. This is evident from equation (35) in the Appendix which shows that the conditional variance matrix of the martingale array consisting of $U_n(s)$ and the martingale part of the partial sum process of u_t is asymptotically block diagonal if and only if

$$\frac{1}{n^{1+\frac{\beta}{2}}} \sum_{t=1}^{\lfloor ns \rfloor} z_{t-1} \rightarrow_p 0,$$

which holds when z_t is a mildly integrated process with $\beta \in (0, 1)$ but fails when $\beta = 1$.

- (ii) Lemma 3.2 shows that instrumentation with a mildly integrated process completely removes the long run endogeneity (typically associated with stochastic integrals of the type $\int_0^1 B_x dB_x'$ and $\int_0^1 J_C dB_x'$) that arises in least squares estimation of cointegrated systems of the form (1) - (2) with integrated and local to unity regressors. These effects still manifest in the relevance matrix limit, as is apparent in (20), but are eliminated from the martingale component (16) that drives the limit distribution theory. On the other hand, as is evident in the centering of (15), there is a bias term of the form $n^{\frac{1-\beta}{2}} \Delta_{0x}$ that carries the effect of simultaneity. This bias term can be estimated and removed by standard nonparametric methods, as we now discuss.

The limit theory established above reveals the possibility of constructing a bias-corrected IVX estimator of the coefficient matrix A in (1) which is asymptotically mixed Gaussian for a very general class of persistent regressors. Using conventional regression notation, let

$$Y = [y_1', \dots, y_n']', \quad X = [x_1', \dots, x_n']' \quad \text{and} \quad \tilde{Z} = [\tilde{z}_1', \dots, \tilde{z}_n']'. \quad (21)$$

The bias corrected IVX estimator has the form

$$\tilde{A}_n = \left(Y' \tilde{Z} - n \hat{\Delta}_{0x} \right) \left(X' \tilde{Z} \right)^{-1}, \quad (22)$$

and is asymptotically mixed Gaussian, as the following theorem shows. The estimator is analogous to the FM-OLS estimator (Phillips and Hansen, 1990) in terms of its built-in bias correction term involving $\hat{\Delta}_{0x}$, but unlike FM-OLS there is no need for an endogeneity correction (either in Y or $\hat{\Delta}_{0x}$). The quantity $\hat{\Delta}_{0x}$ is taken to be a consistent nonparametric estimate of the one sided long run covariance matrix Δ_{0x} , which may be constructed by conventional methods, as given in (27) and considered in Lemma A0 in the Appendix.

3.4 Theorem. *Consider the model (1) - (2) satisfying Assumptions N and LP with instruments \tilde{z}_t defined by (10) with $2/3 < \beta < \min(\alpha, 1)$. Then, the following limit theory applies for the estimator \tilde{A}_n in (22):*

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

as $n \rightarrow \infty$, where

$$\tilde{\Psi}_{xx} = \begin{cases} \Omega_{xx} + \int_0^1 B_x dB_x' & \text{under } N(i) \\ \Omega_{xx} + \int_0^1 J_C dJ_C' & \text{under } N(ii) \\ \Omega_{xx} + V_{xx} C & \text{under } N(iii) \end{cases},$$

J_C is the Ornstein-Uhlenbeck process in (19) and V_{zz}^x and V_{xx} are defined in (14) and (8) respectively.

Having established Theorem 3.4, it remains to discuss the asymptotic behavior of the IVX estimator \tilde{A}_n when the instrument is more persistent than a mildly explosive regressor, i.e. when the vector of regressors x_t satisfies Assumption N(iii) with $\alpha \leq \beta$. In this case, it is useful to apply a different decomposition of \tilde{z}_t than the one presented in (13). In particular, using summation by parts we have

$$\begin{aligned} \tilde{z}_t &= \sum_{j=1}^t R_{nz}^{t-j} \Delta x_j = x_t - R_{nz}^t x_0 - \sum_{j=1}^t (\Delta R_{nz}^{t-j}) x_{j-1} \\ &= x_t - R_{nz}^t x_0 + \frac{C_z}{n^\beta} \psi_{nt}, \end{aligned} \tag{23}$$

where ψ_{nt} is defined in (12). Applying (23) we can show that, when $\alpha < \beta$, the contribution of \tilde{z}_t to the asymptotic behavior of the various sample moments of interest consists exclusively of x_t . This case and the $\alpha = \beta$ case are presented in the following two results which are proved in the Appendix.

3.5 Lemma. *Consider the model (1) - (2) satisfying Assumptions N(iii) and LP and instruments \tilde{z}_t defined by (10) with $\beta \in (1/2, 1)$.*

(i) If $\alpha \in (1/3, \beta)$, then

$$\frac{1}{n^{\frac{1+\alpha}{2}}} \sum_{t=1}^n u_{0t} \tilde{z}'_t = \frac{1}{n^{\frac{1+\alpha}{2}}} \sum_{t=1}^n u_{0t} x'_t + o_p(1) \quad \text{as } n \rightarrow \infty.$$

(ii) For any $\alpha \in (0, \beta)$, the following approximations hold as $n \rightarrow \infty$:

$$\begin{aligned} \frac{1}{n^{1+\alpha}} \sum_{t=1}^n x_t \tilde{z}'_t &= \frac{1}{n^{1+\alpha}} \sum_{t=1}^n x_t x'_t + o_p(1), \\ \frac{1}{n^{1+\alpha}} \sum_{t=1}^n \tilde{z}_t \tilde{z}'_t &= \frac{1}{n^{1+\alpha}} \sum_{t=1}^n x_t x'_t + o_p(1), \end{aligned}$$

so both $n^{-1-\alpha} \sum_{t=1}^n x_t \tilde{z}'_t$ and $n^{-1-\alpha} \sum_{t=1}^n \tilde{z}_t \tilde{z}'_t$ converge to V_{xx} in probability.

3.6 Lemma. Consider the model (1) - (2) satisfying Assumptions N(iii) and LP and instruments \tilde{z}_t defined by (10) with $\alpha = \beta \in (1/2, 1)$. Then, letting

$$\mathbb{V}_{xz} = \int_0^\infty e^{rC} V_{xx} e^{rCz} dr,$$

the following limits apply as $n \rightarrow \infty$:

- (i) $n^{-(1+\alpha)} \sum_{t=1}^n x_t \tilde{z}'_t \rightarrow_p -C \mathbb{V}_{xz}$,
- (ii) $n^{-(1+\alpha)} \sum_{t=1}^n \tilde{z}_t \tilde{z}'_t \rightarrow_p \int_0^\infty e^{sCz} (C \mathbb{V}_{xz} C_z + C_z \mathbb{V}'_{xz} C) e^{sCz} ds$,
- (iii) $n^{-\frac{1+\alpha}{2}} \sum_{t=1}^n \text{vec}(u_{0t} \tilde{z}'_t - \Delta_{0x}) \Rightarrow N(0, \int_0^\infty e^{sCz} (C \mathbb{V}_{xz} C_z + C_z \mathbb{V}'_{xz} C) e^{sCz} ds \otimes \Omega_{00})$.

Using the limit theory of Lemmas 3.5 and 3.6 we can establish an analog of Theorem 3.4 which yields a normal limit distribution for the IVX estimator in the case when the original regressor in (2) is mildly integrated and we employ a more persistent instrument.

3.7 Theorem. Consider the model (1) - (2) satisfying Assumptions N(iii) with $\alpha \in (1/3, \beta]$ and LP and instruments \tilde{z}_t defined by (10) with $\beta \in (2/3, 1)$. Then, the following limit theory applies for the IVX estimator \tilde{A}_n in (22) as $n \rightarrow \infty$:

$$\begin{aligned} n^{\frac{1+\alpha}{2}} \text{vec} \left(\tilde{A}_n - A \right) &\Rightarrow N(0, V_{xx}^{-1} \otimes \Omega_{00}) && \text{if } \alpha < \beta \\ n^{\frac{1+\alpha}{2}} \text{vec} \left(\tilde{A}_n - A \right) &\Rightarrow N(0, \mathbb{V}_{xz}^{-1} C^{-1} V_{xx} C^{-1} (\mathbb{V}'_{xz})^{-1} \otimes \Omega_{00}) && \text{if } \alpha = \beta. \end{aligned}$$

Asymptotic mixed normality of the IVX estimator in Theorems 3.4 and 3.7 implies that conventional Wald tests for testing linear restrictions on A will have standard chi-square limit distributions. As in Park and Phillips (1988), general linear restrictions on the cointegrating coefficients

$$H_0 : H \text{vec}(A) = h, \quad (24)$$

where H is a known $r \times mK$ matrix with rank r and h is a known vector, may be tested using the Wald statistic

$$W_n = \left(H \text{vec} \tilde{A}_n - h \right)' \left[H \left\{ (X' P_{\tilde{Z}} X)^{-1} \otimes \hat{\Omega}_{00} \right\} H' \right]^{-1} \left(H \text{vec} \tilde{A}_n - h \right) \quad (25)$$

where $P_{\tilde{Z}} = \tilde{Z} \left(\tilde{Z}' \tilde{Z} \right)^{-1} \tilde{Z}'$ denotes the projection matrix to the column space of \tilde{Z} . The asymptotic behavior of the normalized sample moment matrix $\tilde{Z}' \tilde{Z}$ is given by Lemma 3.1(iii) and (14) under the assumptions of Theorem 3.4 and by Lemmas 3.5 and 3.6 under the assumptions of Theorem 3.7. The following result shows that the IVX procedure gives rise to robust inferences about A . using (25). It is an immediate corollary of Theorems 3.4 and 3.7. Obvious extensions to analytic restrictions in place of (24) are also covered.

3.8 Theorem. *Consider the model (1) - (2) satisfying Assumptions N and LP with $\alpha > 1/3$ and instruments \tilde{z}_t defined (10) with $\beta \in (2/3, 1)$. The Wald statistic W_n in (25) for testing the linear restrictions (24) satisfies*

$$W_n \Rightarrow \chi^2(r) \quad \text{under } H_0$$

as $n \rightarrow \infty$, where r is the rank of H .

