*Econometric Theory*, **25**, 2009, 1682–1715. Printed in the United States of America.
doi:10.1017/S0266466609990296

# UNIT ROOT AND COINTEGRATING LIMIT THEORY WHEN INITIALIZATION IS IN THE INFINITE PAST

PETER C.B. PHILLIPS
*Yale University*
*University of Auckland*
*University of York*
*and*
*Singapore Management University*

TASSOS MAGDALINOS
*University of Nottingham*

It is well known that unit root limit distributions are sensitive to initial conditions in the distant past. If the distant past initialization is extended to the infinite past, the initial condition dominates the limit theory, producing a faster rate of convergence, a limiting Cauchy distribution for the least squares coefficient, and a limit normal distribution for the $t$-ratio. This amounts to the tail of the unit root process wagging the dog of the unit root limit theory. These simple results apply in the case of a univariate autoregression with no intercept. The limit theory for vector unit root regression and cointegrating regression is affected but is no longer dominated by infinite past initializations. The latter contribute to the limiting distribution of the least squares estimator and produce a singularity in the limit theory, but do not change the principal rate of convergence. Usual cointegrating regression theory and inference continue to hold in spite of the degeneracy in the limit theory and are therefore robust to initial conditions that extend to the infinite past.

## 1. INTRODUCTION

It is well known that limit distributions in models with unit roots and cointegration are sensitive to specific features of the model, such as the presence of intercepts, deterministic trends, breaks in trend, and persistent shifts in variance. This sensitivity in the asymptotics has led to much research and is a distinguishing feature of limit theory for nonstationary time series regression. Paul Newbold contributed to this literature in several ways, cautioning practitioners about the use of the

Our thanks to Robert Taylor (the co-editor) and two referees for helpful comments on the original version of the paper. Phillips acknowledges partial support from a Kelly Fellowship and from the NSF under grant nos. SES 04-142254 and SES 06-47086. Address correspondence to Peter C.B. Phillips, Department of Economics, Yale University, P.O. Box 208268, New Haven, CT 06520-8268; e-mail: peter.phillips@yale.edu.

  0266-4666/09 $15.00

correct limit theory in testing (see, for instance, Kim, Leybourne, and Newbold, 2004) and suggesting modifications where these might be useful in practical work (e.g., Leybourne, Kim, and Newbold, 2005). The present paper has a similar aim, with a focus not so much on model specification but on the initial conditions and their role in impacting the limit theory.

Early research on unit root limit theory revealed that initial conditions could play an important role in the finite sample performance of tests and the form of the limit distribution. The latter role was evident in continuous record asymptotics (Phillips, 1987) and unit root asymptotics developed under distant past initializations (Phillips and Lee, 1996; Uhlig, 1995). The importance of initial conditions in affecting size and power in inference has been particularly emphasized in recent work (Elliott, 1999; Müller and Elliott, 2003; Elliott and Müller, 2006; Harvey, Leybourne, and Taylor, 2009).

For many economic time series that wander randomly like integrated processes, the precise initialization of the sample observations that are used in inference typically has nothing to do with and, in principle at least, should not affect the underlying stochastic properties of the time series. Moreover, the stochastic properties of the initiating observation are often expected to be analogous to those of the terminal observation. Accordingly, just as the time series may wander according to a stochastic trend, the initialization itself may be regarded as the outcome of a similar random wandering process that may have originated in the distant past. In developing asymptotics that embody these properties, it is therefore of some interest to determine the effects of such conditions on the form of the limit theory and on econometric inference.

The present contribution points out that if a distant past initialization is extended to the infinite past, as is frequently the case in stationary series, then the unit root limit theory is dominated by the initial condition. This outcome is equivalent to the tail of the unit root process wagging the dog of the unit root limit theory, an analogy used in an early draft of this paper (Phillips, 2006). Thus, even though an invariance principle still operates, the tail of the process determines the form of the limit theory. In such cases, initial conditions are evidently of great significance.

To fix ideas, consider the simple unit root autoregression

$$x_t = \rho x_{t-1} + u_t, \qquad t \in \{1, \dots, n\}, \qquad \rho = 1, \tag{1}$$

driven by stationary innovations $u_t$ and where $\{x_0, x_1, \dots, x_n\}$ are observed. In order for the process $x_t$ to be uniquely defined by the stochastic difference equation (1), an initial condition is required. In most cases this initial condition is taken to be a constant or a random variable with a specified distribution; see, e.g., White (1958) and Anderson (1959). However, other possibilities may be considered. Much of the theory for the stationary case ($|\rho| < 1$) is based on the Wold decomposition $x_t = \sum_{j=0}^{\infty} \rho^j u_{t-j}$, which entails an initial condition of the form $x_0 = \sum_{j=0}^{\infty} \rho^j u_{-j}$ for (1), so that $x_0$ and $x_t$ are comparable in distribution and order of magnitude. When $\rho = 1$, the infinite series in this initialization for $x_0$ almost surely diverges. We can nonetheless consider an initial condition of the form

$$x_0(n) = \sum_{j=0}^{\kappa_n} u_{-j}, \tag{2}$$

where $\kappa_n$ is an integer-valued sequence increasing to infinity with the sample size. Clearly, the sequence $\kappa_n$ determines how many past innovations are included in the initial condition, with larger values of $\kappa_n$ associated with the more distant past. As shown below, under suitable assumptions on the innovation sequence, $\kappa_n^{-1/2} x_0(n)$ has a limiting form that dominates the rate of convergence and the asymptotic distribution of $\hat{\rho}_n$ when $\kappa_n/n \to \infty$. The limit distribution of $\hat{\rho}_n$ is then Cauchy and bears more similarity to autoregressions with explosive or mildly explosive roots (cf. Phillips and Magdalinos, 2007, 2008) than it does to conventional unit root limit theory. Andrews and Guggenberger (2008) found that a similar result applies for autoregressions with roots very close to unity and infinite past initializations.

On the other hand, when (1) is a vector autoregression, an infinite past initialization for the process gives rise to a singularity in the asymptotic form of the sample moment matrix. This degeneracy is analyzed in this paper by characterizing the degeneracy and rotating the regression coordinates in a direction orthogonal to the initial condition. These reductions produce a limit theory for the least squares estimator that has the usual $n$-rate of convergence but a different analytic form.

In cointegrated models involving integrated processes where initial conditions are in the infinite past, a similar degeneracy occurs in the limiting sample moments. Nonetheless, the usual mixed normal limit theory for estimation of the cointegrating matrix still applies, and inference may proceed as usual in such situations. The effect of infinite past initializations is therefore moderated in multiple regressions when there are some unit roots. These results are relevant in practice and confirm that there is some robustness in cointegrating regression theory to very distant initializations. In this respect, scalar unit root limit theory and cointegration theory are again quite distinct.

The paper is organized as follows: Section 2 outlines the models used and formulates initial conditions into three categories (recent, distant, and infinitely distant) determined by the inherent order of magnitude of the initialization and the extent to which the initialization reaches into the past. Our primary interest in this paper is in the third category, where infinite past initializations are permitted. Some preliminary results for unit root autoregressions and vector autoregressions as well as the new limit theory for infinitely distant past realizations are presented in Section 2. Section 3 develops the corresponding limit theory for cointegrated systems and explores the implications for inference. Section 4 discusses extensions to models with deterministic trend. Proofs are given in the Appendix.

Throughout the paper, standard weak convergence and unit root limit theory notation are employed.

## 2. LIMIT THEORY UNDER EXTENDED INITIALIZATIONS

### 2.1. Model and Assumptions

Consider an $\mathbb{R}^K$-valued integrated process generated by

$$x_t = Rx_{t-1} + u_t, \qquad t \in \{1,\ldots,n\}, \qquad R = I_K, \tag{3}$$

where $u_t$ is a sequence of zero mean, weakly dependent disturbances and $x_0 = x_0(n)$ is an initialization based on past innovations that is possibly dependent on the sample size $n$. The latter dependence enables $x_0$ to have analogous properties to those on the sample trajectory $x_t$ for $t = [nr]$, where $r$ is some fraction of the sample and $[a]$ is the integer part of $a$. The following conditions facilitate the development of a limit theory based on the Phillips-Solo (1992) framework.

**Assumption LP.** Let $F(z) = \sum_{j=0}^{\infty} F_j z^j$, where $F_0 = I_K$ and $F(1)$ has full rank. For each $s \in \mathbb{Z}$, $u_s$ has Wold representation

$$u_s = F(L)\varepsilon_s = \sum_{j=0}^{\infty} F_j \varepsilon_{s-j}, \qquad \sum_{j=0}^{\infty} j^2 \left\|F_j\right\|^2 < \infty, \tag{4}$$

where $(\varepsilon_s)_{s\in\mathbb{Z}}$ is a sequence of independent and identically distributed $(0,\Sigma)$ random vectors with $\Sigma$ positive definite.

We employ the usual notation $\Omega = F(1)\Sigma F(1)'$ for the long-run variance of $u_s$ and $\Lambda = \sum_{h=1}^{\infty} \mathrm{E}\left(u_t u_{t-h}'\right)$, $\Delta = \sum_{h=0}^{\infty} \mathrm{E}\left(u_t u_{t-h}'\right)$ for the one-sided long-run covariance matrices.

Under (3), we may decompose $x_t$ as

$$x_t = x_0(n) + Y_t, \tag{5}$$

where $Y_t := \sum_{j=1}^{t} u_j$ is an integrated process with initial condition $Y_0 = 0$. The asymptotic behavior of $x_t$ is governed by the order of magnitude of the initialization $x_0(n)$, which in turn depends on the behavior of $\kappa_n$ as $n \to \infty$.

**Assumption IC.** The initial condition $x_0(n)$ of the stochastic difference equation (3) is given by (2) with $u_{-j}$ satisfying Assumption LP and $(\kappa_n)_{n\in\mathbb{N}}$ an integer-valued sequence satisfying $\kappa_n \to \infty$ and

$$\frac{\kappa_n}{n} \to \tau \in [0,\infty] \quad \text{as } n \to \infty. \tag{6}$$

The following cases are distinguished:

(i) If $\tau = 0$, $x_0(n)$ is said to be a recent past initialization.