3.9 Remarks.

- (i) Theorem 3.8 provides a robust inferential procedure which is valid for a very general class of persistent regressors. Unlike conventional cointegration methods such as fully modified least squares, *a priori* knowledge of the order of persistence is not required and standard χ^2 inference applies for unit root, local to unity and mildly integrated regressors alike. In this sense, Theorems 3.4, 3.7 and 3.8 provide a solution to the problem of robustness of cointegration methods, highlighted in Elliott (1998).
- (ii) The price paid for the method's robustness is a reduction in the rate of convergence of \tilde{A}_n to $O\left(n^{\frac{1+\beta}{2}}\right)$ from $O(n)$ in the case of applying standard cointegration methods under the knowledge that the regressors are exact unit root processes. Interestingly, \tilde{A}_n remains asymptotically efficient in the case where

the vector of regressors is mildly integrated and we instrument by a more persistent process (Theorem 3.7 with $\alpha < \beta$), and it retains the OLS rate of convergence when $\alpha = \beta$. In any case, a reduction in the convergence rate of \tilde{A}_n should not affect the performance of self normalized tests such as the Wald test of Theorem 3.8.

- (iii) A related point to the above remark is that the approach of removing long run endogeneity by constructing instruments with a reduced rate of convergence does not completely remove asymptotic bias. As earlier work in MP has shown, cointegrated systems with mildly integrated regressors do not suffer from long run endogeneity problems but there is asymptotic bias in least squares estimation and this bias (like simultaneous equations bias) can be more severe than in systems with integrated and local to unity regressors. This observation suggests that, in the present context, there is some trade-off between long run endogeneity and bias effects in terms of practical implementation when choosing the rate of persistence n^β of the IVX instrument. A choice of β close to unity typically reduces the effects of (simultaneity) bias but may exacerbate long run endogeneity effects in least squares regression. On the other hand, the effect of simultaneity bias increases the closer we choose the instrument to the stationary direction. Indeed, the bias becomes too severe for the method to work for $\beta \in (0, 1/2)$. In that event, we need another IV approach that takes advantage of external information, just as in simultaneous equations estimation. The restriction $\beta > 2/3$ is imposed in order to ensure the inclusion of optimal bandwidths associated with non-parametric estimation of the long run covariance matrix Δ (see Lemma A0), which is needed for bias removal.
- (iv) The IVX method presented in this Section is feasible for any cointegrated system generated by Assumption N with $\alpha > 1/3$ in the sense that it can be implemented using only the information provided by the statistical model (1) - (2). The instruments are constructed from the regressors in (2) without assuming any exogenous information. For precisely this reason, the method cannot accommodate mildly integrated regressors that are too close to the stationary region. As earlier work in MP has shown, simultaneity in cointegrated systems with mildly integrated regressors becomes more severe as we approach the stationary region and, eventually for $\alpha \in (0, 1/3]$, it presents similar difficulties to simultaneous equations bias in that it cannot be removed without the use of exogenous instrumental variables. Classical IV procedures that utilize exogenous information and address this problem are considered in the next Section.

4. Classical IV inference with mild integration

In this section we show how to conduct inference on the cointegrated system (1) - (2) when additional information is available in the form of a K_z -vector of mildly integrated instruments z_t , with $K_z \geq K$, satisfying (4), (5) and the following form of the classical IV long-run relevance condition holds. Importantly, orthogonality or long-run orthogonality of the instruments is not required.

Assumption IV(i). *The long run covariance matrix $\Omega_{xz} = F_x(1) \Sigma F_z(1)'$ has full rank equal to K .*

The instruments may be simply taken as given, as is commonly done in simultaneous equations theory, or we may employ a suitable constructive process. The advantage of the latter is that we may control the degree of mild integration in the instrument. One possible approach to constructing such instruments is to employ suitable macroeconomic or financial series that are known to be unit root processes and for which the long run relevance condition holds. Accordingly, let η_t be a K_z -vector unit root process that is correlated with the vector of regressors x_t of the original system (1) - (2). In this case the mildly integrated instruments z_t may be constructed by differencing η_t and forming

$$z_t = \sum_{j=1}^t R_{nz}^{t-j} \Delta \eta_j,$$

for some chosen matrix R_{nz} satisfying (5). Using a unit root process like η_t eliminates the effects of quasi-differencing (see Lemmas 3.1, 3.5 and 3.6 in the previous section) in the earlier construction, but extensions of our theory to that case also apply although they will not be detailed here.

Classical IV inference also imposes incoherence between the innovations of (1) and those of (4). We only make use of such a strong orthogonality condition in order to establish a boundary with stationary simultaneous equation systems. The condition is stated here for convenient subsequent reference in long run orthogonality form.

Assumption IV(ii). *The innovations u_{0t} and u_{zt} of (1) and (4) satisfy the strong (long run) orthogonality condition $\Delta_{0z} = 0$.*

We now return to our program of providing robust inference for the cointegrated system (1) - (2) by means of the mildly integrated instruments z_t in (4). Under both Assumptions IV(i) and IV(ii) this can be achieved by the classical IV estimator

$$\hat{A}_{IV} = Y' P_Z X (X' P_Z X)^{-1}$$

where Y and X are defined in (21), $Z = [z'_1, \dots, z'_n]'$ and $P_Z = Z (Z' Z)^{-1} Z'$.

Given the practical difficulty of assessing the validity of the long run orthogonality condition IV(ii), it is desirable to conduct inference on the basis of Assumption IV(i) only. To this end, we define a bias corrected version of \hat{A}_{IV} as follows

$$\begin{aligned}\tilde{A}_{IV} &= \hat{A}_{IV} - n\hat{\Delta}_{0z} (Z'Z)^{-1} Z'X (X'P_ZX)^{-1} \\ &= \left(Y'Z - n\hat{\Delta}_{0z} \right) (Z'Z)^{-1} Z'X (X'P_ZX)^{-1}.\end{aligned}$$

The estimator \tilde{A}_{IV} is constructed using a similar bias correction principle to that of the IVX estimator \tilde{A}_n in the previous section. The difference between the two estimators consists of the additional information used in the construction of the instrument matrix Z which, in the case of \tilde{A}_{IV} , avoids the restrictions and inefficiency associated with quasi-differencing the regressors in (2).

As in Section 3, the asymptotic behavior of the sample moments $\sum_{t=1}^n z_t z_t'$ and $\sum_{t=1}^n u_{0t} z_t'$ can be deduced directly from MP, and is detailed in the following result.

4.1 Lemma. *Consider the model (1), (2) and (4) satisfying Assumptions N, LP and IV(i). Then, as $n \rightarrow \infty$:*

- (i) $n^{-\frac{1+\beta}{2}} \sum_{t=1}^n (u_{0t} z_t' - \Delta_{0z}) = U_n(1) + o_p(1)$, for each $\beta \in (1/3, 1)$
- (ii) $n^{-(1+\beta)} \sum_{t=1}^n z_t z_t' \rightarrow_p V_{zz}$

where $U_n(\cdot)$ is the martingale array defined in (16) and V_{zz} is given by (9). Under Assumption IV(ii), $\Delta_{0z} = 0$, and part (i) holds for all $\beta \in (0, 1)$.

The asymptotic behavior of the $U_0'Z$ matrix in Lemma 4.1 presents an important difference in comparison to the corresponding sample covariance matrix of Section 3. As there is no quasi-differencing effect like that in Lemmas 3.1 and 3.5, there is no need to restrict α and the validity of IV inference extends over the whole range $\alpha \in (0, 1)$ of mildly integrated regressors allowing linkages to the boundary with stationary simultaneous equation systems.

4.2 Lemma. *Consider the model (1), (2) and (4) satisfying Assumptions N, LP and IV(i). Let $J_n = n^{-1} \sum_{t=1}^n x_{t-1} \varepsilon_t' F_z(1)'$. Then, as $n \rightarrow \infty$:*

- (i) $n^{-1-(\beta \wedge \alpha)} \sum_{t=1}^n \text{vec}(x_{t-1} z_{t-1}') = [n^{\beta \wedge \alpha} (I_{KK_z} - R_{zn} \otimes R_n)]^{-1} \text{vec}(J_n + \Omega_{xz}) + o_p(1)$,
- (ii) $n^{-1-(\beta \wedge \alpha)} X'Z \Rightarrow \Psi_{xz}$, where:

$$\Psi_{xz} = \begin{cases} - \left(\Omega_{xz} + \int_0^1 B_x dB_z' \right) C_z^{-1} & \text{under } N(i) \\ - \left(\Omega_{xz} + \int_0^1 J_C dB_z' \right) C_z^{-1} & \text{under } N(ii) \\ V_{xz} & \text{under } N(iii) \text{ with } \alpha = \beta \\ -\Omega_{xz} C_z^{-1} & \text{under } N(iii) \text{ with } \alpha > \beta \\ -C^{-1} \Omega_{xz} & \text{under } N(iii) \text{ with } \alpha < \beta \end{cases} \quad (26)$$

where V_{xz} is given by (9), B_x and B_z are Brownian motions defined in (7) and J_C is the Ornstein-Uhlenbeck process defined in (19).

Having obtained the asymptotic behavior of the various sample moments in Lemmas 4.1 and 4.5, we can employ a similar approach to that used in Section 3 to derive the limit distribution of the IV estimators \tilde{A}_{IV} and \hat{A}_{IV} . The joint asymptotic behavior of the matrices $\sum_{t=1}^n u_{0t}z_t'$ and $X'Z$ can be deduced by the approximations in Lemma 4.1(i) and Lemma 4.2(i) in conjunction with Proposition A1 in the Appendix: the martingale part $U_n(1)$ of the sample covariance matrix and $\text{vec}(J_n)$ of Lemma 4.2 converge jointly in distribution to independent random vectors. We therefore obtain asymptotic mixed normality for the normalised and centered IV estimators \tilde{A}_{IV} and \hat{A}_{IV} under any degree of regressor persistence specified by Assumption N, as the next result shows. The direct consequences for inference follow in the corollary.