(ii) If $\tau \in (0,\infty)$ , $x_0(n)$ is said to be a distant past initialization.

(iii) If $\tau = \infty$, $x_0(n)$ is said to be an infinite past (or infinitely distant) initialization.

The above rates for the sequence $\kappa_n$ are considered in view of the differing impact of the initial condition on the time series $x_t$ and least squares regression theory on (3). Recent past initializations, where $\tau = 0$, satisfy $x_0(n) = O_p\left(\kappa_n^{1/2}\right) = o_p\left(n^{1/2}\right)$ and do not contribute to the limiting distribution of the least squares coefficient estimator $\hat{R}_n = \sum_{t=1}^n x_t x_{t-1}' \left(\sum_{t=1}^n x_{t-1} x_{t-1}'\right)^{-1}$, in the same way that constant initial conditions are asymptotically negligible. Thus, the limit distribution of the standardized and centered estimator $n(\hat{R}_n - I_K)$ is invariant to recent past initialization of the process and has the standard form given in Phillips and Durlauf (1986). Distant past initializations have asymptotic order $x_0(n) = O_p\left(n^{1/2}\right)$ and are of the same order of magnitude as the partial sum process in the functional limit theory that drives unit root asymptotics. In consequence, the standard approach to unit root limit theory applies but with an additional contribution from the initial condition, as shown in Phillips and Lee (1996) for the near-integrated case.

The main component that determines whether the initialization of $x_t$ contributes to least squares regression limit theory is the joint asymptotic behavior of the initialization $x_0(n)$ and the partial sum process of $(u_t)_{1 \le t \le n}$

$$\left[\kappa_n^{-1/2} x_0(n), n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} u_t\right]' \Rightarrow \Omega[W_0(1), W(r)]', \tag{7}$$

where $W_0$ and $W$ are independent standard Brownian motions on [0, 1]. The condition given in (7) determines the asymptotic behavior of the least squares estimator, which varies according to the three cases presented in Assumption IC.

The effect of infinite past initializations on unit root limit theory is materially different and seems not to have been considered in the published literature, although some results may be familiar.[1] The present paper makes several contributions to this subject. First, we show that an infinite past initialization dominates the unit root limit theory, giving rise to a Cauchy limit distribution for the normalized and centered least squares estimator and a limit normal distribution for the $t$-statistic in the univariate case. These results, which are analogous to those for an explosive Gaussian autoregression, hold under an invariance principle. Second, for multivariate integrated regressors, the effects are shown to be more complex in nature but simpler in terms of their implications. The complexity arises because infinite past initializations produce an asymptotic degeneracy that gives rise to a singular least squares regression limit theory. This singularity carries over to cointegrating regression limit theory, where the effects are important for inference because they ensure robustness of the standard limit theory to infinite past initial conditions, thereby simplifying the effects of initialization on inference. In this respect, there are some major differences between the effects of "large" initial conditions on unit root limits and cointegration regression theory.

## 2.2. Recent and Distant Past Initializations

The following result summarizes limit theory for $\hat{R}_n$ covering recent and distant past initializations and is largely already familiar.

THEOREM 1. *Under model (3) and Assumptions LP and IC, with* $\tau \in [0, \infty)$,

$$n\left(\hat{R}_n - I_K\right) \Rightarrow \left(\int_0^1 dB B_\tau^{+\prime} + \Lambda\right)\left(\int_0^1 B_\tau^+ B_\tau^{+\prime}\right)^{-1}, \quad \text{as } n \to \infty, \tag{8}$$

*where $B$ and $B_0$ are independent $K$-vector Brownian motions with variance matrix $\Omega$ and $B_\tau^+(s) = B(s) + \sqrt{\tau} B_0(1)$.*

**Remark A.**

(i) Under recent past initializations, $\tau = 0$ and the usual least squares regression theory (Phillips and Durlauf, 1986; Phillips, 1988a) apply. Similar results have been obtained (see Müller and Elliott, 2003, and the references therein) for nearly integrated processes with coefficient matrix $R_n = I_K + C/n$, $C = \text{diag}(c_i) < 0$, and an initial condition of the form $x_0(n) = \sum_{j=0}^{\infty} R_n^j u_{-j}$. Of course, when $C = 0$ this infinite series diverges. The integrated processes of the present paper could be nested into a local to unity framework by choosing a more flexible distant past initialization of the form

$$x_0(n) = \sum_{j=0}^{\kappa_n} R_n^j u_{-j}, \tag{9}$$

with $R_n = I_K + C/n$ and $\kappa_n/n \to \tau \in (0, \infty)$, as in Phillips and Lee (1996). Theorem 1 then specializes that limit theory to the case where $C = 0$. In this sense, Theorem 1 is not new and is included for the sake of completeness.

(ii) The Brownian motions $B_0$ and $B$ in Theorem 1 are independent limit processes corresponding to partial sums that involve past and sample period innovations, respectively. These processes are defined by the functional laws

$$\xi_n^0(s) := \frac{F(1)}{\kappa_n^{1/2}} \sum_{j=0}^{\lfloor \kappa_n s \rfloor} \varepsilon_{-j} \Rightarrow B_0(s) \quad \text{and} \quad \xi_n(s) := \frac{F(1)}{n^{1/2}} \sum_{t=1}^{\lfloor ns \rfloor} \varepsilon_t \Rightarrow B(s)$$

given in (A.3) and (A.13) in the Appendix. The composite process $B_\tau^+(s)$ in Theorem 1 then depends on both the limiting sample trajectory $B(s)$ and the component $\sqrt{\tau} B_0(1)$, which carries the effect of the initial conditions.

(iii) Theorem 1 is readily extended to include the case where a nonparametric bias correction (Phillips, 1987) is made to the estimate $\hat{R}_n$ involving a consistent estimator $\hat{\Lambda}$ of the one-sided long-run covariance matrix $\Lambda$ that is constructed in the usual manner from regression residuals. Expression (8) in the limit theory is adjusted accordingly, eliminating the term in the numerator of the matrix quotient that involves $\Lambda$. Evidently, the critical values corresponding to this limit theory differ from those delivered by standard unit root tabulations when $\tau > 0$, partly explaining the size distortions from distant initializations that can occur in such cases.

(iv) If an intercept is included in the regression, i.e., the integrated process $x_t$ is generated by (3) but the least squares estimator is obtained from the regression

$$x_t = \hat{\mu}_n + \hat{R}_n^{\mu} x_{t-1} + \hat{u}_t, \tag{10}$$

then the distribution of $\hat{R}_n^{\mu}$ is invariant to the initial condition $x_0$ even in finite samples. This simple algebraic fact implies, in particular, that the limit theory for least squares regression in this case is given by Theorem 1, with $\tau = 0$ and $B$ replaced by demeaned Brownian motion.

### 2.3. Infinite Past Initializations: Scalar Autoregression

The main contribution of the present paper is the development of a limit theory under infinitely distant initializations, as presented in Theorems 2 and 3 below and in the cointegration limit theory that follows in Sections 3 and 4. We start with the scalar case.

THEOREM 2. *When $K = 1$ and Assumptions LP and IC hold, with $\tau = \infty$, the following limit theory applies as $n \to \infty$:*

*(i) $\sqrt{\kappa_n n}\left(\hat{R}_n - 1\right) \Rightarrow \mathcal{C}$, where $\mathcal{C}$ is a standard Cauchy variate.*

*(ii) Letting $\hat{s}_n^2 = n^{-1}\sum_{t=1}^{n}\left(x_t - \hat{R}_n x_{t-1}\right)^2$, the t-statistic satisfies*

$$\frac{\left(\sum_{t=1}^{n} x_{t-1}^2\right)^{1/2}}{\hat{s}_n}\left(\hat{R}_n - 1\right) \Rightarrow \sqrt{\frac{\Omega}{\sigma^2}}\,W(1), \tag{11}$$

*where $\sigma^2 = \mathrm{E}\left(u_t^2\right)$ and $W$ is standard Brownian motion.*

**Remark B.**

(i) Part (ii) follows immediately from part (i) and the fact that $\hat{s}_n^2 = n^{-1}\sum_{t=1}^{n} u_t^2 + O_p\left(n^{-1}\right) \to_p \mathrm{E}\left(u_t^2\right)$. Theorem 2 shows that integrated processes with infinitely distant initializations do not conform with the usual unit root asymptotics. The asymptotic behavior of the least squares estimator presents more similarities to explosive rather than unit root regression theory in the form of the limiting distribution, its symmetry, and the rate of convergence. The latter can be made to grow arbitrarily fast according to how far in the past of the innovation sequence the initial condition $x_0(n)$ is allowed to reach. If the sequence $\kappa_n$ increases at an exponential rate, the least squares estimator of Theorem 2 may achieve or even exceed the explosive consistency rate. The limit behavior of the $t$-statistic also resembles the standard stationary and Gaussian explosive cases. When the innovation sequence $u_t$ is independent, $\Omega = \mathrm{E}\left(u_1^2\right)$, so the $t$-statistic has a standard normal limit distribution. Obviously, both (i) and (ii) can be used for inference when $u_t$ are weakly dependent innovations, and in the case of the $t$-statistic consistent estimation of $\Omega$ and $\sigma^2$ can be accomplished by standard methods.