4.3 Theorem. *Consider the model (1), (2) and (4) satisfying Assumptions N, LP. Let $\gamma = \min\left(\frac{\beta}{2}, \alpha - \frac{\beta}{2}\right)$.*

(i) *Under Assumption IV(i), we obtain, for each $\beta \in (2/3, 1)$,*

$$n^{\frac{1}{2}+\gamma} \text{vec}\left(\tilde{A}_{IV} - A\right) \Rightarrow MN\left(0, (\Psi_{xz}V_{zz}^{-1}\Psi'_{xz})^{-1} \otimes \Omega_{00}\right) \text{ as } n \rightarrow \infty.$$

(ii) *Under Assumption IV(ii), we obtain, for each $\beta \in (0, 1)$,*

$$n^{\frac{1}{2}+\gamma} \text{vec}\left(\hat{A}_{IV} - A\right) \Rightarrow MN\left(0, (\Psi_{xz}V_{zz}^{-1}\Psi'_{xz})^{-1} \otimes \Omega_{00}\right) \text{ as } n \rightarrow \infty$$

where Ψ_{xz} is given by (26).

4.4 Corollary. *Under the hypotheses of Theorem 4.3(i), the Wald statistic*

$$W_n = \left(H \text{vec} \tilde{A}_{IV} - h\right)' \left[H \left\{ (X'P_Z X)^{-1} \otimes \hat{\Omega}_{00} \right\} H' \right]^{-1} \left(H \text{vec} \tilde{A}_{IV} - h \right)$$

for testing the linear restrictions (24) where H is a known matrix with rank r and h is a known vector has a $\chi^2(r)$ limit distribution under H_0 . Under the hypotheses of Theorem 4.3(ii), the same conclusion applies with \tilde{A}_{IV} replaced by \hat{A}_{IV} in W_n .

Clearly, similar results hold in the case of Wald tests for analytic restrictions under standard regularity conditions.

5. Conclusion

This work develops a theory of econometric estimation and inference for cointegrated systems with regressors that have roots in the general vicinity of unity. Two instrumental variable approaches are developed. One involves internally generated instruments (IVX) and the other uses external instruments (IV), which satisfy certain

relevance conditions but do not necessarily require orthogonality conditions. Both procedures use mildly integrated instruments, both eliminate the endogeneity that is present in conventional cointegration methods when roots are no longer precisely at unity, and both produce asymptotically mixed Gaussian estimators that are convenient for inference. The methods and the resulting inferential techniques are robust to the (generally unknown) persistence properties of the regressors.

Instrumentation by means of a mildly integrated process plays a crucial role in our approach. The underlying reason is that, by virtue of their intermediate persistence rate in comparison to stationary and integrated processes, mildly integrated time series maintain a balance between long run bias and endogeneity. In other words: (i) reduction in the ‘degree of nonstationarity’ through the use of a mildly integrated instrument eliminates the long run endogeneity effect associated with $I(1)$ processes, at the potential cost of some mild reduction in rate of convergence; (ii) the fact that mildly integrated processes are more persistent than stationary processes reduces the impact of simultaneity and, unlike the case of simultaneous equations systems, produces a tractable expression for the asymptotic bias which can be dealt with in estimation, just as in the case of procedures like fully modified least squares.

It is important to mention that the polynomial rates n^α and n^β used throughout this paper in the modeling of mildly integrated series can be replaced by arbitrary sequences (involving slowly varying functions at infinity) that preserve the balance between regressors and instruments. For example, a mildly integrated process in a $\log n/n$ - neighborhood of unity provides equally valid instruments. This type of generalization then follows along the lines of the univariate limit theory given in Phillips and Magdalinos (2007).

The IVX and IV approaches developed here both involve linear estimation methods and are straightforward to implement in practical work. They should be widely applicable and help to alleviate practical concerns about the use of cointegration methodology when roots are in the vicinity of unity rather than precisely at unity. In empirical work, unit root and stationarity pretests are often inconclusive, throwing into doubt the use of conventional cointegration methodology, even when the series manifest forms of stochastic nonstationarity such as randomly wandering behavior with no fixed mean. In such cases, the econometric approaches developed here provide a robust alternative for estimation and inference which allow for a wider class of nonstationarity with roots near unity rather than requiring all series to be precisely $I(1)$.

6. Technical Appendix and Proofs

This Appendix contains some technical results that are useful in the development of the limit theory of the paper as well as proofs of the theorems in the paper. The first result concerns the order of consistency of the commonly used Bartlett (Newey-West)

estimator of Δ , viz.,

$$\hat{\Delta}_n = \frac{1}{n} \sum_{h=0}^M \left(1 - \frac{h}{M+1}\right) \sum_{t=h+1}^n \hat{u}_t \hat{u}'_{t-h}. \quad (27)$$

Related results hold for other commonly used long run and one sided long run covariance matrix estimators and are not given here but may be used without affecting the asymptotic theory of our estimators provided the general bandwidth condition (28) holds. For example, optimal bandwidth choices such as $M = Kn^{1/5}$ for other kernels are included under the general condition that $\beta \in (2/3, 1)$.

Lemma A0. *Consider the estimator $\hat{\Delta}_n$ in (27) with $M = Kn^{1/3}$, for some fixed constant K . Under Assumption LP,*

$$\hat{\Delta}_n - \Delta = o_p\left(n^{-\frac{1-\beta}{2}}\right) \text{ as } n \rightarrow \infty$$

for any $\beta \in (2/3, 1)$.

Proof of Lemma A0. It can be shown by standard methods that, under Assumption LP,

$$\hat{\Delta}_n - \Delta = O_p\left(\max\left\{\frac{M}{n^{1/2}}, \frac{1}{M}\right\}\right) \quad (28)$$

for any bandwidth parameter M increasing to ∞ with n . Therefore, the requirement $n^{\frac{1-\beta}{2}}(\hat{\Delta}_n - \Delta) = o_p(1)$ yields the restriction

$$\max\left\{\frac{M}{n^{\beta/2}}, \frac{n^{\frac{1-\beta}{2}}}{M}\right\} \rightarrow 0$$

which is satisfied for all $\beta \in (2/3, 1)$ if $M = Kn^{1/3}$. ■

Proposition A1. *Consider the system of equations (1), (2) and (4) with autoregressive matrices R_n and R_{nz} given by (3) and (5) respectively and innovations u_t satisfying Assumption LP.*

(i) *Consider the martingale array $U_n(s)$ defined in (16). Then*

$$\left[\begin{array}{c} U_n(s) \\ n^{-1/2} \sum_{t=1}^{\lfloor ns \rfloor} u_t \end{array} \right] \Rightarrow \left[\begin{array}{c} U(s) \\ B(s) \end{array} \right] \quad (29)$$

on the Skorokhod space $D_{\mathbb{R}^{mK_z+m+K+K_z}}[0, 1]$ where U and B are independent Brownian motions with variances $V_{zz} \otimes \Omega_{00}$ and Ω respectively and V_{zz} is defined in (9). If z_t is generated by (11), the above result holds with V_{zz} replaced by the matrix V_{zz}^x defined in (14).

(ii) Under Assumption N(ii),

$$\begin{bmatrix} U_n(1) \\ \text{vec} \left\{ n^{-1} \sum_{t=1}^n x_{t-1} \varepsilon_t' \right\} \\ \text{vec} \left\{ n^{-1} \sum_{t=1}^n x_{t-1} x_{t-1}' \right\} \end{bmatrix} \Rightarrow \begin{bmatrix} U(1) \\ \text{vec} \int_0^1 J_C(s) dW(s)' \\ \text{vec} \int_0^1 J_C(s) J_C(s)' ds \end{bmatrix} \text{ as } n \rightarrow \infty \quad (30)$$

where W is a Brownian motion with variance Σ and $J_C(s) = \int_0^s e^{(s-r)C} dB_x(r)$. Under Assumption N(i), (30) continues to apply with J_C replaced by B_x .

Proof of part (i). Applying the BN decomposition to u_t (see Phillips and Solo, 1992) we find that the left side of (29) is asymptotically equivalent in probability to the \mathcal{F}_{nt} -martingale array $\sum_{t=1}^{\lfloor ns \rfloor} \xi_{nt}$ where

$$\xi_{nt} := n^{-1/2} \left[(n^{-\beta/2} z_{t-1} \otimes F_0(1) \varepsilon_t)', (F(1) \varepsilon_t)' \right]'$$

and $\mathcal{F}_{nt} = \sigma(x_0, z_0, \varepsilon_t, \varepsilon_{t-1}, \dots)$. The result will follow by employing the martingale invariance principle of Jacod and Shiryaev (1987, VIII, Theorem 3.33) to $\sum_{t=1}^{\lfloor ns \rfloor} \xi_{nt}$.

To do so, we first establish the conditional Lindeberg condition

$$\sum_{t=1}^{\lfloor ns \rfloor} E_{\mathcal{F}_{nt-1}} (\|\xi_{nt}\|^2 \mathbf{1} \{\|\xi_{nt}\| > \delta\}) \rightarrow_p 0, \quad (31)$$

for all $s \in [0, 1]$ and $\delta > 0$. Using the inequality $(a + b)^{1/2} \leq a^{1/2} + b^{1/2}$ for all $a, b > 0$ and the fact that $\|F(1)\| \geq \|F_0(1)\|$, we obtain

$$\begin{aligned} \|\xi_{nt}\| &\leq \frac{1}{n^{1/2}} \left(\frac{\|F_0(1)\|}{n^{\beta/2}} \|z_{t-1}\| \|\varepsilon_t\| + \|F(1)\| \|\varepsilon_t\| \right) \\ &\leq \frac{2\|F(1)\|}{n^{1/2}} \left(\frac{\|z_{t-1}\|}{n^{\beta/2}} + 1 \right) \|\varepsilon_t\| \end{aligned} \quad (32)$$