(ii) Andrews and Guggenberger (2008) derived a result related to Theorem 2 by considering autoregressions with a root very close to unity and infinite past initialization based on i.i.d. innovations. The present result extends that theory to the unit root case with weakly dependent innovations. It is possible to nest the results of the two papers by considering the following (possibly vector valued) autoregressive process:

$$x_t = R_n x_{t-1} + u_t, \qquad R_n = I_K + C/\lambda_n \tag{12}$$

with initial condition of the form (9), where $\min\{\lambda_n, \kappa_n\}/n \to \infty$. Then the case $C = 0$ reduces to the process analyzed in this paper, whereas the case $C \neq 0$ and $K = 1$ gives rise to a process with identical asymptotic properties to that considered in Andrews and Guggenberger (2008) and a Cauchy limit theory for $\sqrt{\lambda_n n}\left(\hat{R}_n - 1\right)$. Note that in the latter case the rate of convergence of the least squares estimator is determined by the proximity of $R_n$ to a unit root rather than the number of past innovations included in the initialization (one may set $\kappa_n = \infty$). For this reason, Theorem 2 cannot be extended to include local to unity processes: When $\lambda_n = n$ and $C < 0$, we have $\sum_{j=0}^{\infty} (I_K + C/n)^j u_{-j} = O_p\left(n^{1/2}\right)$; i.e., the infinite past initialization has the same order of magnitude as the partial sum process of the innovations $(u_t)_{1 \le t \le n}$, and hence it contributes but does not dominate least squares regression theory. In this sense, the situation of Theorem 2 where an infinitely distant initial condition dominates the asymptotic behavior of the ordinary least squares (OLS) estimator is a unit root rather than a local to unity phenomenon.

(iii) It is worth pointing out that the effect of the dominating initial condition in Theorem 2 is analogous to the effect of the initial condition and initial shocks in an explosive autoregression. In that case, the initial condition and shocks also play a dominant role in determining the form of the signal, which behaves like the square of a one-dimensional random variable whose distribution depends on the distribution of the shocks in the pure explosive case but not in the mildly explosive case (see Phillips and Magdalinos, 2007). In the present case, the centered least squares estimator again behaves like the ratio of two independent random variables, one determined by the past (through $B_0$) and one by the future (through $B$). Unlike the explosive case, the limit theory involves an invariance principle, because the dominating initial condition effect arises from the functional limit law (A.4).

(iv) The heuristic explanation for the result in Theorem 2(i) is that when $\tau = \infty$ the behavior of the time series $x_{t=\lfloor nr \rfloor}$ is overtaken by the one-dimensional normal random variable $B_0$, which is not dependent on $r$, the limiting point in the sample trajectory corresponding to $t$. So the limiting trajectory of the process over the sample period is dominated by the infinitely distant initialization. Hence, upon suitable scaling, the numerator of the centered least squares estimate is a product of independent normals and the denominator is the square of one of these normals, thereby producing a Cauchy limit distribution for the centered coefficient. Upon random normalization in the case of the $t$-ratio, the effect of the infinitely distant initialization cancels from the numerator and denominator, producing a Gaussian limit. In both cases, the tail of the unit root process wags the trajectory of the process and in doing so defines the limit theory when $\tau = \infty$. Notably, in this event, the length of the unobserved tail of the process grows faster than the observed trajectory as $n \to \infty$.

(v) Theorem 2 holds for regressions through the origin. If an intercept is included in the regression as in (10), the effect of the initial condition is eliminated and standard unit root limit theory involving demeaned Brownian motion applies. However, including an irrelevant intercept leads to (infinite) efficiency loss as the consistency rate of the least squares estimator reduces from $\sqrt{\kappa_n n}$ to $n$.

(vi) The Cauchy limiting distribution of Theorem 2 requires that the initial condition is a random process. If the initial condition is a nonstochastic sequence increasing at a rate faster than $O\left(n^{1/2}\right)$, i.e., $x_0(n)/\sqrt{\kappa_n} \to \theta \in \mathbb{R} \setminus \{0\}$ with $\kappa_n/n \to \infty$ as $n \to \infty$, the tail of the unit root process again wags the trajectory of the process. The decomposition (5) and the central limit theorem then imply that $x_t$ is dominated by $x_0(n)$ in such a way that the signal is nonrandom in the limit and

$$\hat{R}_n - 1 = \frac{\sum_{t=1}^n x_{t-1} u_t}{\sum_{t=1}^n x_{t-1}^2} \sim \frac{x_0(n) \sum_{t=1}^n u_t}{n x_0(n)^2} = \frac{1}{\sqrt{n\kappa_n}} \frac{\frac{1}{\sqrt{n}} \sum_{t=1}^n u_t}{x_0(n)/\sqrt{\kappa_n}},$$

implying that $\sqrt{\kappa_n n}\left(\hat{R}_n - 1\right) \Rightarrow N\left(0, \frac{\Omega}{\theta^2}\right)$. This Gaussian limit theory corresponds to results originally obtained for large initializations in Phillips (1987) and later in Perron (1991). Of course, this specification for $x_0(n)$ is rather unrealistic because the initialization does not carry any information about past innovations, unlike initializations such as those given in (2), which carry long-range memory effects of the past innovation sequence.

(vii) Since the rate of convergence of $\hat{R}_n$ in Theorem 2 (i) is of order $\sqrt{n\kappa_n}$, which exceeds the order $n$ rate of conventional unit root theory (Phillips, 1987), it is apparent that conventional coefficient-based unit root tests will produce conservative tests, thereby underrejecting the null hypothesis of a unit root asymptotically. Hence, as indicated in Andrews and Guggenberger (2008), the usual unit root tests are not robust to infinite past initializations with $\tau = \infty$.