which, in turn, implies that

$$\begin{aligned} \mathbf{1} \{\|\xi_{nt}\| > \delta\} &\leq \mathbf{1} \left\{ \frac{2\|F(1)\|}{n^{\frac{1}{2}-\frac{\beta}{4}}} \left(\frac{\|z_{t-1}\|}{n^{\beta/2}} + 1 \right) > \sqrt{\delta} \right\} + \mathbf{1} \left\{ \frac{\|\varepsilon_t\|}{n^{\beta/4}} > \sqrt{\delta} \right\} \\ &\leq \mathbf{1} \left\{ \frac{\|z_{t-1}\|}{n^{\frac{1}{2}+\frac{\beta}{4}}} > \frac{\delta'}{2} \right\} + \mathbf{1} \left\{ \frac{1}{n^{\frac{1}{2}-\frac{\beta}{4}}} > \frac{\delta'}{2} \right\} + \mathbf{1} \left\{ \frac{\|\varepsilon_t\|}{n^{\beta/4}} > \sqrt{\delta} \right\} \end{aligned} \quad (33)$$

where $\delta' = \sqrt{\delta}/2 \|F(1)\|$. Using (32) and the \mathcal{F}_{nt-1} -measurability of z_{t-1} we obtain that

$$\sum_{t=1}^{\lfloor ns \rfloor} E_{\mathcal{F}_{nt-1}} (\|\xi_{nt}\|^2 \mathbf{1} \{\|\xi_{nt}\| > \delta\}) \leq \max_{1 \leq t \leq n} E_{\mathcal{F}_{nt-1}} (\|\varepsilon_t\|^2 \mathbf{1} \{\|\xi_{nt}\| > \delta\}) O_p(1)$$

because $n^{-1-\beta} \sum_{t=1}^{\lfloor ns \rfloor} \|z_{t-1}\|^2 = O_p(1)$. Therefore,

$$\max_{1 \leq t \leq n} E_{\mathcal{F}_{nt-1}} (\|\varepsilon_t\|^2 \mathbf{1} \{\|\xi_{nt}\| > \delta\}) = o_p(1) \quad (34)$$

is sufficient for (31). To show (34), using (33) and the fact that ε_t is an i.i.d. sequence with $E \|\varepsilon_1\|^2 < \infty$ we obtain

$$\begin{aligned} \mathbf{1} \{\|\xi_{nt}\| > \delta\} &\leq \mathbf{1} \left\{ \frac{\|z_{t-1}\|}{n^{\frac{1}{2} + \frac{\beta}{4}}} > \frac{\delta'}{2} \right\} + \mathbf{1} \left\{ \frac{1}{n^{\frac{1}{2} - \frac{\beta}{4}}} > \frac{\delta'}{2} \right\} + \mathbf{1} \left\{ \frac{\|\varepsilon_t\|}{n^{\beta/4}} > \sqrt{\delta} \right\} \\ \max_{1 \leq t \leq n} E_{\mathcal{F}_{nt-1}} (\|\varepsilon_t\|^2 \mathbf{1} \{\|\xi_{nt}\| > \delta\}) &\leq \max_{1 \leq t \leq n} \mathbf{1} \left\{ \frac{\|z_{t-1}\|}{n^{\frac{1}{2} + \frac{\beta}{4}}} > \frac{\delta'}{2} \right\} E \|\varepsilon_1\|^2 \\ &\quad + E \left(\|\varepsilon_1\|^2 \mathbf{1} \left\{ \|\varepsilon_1\| > n^{\beta/4} \sqrt{\delta} \right\} \right) + o(1) \\ &\leq \mathbf{1} \left\{ \frac{1}{n^{\frac{1}{2} + \frac{\beta}{4}}} \max_{1 \leq t \leq n} \|z_{t-1}\| > \frac{\delta'}{2} \right\} E \|\varepsilon_1\|^2 + o(1) \\ &= o_p(1) \end{aligned}$$

since equation (53) of MP shows that $n^{-\frac{1}{2} - \frac{\beta}{4}} \max_{1 \leq t \leq n} \|z_{t-1}\| = o_p(1)$. This establishes (31).

Letting $Q = F(1) \Sigma F_0(1)'$, the conditional variance of the martingale array $\sum_{t=1}^{\lfloor ns \rfloor} \xi_{nt}$ is given by

$$\begin{aligned} \sum_{t=1}^{\lfloor ns \rfloor} E_{\mathcal{F}_{nt-1}} \xi_{nt} \xi_{nt}' &= \begin{bmatrix} \left(\frac{1}{n^{1+\beta}} \sum_{t=1}^{\lfloor ns \rfloor} z_{t-1} z_{t-1}' \right) \otimes \Omega_{00} & \left(\frac{1}{n^{1+\frac{\beta}{2}}} \sum_{t=1}^{\lfloor ns \rfloor} z_{t-1} \right) \otimes Q' \\ \left(\frac{1}{n^{1+\frac{\beta}{2}}} \sum_{t=1}^{\lfloor ns \rfloor} z_{t-1}' \right) \otimes Q & \frac{\lfloor ns \rfloor}{n} \Omega \end{bmatrix} \\ &\rightarrow_p s \begin{bmatrix} V_{zz} \otimes \Omega_{00} & 0 \\ 0 & \Omega \end{bmatrix}, \end{aligned} \quad (35)$$

because we know from MP that, for all $s \in [0, 1]$, and

$$n^{-1-\beta} \sum_{t=1}^{\lfloor ns \rfloor} z_t z_t' \rightarrow_p s V_{zz} \quad \text{and} \quad \sum_{t=1}^{\lfloor ns \rfloor} z_{t-1} = O_p \left(n^{\frac{1}{2} + \beta} \right).$$

Therefore, applying Theorem 3.33 VIII of Jacod and Shiryaev (1987) to $\sum_{t=1}^{\lfloor ns \rfloor} \xi_{nt}$, there exists a continuous Gaussian martingale $\xi(s)$ with quadratic variation

$$\langle \xi \rangle_s = s \operatorname{diag} (V_{zz} \otimes \Omega_{00}, \Omega)$$

such that $\sum_{t=1}^{\lfloor ns \rfloor} \xi_{nt} \Rightarrow \xi(s)$ on the Skorokhod space $D_{\mathbb{R}^{mK_z + m + K + K_z}} [0, 1]$. By Levy's characterisation of Brownian motion (e.g. Theorem 4.4 II of Jacod and Shiryaev (1987)), $\xi(s)$ is a Brownian motion on $D_{\mathbb{R}^{mK_z + m + K + K_z}} [0, 1]$ with covariance matrix $\operatorname{diag}(V_{zz} \otimes \Omega_{00}, \Omega)$. Partitioning $\xi(s) = [U(s)', B(s)']'$ conformably with its covariance matrix we conclude that U is independent of B .

Proof of part (ii). Letting

$$B_{nx}(s) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor ns \rfloor} u_{xj}, \quad W_n(s) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor ns \rfloor} \varepsilon_j,$$

Proposition A1 (ii) implies that

$$\left(B_{nx}(s), \begin{bmatrix} U_n(s) \\ W_n(s) \end{bmatrix} \right) \Rightarrow \left(B_x(s), \begin{bmatrix} U(s) \\ W(s) \end{bmatrix} \right) \quad \text{on } D_{\mathbb{R}^K \times \mathbb{R}^{(m+1)K_z + K}} [0, 1]. \quad (36)$$

Since $x_{\lfloor n \cdot \rfloor}$ can be represented as a continuous functional of $B_{nx}(\cdot)$, viz.

$$\frac{1}{\sqrt{n}} x_{\lfloor ns \rfloor} = B_{nx}(s) - C \int_0^s e^{C(s-r)} B_{nx}(s) dr + o_p(1),$$

and $n^{-1/2} x_{\lfloor ns \rfloor} \Rightarrow J_C(s)$ on $D_{\mathbb{R}^K} [0, 1]$, (36) and the continuous mapping theorem imply that

$$\left(\frac{1}{\sqrt{n}} x_{\lfloor ns \rfloor}, \begin{bmatrix} U_n(s) \\ W_n(s) \end{bmatrix} \right) \Rightarrow \left(J_C(s), \begin{bmatrix} U(s) \\ W(s) \end{bmatrix} \right) \quad \text{on } D_{\mathbb{R}^K \times \mathbb{R}^{(m+1)K_z + K}} [0, 1]. \quad (37)$$

The identity

$$\text{vec} \int_0^1 \frac{x_{\lfloor ns \rfloor}}{\sqrt{n}} dW_n(s)' = \int_0^1 \left\{ I_{K_z} \otimes \frac{x_{\lfloor ns \rfloor}}{\sqrt{n}} \right\} dW_n(s)$$

yields

$$\begin{aligned} \begin{bmatrix} U_n(1) \\ \text{vec} \left\{ n^{-1} \sum_{t=1}^n x_{t-1} \varepsilon_t' \right\} \\ \text{vec} \left\{ n^{-1} \sum_{t=1}^n x_{t-1} x_{t-1}' \right\} \end{bmatrix} &= \begin{bmatrix} \int_0^1 dU_n(s) \\ \int_0^1 \left\{ I_{K_z} \otimes \frac{x_{\lfloor ns \rfloor}}{\sqrt{n}} \right\} dW_n(s) \\ \int_0^1 \left(\frac{x_{\lfloor ns \rfloor}}{\sqrt{n}} \otimes \frac{x_{\lfloor ns \rfloor}}{\sqrt{n}} \right) ds \end{bmatrix} + o_p(1) \\ &= \int_0^1 G_n(s) dV_n(s), \end{aligned} \quad (38)$$

where

$$G_n(s) = \text{diag} \left(I_{mK_z}, I_{K_z} \otimes \frac{x_{\lfloor ns \rfloor}}{\sqrt{n}}, \frac{x_{\lfloor ns \rfloor}}{\sqrt{n}} \otimes \frac{x_{\lfloor ns \rfloor}}{\sqrt{n}} \right) \quad \text{and} \quad V_n(s) = [U_n(s)', W_n(s)', s]'$$

Letting

$$G(s) = \text{diag} [I_{mK_z}, I_{K_z} \otimes J_C(s), J_C(s) \otimes J_C(s)] \quad \text{and} \quad V(s) = [U(s)', W(s)', s]',$$

joint convergence $[G_n(s), V_n(s)] \Rightarrow [G(s), V(s)]$ on the relevant Skorokhod space is guaranteed by (37). Note that $V_n(s)$ is a semimartingale with Doob-Meyer decomposition $V_n(s) = M_n(s) + A(s)$ where $M_n(s) = [U_n(s)', W_n(s)', 0]'$ is a martingale

and $A(s) = [0, 0, s]'$ has bounded variation for all $s \in [0, 1]$. Thus, Theorem 2.7 of Kurtz and Protter (1991) implies that

$$\int_0^1 G_n(s) dV_n(s) \Rightarrow \int_0^1 G(s) dV(s) \quad (39)$$

obtains, provided that $\sup_n E[M_n]_s < \infty$. The later condition holds trivially, since

$$E[M_n]_s = \text{diag} \left[\left(n^{-1-\beta} \sum_{t=1}^{\lfloor ns \rfloor} E z_{t-1} z'_{t-1} \right) \otimes \Omega_{00}, \Sigma, 0 \right]$$

and $\sup_n n^{-1-\beta} \left\| \sum_{t=1}^{\lfloor ns \rfloor} E z_{t-1} z'_{t-1} \right\| \leq \sup_n n^{-1-\beta} \sum_{t=1}^n E \|z_{t-1}\|^2 < \infty$ for all $s \in [0, 1]$.

Now part (ii) can be established by combining (38) and (39).

Proposition A2. *Under Assumptions N and LP, the process ψ_{nt} in (12) satisfies*

$$\sup_{1 \leq t \leq n} E \|\psi_{nt}\|^2 = O\left(n^{(\alpha \vee \beta) + 2(\alpha \wedge \beta)}\right) \quad \text{as } n \rightarrow \infty \quad (40)$$

and, for any $\beta \in (1/2, 1)$ with $\beta \neq \alpha$,

$$\frac{1}{n^{\frac{1+(\alpha \wedge \beta)}{2} + (\alpha \vee \beta)}} \sum_{t=1}^n u_t \psi'_{nt} \rightarrow_p 0 \quad \text{as } n \rightarrow \infty. \quad (41)$$

When $\beta = \alpha \in (1/2, 1)$, the left side of (41) is bounded in probability.

Proof. For the sake of clarity, we differentiate between the Euclidian matrix norm $\|M\|_E = (\text{tr} M' M)^{1/2}$ and the spectral norm $\|M\|$ defined as the square root of the maximal eigenvalue of $M' M$. Recall that

$$x_{j-1} = R_n^{j-1} x_0 + \sum_{k=2}^j R_n^{j-k} u_{xk-1}$$

and denote the autocovariance matrix of u_{xt} by $\Gamma_{u_x}(h) = E(u_{xt} u'_{xt-h})$. Since

$$E(x_j x'_i) = \sum_{k=2}^j \sum_{l=2}^i R_n^{j-k} \Gamma_{u_x}(k-l) R_n^{i-l},$$

using the trace version of the Cauchy Schwarz inequality (see e.g. Abadir and Magnus, 2005) we obtain

$$\begin{aligned}
E \|\psi_{nt}\|^2 &= \text{tr} \left\{ \sum_{i,j=1}^t R_{nz}^{t-j} R_{nz}^{t-i} E(x_{j-1} x'_{i-1}) \right\} \\
&= \sum_{i,j=1}^t \sum_{k=2}^j \sum_{l=2}^i \text{tr} \{ R_{nz}^{2t-j-i} R_n^{j+i-k-l} \Gamma_{u_x}(k-l) \} \\
&\leq \sum_{i,j=1}^t \sum_{k=2}^j \sum_{l=2}^i \|R_{nz}^{2t-j-i} R_n^{j+i-k-l}\|_E \|\Gamma_{u_x}(k-l)\|_E \\
&\leq \sum_{i,j=1}^t \sum_{k=2}^j \sum_{l=2}^i \|R_{nz}^{2t-j-i}\|_E \|R_n^{j+i-k-l}\|_E \|\Gamma_{u_x}(k-l)\|_E \\
&\leq K \sum_{i,j=1}^t \sum_{k=2}^j \sum_{l=2}^i \|R_{nz}\|^{2t-j-i} \|R_n\|^{j+i-k-l} \|\Gamma_{u_x}(k-l)\|_E \quad (42)
\end{aligned}$$

where the last inequality follows from the fact that R_n and R_{nz} are diagonal $K \times K$ matrices, so $\|R_n^m\|_E \leq K^{1/2} \|R_n\|^m$ for any $m \in \mathbb{Z}$ and the same inequality holds for R_{nz} . In order to estimate the right side (42) we make use of the bounds

$$\sup_{1 \leq t \leq n} \sum_{j=1}^t \|R_n\|^{t-j} = O(n^{\alpha \wedge 1}) \quad \text{and} \quad \sup_{1 \leq t \leq n} \sum_{j=1}^t \|R_{nz}\|^{t-j} = O(n^\beta) \quad (43)$$

and of the fact that Assumption LP guarantees $\sum_{h \in \mathbb{Z}} \|\Gamma_{u_x}(h)\|_E < \infty$.

We consider first the case $\beta < \alpha$. Using the fact that $\|R_n\|^{i-l} \leq 1$ for all $l \leq i$, (42) and (43) yield

$$\begin{aligned}
\sup_{1 \leq t \leq n} E \|\psi_{nt}\|^2 &\leq K \left(\sup_{1 \leq t \leq n} \sum_{i=1}^t \|R_{nz}\|^{t-i} \right)^2 \left(\sup_{1 \leq j \leq n} \sum_{k=1}^j \|R_n\|^{j-k} \right) \sum_{l=-\infty}^{\infty} \|\Gamma_{u_x}(l)\|_E \\
&= O(n^{\alpha+2\beta}).
\end{aligned}$$

When $\beta \geq \alpha$, letting $j - k = m$ in (42) we obtain that

$$E \|\psi_{nt}\|^2 \leq K \sum_{i,j=1}^t \sum_{m=0}^{j-2} \sum_{l=1}^i \|R_{nz}\|^{2t-j-i} \|R_n\|^{m+i-l} \|\Gamma_{u_x}(j-m-l)\|_E.$$

Using the fact that $\|R_{nz}\|^{t-j} \leq 1$ for all $j \leq t$,

$$\begin{aligned}
\sup_{1 \leq t \leq n} E \|\psi_{nt}\|^2 &\leq K \left(\sup_{1 \leq t \leq n} \sum_{i=1}^t \|R_{nz}\|^{t-i} \right) \left(\sum_{m=0}^{\infty} \|R_n\|^m \right)^2 \sum_{j=-\infty}^{\infty} \|\Gamma_{u_x}(j)\|_E \\
&= O(n^{\beta+2\alpha})
\end{aligned}$$

and (40) follows.

To show (41), note that ψ_{nt} in (12) satisfies the recursive formula

$$\psi_{nt} = R_{nz}\psi_{n,t-1} + x_{t-1} \quad (44)$$

and using the BN decomposition

$$\frac{1}{n^{\frac{1+(\alpha\wedge\beta)}{2}+(\alpha\vee\beta)}} \sum_{t=1}^n u_t \psi'_{nt} = \frac{1}{n^{\frac{1+(\alpha\wedge\beta)}{2}+(\alpha\vee\beta)}} \left\{ F(1) \sum_{t=1}^n \varepsilon_t \psi'_{nt} - \sum_{t=1}^n \Delta \tilde{\varepsilon}_t \psi'_{nt} \right\}. \quad (45)$$

Since ψ_{nt} is $\sigma(x_0, \dots, x_{t-1})$ -measurable, the first term of (45) is a matrix martingale array with

$$\begin{aligned} E \left\| \frac{1}{n^{\frac{1+(\alpha\wedge\beta)}{2}+(\alpha\vee\beta)}} \sum_{t=1}^n (\psi_{nt} \otimes \varepsilon_t) \right\|^2 &= \frac{E \|\varepsilon_1\|^2}{n^{1+\alpha\wedge\beta+2(\alpha\vee\beta)}} \sum_{t=1}^n E \|\psi_{nt}\|^2 \\ &\leq \frac{E \|\varepsilon_1\|^2}{n^{\alpha\wedge\beta+2(\alpha\vee\beta)}} \sup_{1 \leq t \leq n} E \|\psi_{nt}\|^2 \\ &= O\left(\frac{n^{\alpha\wedge\beta}}{n^{\alpha\vee\beta}}\right) \end{aligned}$$

by (40). The last order is $o(1)$ when $\beta \neq \alpha$ and exactly $O(1)$ when $\beta = \alpha$.

Next, we show that, under the condition $\beta \in (1/2, 1)$, the second term of (45) is $o_p(1)$ both when $\beta \neq \alpha$ and when $\beta = \alpha$. Summation by parts, (40) and (44) yield

$$\begin{aligned} \sum_{t=1}^n \Delta \tilde{\varepsilon}_t \psi'_{nt} &= - \sum_{t=1}^n \tilde{\varepsilon}_t \Delta \psi'_{nt+1} + O_p\left(n^{\frac{\alpha\vee\beta}{2}+(\alpha\wedge\beta)}\right) \\ &= - \frac{1}{n^\beta} \sum_{t=1}^n \tilde{\varepsilon}_t \psi'_{nt} C'_z - \sum_{t=1}^n \tilde{\varepsilon}_t x'_t + O_p\left(n^{\frac{\alpha\vee\beta}{2}+(\alpha\wedge\beta)}\right) \\ &= - \frac{1}{n^\beta} \sum_{t=1}^n \tilde{\varepsilon}_t \psi'_{nt} C'_z + O_p(n) + O_p\left(n^{\frac{\alpha\vee\beta}{2}+(\alpha\wedge\beta)}\right) \end{aligned}$$

since $\sum_{t=1}^n \tilde{\varepsilon}_t x'_t = O_p(n)$ for all $\alpha > 0$ under Assumptions N and LP by Phillips (1987) and MP. Now $\beta \in (1/2, 1)$ implies that $n^{1-\frac{1+(\alpha\wedge\beta)}{2}-(\alpha\vee\beta)} \rightarrow 0$, so

$$\frac{1}{n^{\frac{1+(\alpha\wedge\beta)}{2}+(\alpha\vee\beta)}} \sum_{t=1}^n \Delta \tilde{\varepsilon}_t \psi'_{nt} = - \frac{1}{n^{\frac{1+(\alpha\wedge\beta)}{2}+(\alpha\vee\beta)}} \frac{1}{n^\beta} \sum_{t=1}^n \tilde{\varepsilon}_t \psi'_{nt} C'_z + o_p(1).$$