(viii) A practical limitation of Theorem 2 (i) is that there is typically little information about the properties of the initialization, including its extent into the past. The distant past parameter $\kappa_n$ and the convergence rate $\sqrt{n\kappa_n}$ are therefore generally unknown. An exception occurs in the case of panel data. As discussed in Moon and Phillips (2000, Sect. 4.3), if $\frac{\kappa_n}{n} \to \tau$ and the parameter $\tau$ is common across individuals in the panel, it is possible to consistently estimate $\tau$ using cross-section information. In effect, the observed variation in the first sample data point across the panel carries identifying information about $\tau$ and can be used to construct a consistent estimate. This result holds for models with unit roots and roots that are local to unity.

## 2.4. Infinite Past Initializations: Vector Autoregression

When the initialization is in the infinite past ($\tau = \infty$), the sample moment matrix is shown in (A.19) to behave asymptotically as

$$\kappa_n^{-1} n^{-1} \sum_{t=1}^{n} x_{t-1} x_{t-1}' \Rightarrow B_0(1) B_0(1)' \quad \text{as } n \to \infty,$$

where $B_0 \equiv BM(\Omega)$ obtained from the functional law (A.3), so the limit is singular if $K \geq 2$. A similar situation occurs in explosively cointegrated systems with repeated roots, i.e., systems with a (possibly mildly) explosive coefficient matrix that does not have distinct latent roots (see Phillips and Magdalinos, 2008, and Magdalinos and Phillips, 2009, for details). The asymptotic singularity of the sample moment matrix may be treated by rotating the regression coordinate system to isolate the effects of the dominant component (here the initialization $x_0(n)$). This coordinate rotation is analogous to that used in Park and Phillips (1988) and Phillips (1989) for systems with cointegrated regressors, but in the present case the rotation matrix is a random matrix in the limit, corresponding to the random limit of $x_0(n)$, a feature that causes some technical complications.

To fix ideas, define

$$H(n) = \frac{x_0(n)}{[x_0(n)' x_0(n)]^{1/2}}, \tag{13}$$

and consider a $K \times (K-1)$ random orthogonal complement $H_\perp(n)$ to $H(n)$ satisfying $H_\perp(n)' H(n) = 0$ and $H_\perp(n)' H_\perp(n) = I_{K-1}$ almost surely. Although

$H_{\perp}(n)$ is not unique, its outer product is uniquely defined by the well-known identity (e.g., 8.67 in Abadir and Magnus, 2005):

$$H_{\perp}(n)H_{\perp}(n)' + H(n)H(n)' = I_K \quad \text{a.s.} \tag{14}$$

Then $M(n) = [H(n), H_{\perp}(n)]$ is a $K \times K$ orthogonal matrix that may be used to transform $x_t$ into a vector with the property that all but one of the regressors has a zero initialization. Specifically, define

$$z_t := M(n)'x_t = \begin{bmatrix} H(n)'x_t \\ H_{\perp}(n)'x_t \end{bmatrix} =: \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix}. \tag{15}$$

Then, using (5), we can write $z_{2t} = H_{\perp}(n)'x_0(n) + H_{\perp}(n)'Y_t = H_{\perp}(n)'Y_t$, which implies that $z_{2t}$ has initial condition zero, and

$$\begin{aligned} z_{1t} &= H(n)'x_t = H(n)'(x_0(n) + Y_t) = \left[x_0(n)'x_0(n)\right]^{1/2} + H(n)'Y_t \\ &= \left[x_0(n)'x_0(n)\right]^{1/2}\left\{1 + O_p\left(\sqrt{\frac{n}{\kappa_n}}\right)\right\}, \end{aligned} \tag{16}$$

under infinite past initialization ($\tau = \infty$). Thus, for large $n$, $z_{1t}$ behaves like the quantity $\left[x_0(n)'x_0(n)\right]^{1/2}$ and is independent of $t$ as $\frac{n}{\kappa_n} \to 0$. Thus, the new coordinate system reveals that in one direction the time series behaves like an integrated process originating at the origin (i.e., $H_{\perp}(n)'Y_t$), whereas in the other direction the time series behaves like a "constant" (over $t$) intercept but one that has a random diverging value as $n \to \infty$, viz., $\left[x_0(n)'x_0(n)\right]^{1/2} \sim \kappa_n^{1/2}\left\{B_0(1)'B_0(1)\right\}^{1/2}$.

The differing behavior of these components leads to a singular regression limit theory that corresponds to a unit root limit theory of reduced dimension $(K-1)$ in one direction and an explosive limit theory in the other. The outcome is presented in the following result.