The last term on the right converges to 0 in L_1 for any $\beta > 1/2$ since, letting

$\kappa = E \|\tilde{\varepsilon}_1\|^2$, the Cauchy Schwarz inequality and (40) yield

$$\begin{aligned}
\left\| \frac{1}{n^{\frac{1+(\alpha\wedge\beta)}{2}+(\alpha\vee\beta)}} \frac{1}{n^\beta} \sum_{t=1}^n \tilde{\varepsilon}_t \psi'_{nt} \right\|_{L_1} &\leq \frac{1}{n^{\frac{1+(\alpha\wedge\beta)}{2}+(\alpha\vee\beta)}} \frac{1}{n^\beta} \sum_{t=1}^n E (\|\psi_{nt}\| \|\tilde{\varepsilon}_t\|) \\
&\leq \frac{\kappa^{1/2}}{n^{\frac{1+(\alpha\wedge\beta)}{2}+(\alpha\vee\beta)}} \frac{1}{n^\beta} \sum_{t=1}^n (E \|\psi_{nt}\|^2)^{1/2} \\
&\leq \kappa^{1/2} \frac{n^{1/2}}{n^{\frac{(\alpha\wedge\beta)}{2}+(\alpha\vee\beta)+\beta}} \left(\sup_{1 \leq t \leq n} E \|\psi_{nt}\|^2 \right)^{1/2} \\
&= O \left(\frac{1}{n^{\beta-1/2}} \frac{n^{\frac{\alpha\vee\beta}{2}+(\alpha\wedge\beta)}}{n^{\frac{(\alpha\wedge\beta)}{2}+(\alpha\vee\beta)}} \right) = O \left(\frac{1}{n^{\beta-1/2}} \right).
\end{aligned}$$

This shows that the second term of (45) is $o_p(1)$ and establishes (41).

Proof of Lemma 3.1. Using the fact that $\sup_{t \in [0,1]} \|x_{[nt]}\| = O_p(n^{(\alpha\wedge 1)/2})$ and (43) we obtain the following uniform bound for $\|\psi_{nt}\|$:

$$\sup_{1 \leq t \leq n} \|\psi_{nt}\| \leq \sup_{1 \leq j \leq n} \|x_{j-1}\| \sup_{1 \leq t \leq n} \sum_{j=1}^t \|R_{nz}\|^{t-j} = O_p(n^{\alpha/2+\beta}). \quad (46)$$

For part (i), using (13) we obtain

$$\frac{1}{n^{\frac{1+\beta}{2}}} \left(\sum_{t=1}^n u_{0t} \tilde{z}'_t - \sum_{t=1}^n u_{0t} z'_t \right) = \frac{1}{n^{\frac{1+\beta}{2}+\alpha}} \sum_{t=1}^n u_{0t} \psi'_{nt} C = o_p(1)$$

by (41) since $\beta < \alpha$.

For part (ii), using the recursive formulae (2) and (44) we obtain

$$x_t \psi'_{nt} = R_n x_{t-1} \psi'_{n,t-1} R_{nz} + R_n x_{t-1} x'_{t-1} + u_{xt} \psi'_{n,t-1} R_{nz} + u_{xt} x'_{t-1}.$$

Vectorising and summing along $t \in \{1, \dots, n\}$ we obtain

$$\begin{aligned}
& [I_{K^2} - R_{nz} \otimes R_n] \sum_{t=1}^n (\psi_{n,t-1} \otimes x_{t-1}) \\
&= \psi_{n,0} \otimes x_0 - \psi_{n,n} \otimes x_n + (I_K \otimes R_n) \sum_{t=1}^n (x_{t-1} \otimes x_{t-1}) \\
&\quad + (R_{nz} \otimes I_K) \sum_{t=1}^n (\psi_{n,t-1} \otimes u_{xt}) + (I_K \otimes R_n) \sum_{t=1}^n (x_{t-1} \otimes u_{xt}) \\
&= (I_K \otimes R_n) \sum_{t=1}^n (x_{t-1} \otimes x_{t-1}) + o_p \left(n^{\frac{1+\beta}{2}+\alpha} \right) + O_p(n),
\end{aligned}$$

since $\sum_{t=1}^n (\psi_{n,t-1} \otimes u_{xt}) = o_p \left(n^{\frac{1+\beta}{2} + \alpha} \right)$ by (41) and $\sum_{t=1}^n (x_{t-1} \otimes u_{xt}) = O_p(n)$ for any $\alpha > 0$ (see equation (10) of MP). Since

$$I_{K^2} - R_{nz} \otimes R_n = \frac{1}{n^\beta} \left[-C_z \otimes I_K + O \left(\frac{1}{n^{\alpha-\beta}} \right) \right],$$

multiplying the above expression by $n^{-1} (n^{-\alpha} C \otimes I_K)$ yields

$$\frac{1}{n^{1+\beta}} \sum_{t=1}^n \left(\frac{C}{n^\alpha} \psi_{n,t-1} \otimes x_{t-1} \right) = - (C_z^{-1} C \otimes I_K) \frac{1}{n^{1+\alpha}} \sum_{t=1}^n (x_{t-1} \otimes x_{t-1}) + o_p(1).$$

The result now follows by undoing the vectorisation and using (13).

For part (iii), (13) yields

$$\begin{aligned} \frac{1}{n^{1+\beta}} \left\| \sum_{t=1}^n \tilde{z}_t \tilde{z}_t' - \sum_{t=1}^n z_t z_t' \right\| &= \frac{1}{n^{1+\beta+\alpha}} \left\| \frac{1}{n^\alpha} \sum_{t=1}^n C \psi_{nt} \psi_{nt}' C + \sum_{t=1}^n z_t \psi_{nt}' C + C \sum_{t=1}^n \psi_{nt} z_t' \right\| \\ &\leq \frac{\|C\|^2}{n^{1+\beta+2\alpha}} \sum_{t=1}^n \|\psi_{nt}\|^2 + \frac{2\|C\|}{n^{1+\beta+\alpha}} \sum_{t=1}^n \|\psi_{nt}\| \|z_t\| \\ &\leq \|C\|^2 \left(\frac{\sup_{1 \leq t \leq n} \|\psi_{nt}\|}{n^{\beta/2+\alpha}} \right)^2 + \left(\frac{\sup_{1 \leq t \leq n} \|\psi_{nt}\|}{n^{\beta/2+\alpha}} \right) O_p(1) \\ &= o_p(1) \end{aligned}$$

by (46) since the Lyapounov inequality gives

$$E \left(\frac{1}{n^{1+\beta/2}} \sum_{t=1}^n \|z_t\| \right) \leq \left(\frac{1}{n^\beta} \sup_{1 \leq t \leq n} E \|z_t\|^2 \right)^{1/2} = O(1).$$

Proof of Theorem 3.4. For any $\beta \in (2/3, 1)$, Lemma 2.1 and (15) yield

$$\begin{aligned} n^{\frac{1+\beta}{2}} (\tilde{A}_n - A) &= \frac{1}{n^{\frac{1+\beta}{2}}} \left(U_0' \tilde{Z} - n \hat{\Delta}_{0x} \right) \left(\frac{1}{n^{1+\beta}} X' \tilde{Z} \right)^{-1} \\ &= \frac{1}{n^{\frac{1+\beta}{2}}} (U_0' Z - n \Delta_{0x}) \left(\frac{1}{n^{1+\beta}} X' \tilde{Z} \right)^{-1} + O_p \left(n^{\frac{1-\beta}{2}} (\hat{\Delta}_{0x} - \Delta_{0x}) \right) \\ &= U_n(1) \left(\frac{1}{n^{1+\beta}} X' \tilde{Z} \right)^{-1} + o_p(1). \end{aligned}$$

For $n^{-1-\beta} \tilde{Z}' X$, Lemma 3.1(ii) shows that $n^{-1-\beta} \tilde{Z}' X = n^{-1-\beta} Z' X - L_n$, where

$$L_n = \frac{1}{n^{1+\alpha}} \sum_{t=1}^n x_{t-1} x_{t-1}' C C_z^{-1},$$

and the asymptotic behavior of $n^{-1-\beta}Z'X$ is given by (20). Under Assumption N(i), $L_n = o_p(1)$. Under N(ii), $L_n \Rightarrow \int_0^1 J_C J'_C ds C C_z^{-1}$, giving

$$\begin{aligned} \frac{1}{n^{1+\beta}} \tilde{Z}'X &\Rightarrow - \left(\int_0^1 J_C dB'_x + \Omega_{xx} \right) C_z^{-1} - \int_0^1 J_C J'_C ds C C_z^{-1} \\ &= - \left(\Omega_{xx} + \int_0^1 J_C dJ'_C \right) C_z^{-1}. \end{aligned}$$

Under N(iii), equation (7) in MP gives $L_n \rightarrow_p V_{xx} C C_z^{-1}$. Thus, in all of the above cases, $n^{-1-\beta} \tilde{Z}'X \Rightarrow -\tilde{\Psi}_{xx} C_z^{-1}$. Now, by Lemma 3.1(ii), (17), (18) and Lemma 3.2, $n^{-1-\beta} \tilde{Z}'X$ and $U_n(1)$ converge jointly in distribution and are asymptotically independent. Thus,

$$n^{\frac{1+\beta}{2}} \text{vec} \left(\tilde{A}_n - A \right) = \left[\left(\frac{1}{n^{1+\beta}} \tilde{Z}'X \right)^{-1} \otimes I_m \right] U_n(1) + o_p(1)$$

has the required mixed Gaussian limit distribution.