THEOREM 3. *For the multivariate integrated process generated by (3) with $K \geq 2$ under Assumptions LP and IC with $\tau = \infty$, the following limit theory applies as $n \to \infty$:*

$$\begin{aligned} n\left(\hat{R}_n - I_K\right) &\Rightarrow \Psi(B, B_0) \\ &:= \left[\int_0^1 dB\underline{B}' + \Lambda\right] H_{\perp}\left(H_{\perp}'\int_0^1 \underline{B}\,\underline{B}' H_{\perp}\right)^{-1} H_{\perp}', \end{aligned} \tag{17}$$

$$\begin{aligned} &\sqrt{n\kappa_n}\left(\hat{R}_n - I_K\right)H(n) \\ &\quad \Rightarrow \left[B_0(1)'B_0(1)\right]^{-1/2}\left[B(1) - \Psi(B, B_0)\int_0^1 B\right], \end{aligned} \tag{18}$$

*where $H_\perp$ is a $K \times (K-1)$ random orthogonal complement to $B_0(1)$ satisfying (A.2), $B$ and $B_0$ are independent $K$-vector Brownian motions with variance matrix $\Omega$, and $\underline{B}(s) = B(s) - \int_0^1 B(s)\,ds$.*

**Remark C.**

(i) Theorem 3 reveals that the least squares estimator has the usual $n$-rate of convergence and that the initialization contributes to the asymptotic distribution (through $H_\perp H_\perp'$) but does not dominate the limit theory. Thus, the effect of an infinite past initial condition on multivariate unit root regression theory is moderated by higher dimensional effects in comparison with the univariate case. The result of Theorem 3 bears some similarity to regression theory under distant past initializations, where both the initial condition and the sample moments of the integrated process contribute to the limiting distribution of the least squares estimator without one dominating the other. Of course, in the direction $H(n)$, where the initialization dominates, the limit theory is accelerated to the rate $\sqrt{n\kappa_n}$. When $K = 1$, (18) reduces to the result for the scalar case given in Theorem 2(i) because in this case $H_\perp(n) = H_\perp = 0$, $H(n) = \operatorname{sign}(x_0(n)) \Rightarrow \operatorname{sign}(B_0(1))$, $\left\{B_0(1)' B_0(1)\right\}^{1/2} = |B_0(1)|$, and then (18) is simply $\sqrt{\kappa_n n}\left(\hat{R}_n - 1\right) \Rightarrow \mathcal{C}$.

(ii) Interestingly, the unit root limit theory given in (17) and (18) involves the demeaned process $\underline{B}(s)$ even though there is no intercept in the regression. The demeaning effect arises because, as shown in (16), in the direction of the initial condition, the time series is dominated by a component that behaves like a "constant," i.e., $z_{1t} \sim \left[x_0(n)' x_0(n)\right]^{1/2}$. Thus, using the identity (14) and the definition of $H(n)$ in (13), we can write the fitted regression as

$$
\begin{aligned}
x_t &= \hat{R}_n x_{t-1} + \hat{u}_t \\
&= \hat{R}_n H(n) z_{1t-1} + \hat{R}_n H_\perp(n) z_{2t-1} + \hat{u}_t \\
&\sim \hat{R}_n x_0(n) + \hat{R}_n H_\perp(n) z_{2t-1} + \hat{u}_t
\end{aligned}
$$

as $n \to \infty$. Thus, the fitted regression in the direction of $H_\perp(n)$ is given by

$$
z_{2t} \sim H_\perp(n)' \hat{R}_n x_0(n) + H_\perp(n)' \hat{R}_n H_\perp(n) z_{2t-1} + H_\perp(n)' \hat{u}_t. \tag{19}
$$

It is the regression in (19) that gives rise to the limit theory in (17), the term $H_\perp(n)' \hat{R}_n x_0(n)$ producing the demeaning effect of an intercept. Of course, this random intercept does not appear in the data generating process, since $H_\perp(n)' R x_0(n) = H_\perp(n)' x_0(n) = 0$.

(iii) The limiting distribution in Theorem 3 is singular, since the matrix

$$H_\perp \left( H_\perp' \int_0^1 \underline{B}\underline{B}' H_\perp \right)^{-1} H_\perp'$$

has rank equal to $K-1$. This is a manifestation in the limit theory of the asymptotic singularity of the sample moment matrix in the original regression coordinates.

(iv) The matrix $H_\perp \left( H_\perp' \int_0^1 \underline{B}\underline{B}' H_\perp \right)^{-1} H_\perp'$ is invariant to the coordinate system defining $H_\perp$. Thus, the limit theory of Theorem 3 is also invariant to the choice of coordinates.

(v) When $\Lambda$ is estimated nonparametrically and a corresponding bias corrected estimate $\hat{R}_n^+$ constructed, then the limit theory for this estimate is given by expression (17) with $\Lambda = 0$. This limit theory is analogous to that of a first-order vector autoregression with a fitted intercept and $K-1$ unit roots. The reason for the fitted intercept in this correspondence is that the implicit regression on $z_{1t}$ in the new coordinate system is equivalent to regression on a constant because $z_{1t} = H(n)'x_t$ behaves like $x_0(n)$ asymptotically.

(vi) As in Remark B(ii), Theorem 3 applies for autoregressive processes of the form (12) with each diagonal element of the autoregressive matrix $R_n$ lying in a small neighborhood of unity with radius defined by the condition $\lambda_n/n \to \infty$. In the local to unity case, ((12) with $\lambda_n = n$ and $C < 0$), $x_0(n)$ has order of magnitude $O_p\left(n^{1/2}\right)$ (the same as distant-past initializations), and the sample moment matrix is no longer asymptotically singular: $n^{-2}\sum_{t=1}^n x_{t-1}x_{t-1}' \Rightarrow \int_0^1 J_C^*(r)J_C^*(r)'dr$ with $J_C^*(r) = J_C(r) + J_C(1)$ where $J_C(r)$ is a vector Ornstein-Uhlenbeck process with covariance matrix $\Omega$.