Proof of Lemma 3.5. For part (i), we can use (23) to write

$$\begin{aligned} \frac{1}{n^{\frac{1+\alpha}{2}}} \left\{ \sum_{t=1}^n u_{0t} \tilde{z}'_t - \sum_{t=1}^n u_{0t} x'_t \right\} &= \frac{1}{n^{\frac{1+\alpha}{2}}} \sum_{t=1}^n u_{0t} \left(\frac{C_z}{n^\beta} \psi_{nt} - R_{nz}^t x_0 \right)' \\ &= \frac{1}{n^{\frac{1+\alpha}{2} + \beta}} \sum_{t=1}^n u_{0t} \psi'_{nt} C_z + o_p \left(\frac{1}{n^{\frac{1-\beta}{2}}} \right), \end{aligned}$$

because $x_0 = o_p(n^{\alpha/2})$ and $\sum_{t=1}^n (R_{nz}^{t-1} \otimes u_{0t}) = O_p(n^{\beta/2})$ by MP. The leading term in the above display is $o_p(1)$ by (41) since $\alpha < \beta$.

For part (ii), we make repeated use of the decomposition (23) and Proposition A2. By (23),

$$\begin{aligned} \frac{1}{n^{1+\alpha}} \left(\sum_{t=1}^n x_t \tilde{z}'_t - \sum_{t=1}^n x_t x'_t \right) &= \frac{1}{n^{1+\alpha}} \left[\frac{1}{n^\beta} \sum_{t=1}^n x_t \psi'_{nt} C_z - \sum_{t=1}^n x_t x'_0 R_{nz}^t \right] \\ &= \frac{1}{n^{1+\alpha+\beta}} \sum_{t=1}^n x_t \psi'_{nt} C_z + o_p \left(\frac{1}{n^{1-\beta}} \right), \end{aligned}$$

since $\sup_{1 \leq t \leq n} \|x_t\| = O_p(n^{\alpha/2})$ implies that

$$\left\| \frac{1}{n^{1+\alpha}} \sum_{t=1}^n x_t x'_0 R_{nz}^t \right\| \leq \frac{\|x_0\|}{n^{\alpha/2}} \frac{1}{n^{1+\alpha/2}} \sup_{1 \leq t \leq n} \|x_t\| \sum_{t=1}^n \|R_{nz}\|^t = \frac{\|x_0\|}{n^{\alpha/2}} O_p \left(\frac{1}{n^{1-\beta}} \right).$$

For the leading term, we know from MP that $\sup_{1 \leq t \leq n} E(\|x_t\|^2) = O(n^\alpha)$. Thus, the Cauchy Schwarz inequality and (40) yield

$$\begin{aligned}
\left\| \frac{1}{n^{1+\alpha+\beta}} \sum_{t=1}^n (\psi_{nt} \otimes x_t) \right\|_{L_1} &\leq \frac{1}{n^{1+\alpha+\beta}} \sum_{t=1}^n E(\|\psi_{nt}\| \|x_t\|) \\
&\leq \frac{1}{n^{1+\alpha+\beta}} \sum_{t=1}^n (E\|\psi_{nt}\|^2)^{1/2} (E\|x_t\|^2)^{1/2} \\
&\leq \frac{1}{n^{\alpha+\beta}} \left\{ \sup_{1 \leq t \leq n} E\|\psi_{nt}\|^2 \sup_{1 \leq j \leq n} E\|x_j\|^2 \right\}^{1/2} \\
&= O\left(\frac{1}{n^{(\beta-\alpha)/2}}\right).
\end{aligned}$$

Since $\beta > \alpha$, this shows the result for $n^{-1-\alpha} \sum_{t=1}^n x_t \tilde{z}'_t$. It remains to show the result for $n^{-1-\alpha} \sum_{t=1}^n \tilde{z}_t \tilde{z}'_t$. By (23) and given the derivations for $n^{-1-\alpha} \sum_{t=1}^n x_t \tilde{z}'_t$, it is sufficient to prove that

$$\frac{1}{n^{1+\alpha}} \sum_{t=1}^n R_{nz}^t x_0 x_0' R_{nz}^t, \quad \frac{1}{n^{1+\alpha+\beta}} \sum_{t=1}^n \psi_{nt} x_0' R_{nz}^t \quad \text{and} \quad \frac{1}{n^{1+\alpha+2\beta}} \sum_{t=1}^n \psi_{nt} \psi_{nt}'$$

all converge to 0 in probability. The first term in the above display is clearly $O_p(n^{-(1-\beta)})$. For the second term, (46) gives

$$\begin{aligned}
\left\| \frac{1}{n^{1+\alpha+\beta}} \sum_{t=1}^n \psi_{nt} x_0' R_{nz}^t \right\| &\leq \frac{\|x_0\|}{n^{1+\alpha+\beta}} \sup_{1 \leq t \leq n} \|\psi_{nt}\| \sum_{t=1}^n \|R_{nz}^t\| \\
&= o_p\left(\frac{n^{\alpha/2+\beta}}{n^{1+\alpha/2}}\right) = o_p\left(\frac{1}{n^{1-\beta}}\right).
\end{aligned}$$

For the third term, (40) implies that

$$\begin{aligned}
E \left\| \frac{1}{n^{1+\alpha+2\beta}} \sum_{t=1}^n \psi_{nt} \psi_{nt}' \right\| &\leq \frac{1}{n^{1+\alpha+2\beta}} \sum_{t=1}^n E\|\psi_{nt}\|^2 \\
&\leq \frac{1}{n^{\alpha+2\beta}} \sup_{1 \leq t \leq n} E\|\psi_{nt}\|^2 = O\left(\frac{1}{n^{\beta-\alpha}}\right).
\end{aligned}$$

This completes the proof of the lemma.

Proof of Lemma 3.6. Proposition A2 and (13) imply that

$$\frac{1}{n} \sum_{t=1}^n (\tilde{z}_{t-1} \otimes u_t) = \frac{1}{n} \sum_{t=1}^n (z_{t-1} \otimes u_t) + o_p(1). \tag{47}$$

Note that, by construction, \tilde{z}_t satisfies the recursive formula

$$\tilde{z}_t = R_{nz}\tilde{z}_{t-1} + v_t, \quad v_t = u_{xt} + \frac{C}{n^\alpha}x_{t-1}. \quad (48)$$

For part (i), using (48) together with the recursive property of x_t , we obtain

$$x_t \tilde{z}'_t = R_n x_{t-1} \tilde{z}'_{t-1} R_{nz} + R_n x_{t-1} v'_t + u_{xt} \tilde{z}'_{t-1} R_{nz} + u_{xt} v'_t.$$

Vectorising and summing over $\{1, \dots, n\}$ we obtain

$$\begin{aligned} & (I_{K^2} - R_{zn} \otimes R_n) \frac{1}{n} \sum_{t=1}^n (\tilde{z}_{t-1} \otimes x_{t-1}) \\ &= [I_{K^2} + o_p(1)] \left\{ \frac{1}{n} \sum_{t=1}^n (v_t \otimes x_{t-1}) + \frac{1}{n} \sum_{t=1}^n (\tilde{z}_{t-1} \otimes u_{xt}) + \frac{1}{n} \sum_{t=1}^n (v_t \otimes u_{xt}) \right\} \\ &= \frac{1}{n} \sum_{t=1}^n (u_{xt} \otimes x_{t-1}) + \frac{1}{n^{1+\alpha}} \sum_{t=1}^n (C x_{t-1} \otimes x_{t-1}) + \frac{1}{n} \sum_{t=1}^n (z_{t-1} \otimes u_{xt}) \\ &\quad + \frac{1}{n} \sum_{t=1}^n (u_{xt} \otimes u_{xt}) + o_p(1) \\ &= \text{vec} \left\{ \frac{1}{n} \sum_{t=1}^n x_{t-1} u'_{xt} + \frac{1}{n} \sum_{t=1}^n u_{xt} z'_{t-1} + \frac{1}{n} \sum_{t=1}^n u_{xt} u'_{xt} + \frac{1}{n^{1+\alpha}} \sum_{t=1}^n x_{t-1} x'_{t-1} C \right\} \end{aligned}$$

where the second asymptotic equivalence follows from (47) and $\sum_{t=1}^n x_{t-1} u'_{xt} = O_p(n)$. Now x_t and z_t are both mildly integrated processes with innovations u_{xt} , so Lemma 3.1(d) of MP implies that $n^{-1} \sum_{t=1}^n u_{xt} z'_{t-1} \rightarrow_p \Lambda_{xx}$ and $n^{-1} \sum_{t=1}^n x_{t-1} u'_{xt} \rightarrow_p \Lambda'_{xx}$. Also, applying the integration by parts formula to $V_{xx} = \int_0^\infty e^{rC} \Omega_{xx} e^{rC} dr$, we obtain

$$C V_{xx} + V_{xx} C = -\Omega_{xx}. \quad (49)$$

Therefore,

$$\begin{aligned} \frac{1}{n^{1+\alpha}} \sum_{t=1}^n (\tilde{z}_{t-1} \otimes x_{t-1}) &\rightarrow_p - (C_z \otimes I_K + I_K \otimes C)^{-1} \text{vec} (\Omega_{xx} + V_{xx} C) \\ &= (C_z \otimes I_K + I_K \otimes C)^{-1} \text{vec} (C V_{xx}) \\ &= - \int_0^\infty (e^{rC_z} \otimes e^{rC}) dr \text{vec} (C V_{xx}) \\ &= \text{vec} (-C \mathbb{V}_{xz}) \end{aligned}$$

which shows part (i).

For part (ii), (48) yields

$$(I_{K^2} - R_{zn} \otimes R_{nz}) \frac{1}{n} \sum_{t=1}^n (\tilde{z}_{t-1} \otimes \tilde{z}_{t-1}) = [I_{K^2} + o_p(1)] \\ \times \frac{1}{n} \text{vec} \left\{ \sum_{t=1}^n v_t \tilde{z}'_{t-1} + \sum_{t=1}^n \tilde{z}_{t-1} v'_t + \sum_{t=1}^n v_t v'_t \right\}.$$

Now since $\sum_{t=1}^n x_{t-1} u'_{xt} = O_p(n)$ and $\sum_{t=1}^n x_{t-1} x'_{t-1} = O_p(n^{1+\alpha})$,

$$\frac{1}{n} \sum_{t=1}^n v_t v'_t = \frac{1}{n} \sum_{t=1}^n u_{xt} u'_{xt} + O_p(n^{-\alpha}) \rightarrow_p E(u_{x1} u'_{x1}).$$

Also, using (47) and part (i),

$$\frac{1}{n} \sum_{t=1}^n \tilde{z}_{t-1} v'_t = \frac{1}{n} \sum_{t=1}^n \tilde{z}_{t-1} u'_{xt} + \frac{1}{n^{1+\alpha}} \sum_{t=1}^n \tilde{z}_{t-1} x'_{t-1} C \\ = \frac{1}{n} \sum_{t=1}^n z_{t-1} u'_{xt} - \mathbb{V}'_{xz} C^2 + o_p(1) \\ = \Lambda'_{xx} - \mathbb{V}'_{xz} C^2 + o_p(1).$$

Collecting the above asymptotic results we obtain

$$\frac{1}{n^{1+\alpha}} \sum_{t=1}^n (\tilde{z}_{t-1} \otimes \tilde{z}_{t-1}) \rightarrow_p - (C_z \otimes I_K + I_K \otimes C_z)^{-1} \text{vec} \{ \Omega_{xx} - C^2 \mathbb{V}_{xz} - \mathbb{V}'_{xz} C^2 \}.$$

Applying the integration by parts formula to $\mathbb{V}_{xz} = \int_0^\infty e^{rC} V_{xx} e^{rC_z} dr$ and using (49)

$$C^2 \mathbb{V}_{xz} + \mathbb{V}'_{xz} C^2 = -V_{xx} C - C_z \mathbb{V}'_{xz} C - C V_{xx} - C \mathbb{V}_{xz} C_z \\ = \Omega_{xx} - (C_z \mathbb{V}'_{xz} C + C \mathbb{V}_{xz} C_z).$$

Thus,

$$\frac{1}{n^{1+\alpha}} \sum_{t=1}^n (\tilde{z}_{t-1} \otimes \tilde{z}_{t-1}) \rightarrow_p \left(\int_0^\infty e^{sC_z} \otimes e^{sC_z} ds \right) \text{vec} (C_z \mathbb{V}'_{xz} C + C \mathbb{V}_{xz} C_z) \\ = \text{vec} \int_0^\infty e^{sC_z} (C_z \mathbb{V}'_{xz} C + C \mathbb{V}_{xz} C_z) e^{sC_z} ds,$$

and part (ii) follows.

For part (iii), using (48), (13), the fact that $\sum_{t=1}^n u_{0t}x'_{t-1} = O_p(n)$ and the BN decomposition we obtain

$$\begin{aligned}
\sum_{t=1}^n u_{0t}\tilde{z}'_t &= \sum_{t=1}^n u_{0t}\tilde{z}'_{t-1}R_{zn} + \sum_{t=1}^n u_{0t}u'_{xt} + O_p(n^{1-\alpha}) \\
&= F_0(1) \sum_{t=1}^n \varepsilon_t \tilde{z}'_{t-1} R_{zn} + \sum_{t=1}^n u_{0t}u'_{xt} - \sum_{t=1}^n \Delta \tilde{\varepsilon}_{0t} \tilde{z}'_{t-1} R_{zn} \\
&\quad - \frac{1}{n^\alpha} \sum_{t=1}^n \Delta \tilde{\varepsilon}_{0t} \psi'_{nt-1} C R_{zn} + O_p(n^{1-\alpha}). \tag{50}
\end{aligned}$$

As shown in the proof of Proposition A2 (where asymptotic negligibility of the second term of (45) was established) $n^{-\alpha} \sum_{t=1}^n \Delta \tilde{\varepsilon}_{0t} \psi'_{nt-1} = o_p\left(n^{\frac{1+\alpha}{2}}\right)$ for any $\alpha > 1/2$. For the third term of (50), summation by parts yields

$$\begin{aligned}
\sum_{t=1}^n \Delta \tilde{\varepsilon}_{0t} \tilde{z}'_{t-1} &= \sum_{t=1}^n \tilde{\varepsilon}_{0t} \Delta \tilde{z}'_t + O_p(n^{\beta/2}) = \sum_{t=1}^n \tilde{\varepsilon}_{0t} u'_{xt} + \frac{1}{n^\beta} \sum_{t=1}^n \tilde{\varepsilon}_{0t} \tilde{z}'_{t-1} C_z + O_p(n^{\beta/2}) \\
&= \sum_{t=1}^n \tilde{\varepsilon}_{0t} u'_{xt} + O_p(n^{1-\beta}) + O_p(n^{\beta/2})
\end{aligned}$$

by MP. Since $\alpha = \beta > 1/2$, and $\sum_{t=1}^n \{u_{0t}u'_{xt} - E(u_{0t}u'_{xt})\}$ and $\sum_{t=1}^n \{\tilde{\varepsilon}_{0t}u'_{xt} - \Lambda_{0x}\}$ both have rate $O_p(n^{-1/2})$ by the stationary ergodic CLT, (50) implies that

$$\frac{1}{n^{\frac{1+\alpha}{2}}} \sum_{t=1}^n \text{vec}(u_{0t}\tilde{z}'_t - \Delta_{0x}) = \frac{[I_K \otimes F_0(1)]}{n^{\frac{1+\alpha}{2}}} \sum_{t=1}^n (\tilde{z}_{t-1} \otimes \varepsilon_t) + o_p(1).$$

A standard martingale CLT implies that the right side of the above expression converges in distribution to

$$N\left(0, \left\{ \text{plim}_{n \rightarrow \infty} n^{-(1+\alpha)} \sum_{t=1}^n \tilde{z}_{t-1} \tilde{z}'_{t-1} \right\} \otimes \Omega_{00}\right)$$

which, in view of part (ii), yields the required limit distribution.

Proof of Lemma 4.2. Part (i) can be deduced by an identical method to that used in establishing (17) (18):

$$(I_{KK_z} - R_{zn} \otimes R_n) \frac{1}{n} \sum_{t=1}^n \text{vec}(x_{t-1}z'_{t-1}) = \text{vec}(J_n + \Omega_{xz}) + o_p(1).$$

When $\alpha > \beta$,

$$I_{KK_z} - R_{zn} \otimes R_n = -\frac{1}{n^\beta} (C_z \otimes I_K) \left[I_K + O_p\left(\frac{1}{n^{\alpha-\beta}}\right) \right] \text{ as } n \rightarrow \infty$$

and the lemma follows by standard unit root asymptotics and, in the $1 > \alpha > \beta$ case, from the fact that $J_n = o_p(1)$. It remains to show the lemma when $\alpha \leq \beta < 1$. When $\beta = \alpha$, $J_n = o_p(1)$ and $I_{KK_z} - R_{zn} \otimes R_n \rightarrow -(C_z \otimes I_K + I_{K_z} \otimes C)$, so

$$\frac{1}{n^{1+\beta}} \sum_{t=1}^n \text{vec} (x_{t-1} z'_{t-1}) = \int_0^\infty (e^{rC_z} \otimes e^{rC}) dr \text{vec} \Omega_{xz} + o_p(1) = \text{vec} V_{xz}.$$

When $\beta > \alpha$, $J_n = o_p(1)$ and $n^\alpha (I_{KK_z} - R_{zn} \otimes R_n) \rightarrow -I_{K_z} \otimes C$ so

$$\begin{aligned} \frac{1}{n^{1+\alpha}} \sum_{t=1}^n \text{vec} (x_{t-1} z'_{t-1}) &= -(I_{K_z} \otimes C^{-1}) \text{vec} \Omega_{xz} + o_p(1) \\ &= \text{vec} (C^{-1} \Omega_{xz}). \end{aligned}$$

7. References

- Abadir, K.M. and J. R. Magnus (2005). *Matrix Algebra*. Econometric Exercises, vol.1. Cambridge University Press.
- Chan, N. H. and C. Z. Wei (1988). "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes", *Annals of Statistics* 16, 367-401.
- Elliott, G. (1998). "On the Robustness of Cointegration Methods When Regressors Almost Have Unit Roots." *Econometrica* 66, 149-158.
- Elliott G., J. Rothenberg and J. H. Stock (1996). Efficient tests for an autoregressive unit root, *Econometrica* 64, 813-836.
- Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and its Application*. Academic Press.
- Kurtz, T.G. and P. Protter (1991). "Weak limit theorems for stochastic integrals and stochastic differential equations". *Annals of Probability* 19, 1035-1070.
- Jacod, J. and A.N. Shiryaev (1987). *Limit Theorems for Stochastic Processes*. Springer-Verlag.
- Johansen, S. (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551-80.
- Magdalinos, T. and P. C. B Phillips (2009). "Limit Theory for Cointegrated Systems with Moderately Integrated and Moderately Explosive Regressors", *Econometric Theory*, forthcoming.

- Newey, W.K. and K.D. West (1987). "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix". *Econometrica*, 55, 703–708.
- Park, J. Y. and P.C.B. Phillips (1989). "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory* 4, 468-497.
- Phillips, P. C. B. (1987). "Towards a unified asymptotic theory for autoregression," *Biometrika* 74, 535–547.
- Phillips, P. C. B. (1988a). "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56, 1021-1044.
- Phillips, P. C. B. (1988b). "Multiple regression with integrated processes." In N. U. Prabhu, (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics* 80, 79–106.
- Phillips, P. C. B. and S. N. Durlauf (1986). "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies* 53, 473–496.
- Phillips, P. C. B. and B. E. Hansen (1990). "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies* 57, 99–125.
- Phillips, P. C. B. and T. Magdalinos (2007), "Limit theory for Moderate deviations from a unit root." *Journal of Econometrics*, 136, 115-130.
- Phillips, P. C. B. and V. Solo (1992), "Asymptotics for Linear Processes," *Annals of Statistics* 20, 971–1001.
- Saikkonen, P. (1991). "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory* 7, 1-21.
- Stock, J. H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegration Vectors", *Econometrica* 55, 1035 - 1056.
- Stock, J.H. and M.W. Watson (1993). "A simple estimator of cointegrating vectors in higher order integrated systems", *Econometrica* 61, 783-821.