## 3. COINTEGRATION UNDER EXTENDED INITIALIZATION

This section considers the cointegrated system

$$y_t = Ax_t + u_{yt}, \qquad x_t = x_{t-1} + u_{xt}, \qquad t \in \{1, \ldots, n\}, \tag{20}$$

where $u_t = \left(u_{yt}', u_{xt}'\right)'$ is an $m+K$-vector of innovations satisfying Assumption LP, $A$ is an $m \times K$ matrix of cointegrating coefficients, $x_t$ is a $K$-vector of integrated time series, and the system is initialized at some $x_0(n) = \sum_{j=0}^{\kappa_n} u_{x,-j}$ that satisfies Assumption IC. Under LP, the functional law $n^{-1/2}\sum_{j=1}^{\lfloor n\cdot \rfloor} u_j \Rightarrow B(\cdot)$ applies, with $B$ an $m+K$-vector Brownian motion with variance matrix $\Omega$. We partition the limiting Brownian motion and the various matrices associated with its variance conformably with $u_t$ as follows: $B = \left(B_y', B_x'\right)'$,

$F(1) = \left(F_y(1)', F_x(1)'\right)'$,

$$\Omega = \begin{bmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{bmatrix}, \quad \text{and} \quad \Delta = \begin{bmatrix} \Delta_{yy} & \Delta_{yx} \\ \Delta_{xy} & \Delta_{xx} \end{bmatrix}.$$

Finally, we let $B_0$ denote a $K$-vector Brownian motion with variance matrix $\Omega_{xx}$ defined by the functional law $\kappa_n^{-1/2} \sum_{j=0}^{\lfloor \kappa_n \cdot \rfloor} u_{x,-j} \Rightarrow B_0(\cdot)$.

We will be concerned with the effect of the initialization on the limit theory of cointegration estimators and tests. These effects are demonstrated in terms of the fully modified (FM) regression procedure (Phillips and Hansen, 1990), and the same results apply for other commonly used cointegration procedures. Of course, under IC(i) or recent past initializations, the limit theory is well known to be invariant to the effects of $x_0(n)$. Under IC(ii), the effects are manifest in the mixture process in the limit theory, so that

$$\operatorname{vec}\left\{n\left(\hat{A}^+ - A\right)\right\} \Rightarrow MN\left(0, \left(\int_0^1 B_\tau^+ B_\tau^{+\prime}\right)^{-1} \otimes \Omega_{yy.x}\right), \tag{21}$$

where $\hat{A}^+$ is the FM regression estimator, $\Omega_{yy.x} = \Omega_{yy} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy}$ is the conditional long-run covariance matrix of $u_{yt}$ given $u_{xt}$, and $B_\tau^+(s) = B_x(s) + \sqrt{\tau} B_0(1)$ as in Theorem 2, so that $B_x$ and $B_0$ are independent $K$-vector Brownian motions with variance matrix $\Omega_{xx}$. Result (21) follows in a straightforward way using results obtained in the proof of Theorem 2. Since $\int_0^1 B_\tau^+ B_\tau^{+\prime}$ is the weak limit of the standardized sample moment matrix $n^{-2}\sum_{t=1}^n x_t x_t'$, as shown earlier, the limit theory (21) leads to the usual inferential theory based on the estimate $\hat{A}^+$. Thus, the conventional approach to inference in cointegrated systems is robust to both recent and distant initializations. We therefore focus our attention in this section on infinitely distant initial conditions.

The FM regression estimator has the explicit form $\hat{A}^+ = \left(\hat{Y}^{+\prime}X - n\hat{\Delta}_{yx}^+\right) \times \left(X'X\right)^{-1}$, where $X = \left[x_1', \ldots, x_n'\right]'$, $\hat{Y}^+ = \left[\hat{y}_1^{+\prime}, \ldots, \hat{y}_n^{+\prime}\right]'$ is an $n \times m$ matrix of observations of corrected variates $\hat{y}_t^+ = y_t - \hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1}\Delta x_t$, where $\hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1}$ and $\hat{\Delta}_{yx}^+$ are consistent estimates of $\Omega_{yx}\Omega_{xx}^{-1}$ and $\Delta_{yx}^+ = \Delta_{yx} - \Omega_{yx}\Omega_{xx}^{-1}\Delta_{xx}$, all of which may be constructed in the familiar fashion using semiparametric lag kernel methods with residuals from a preliminary cointegrating least squares regression on (20). The limit theory for $\hat{A}^+$ under infinitely distant initial conditions as given in IC(iii) is as follows:

THEOREM 4. *Under model (20) and Assumptions LP and IC(iii) with $\tau = \infty$, we have, as $n \to \infty$,*

$$\operatorname{vec}\left\{n\left(\hat{A}^+ - A\right)\right\} \Rightarrow MN\left(0, H_\perp\left(H_\perp' \int_0^1 \underline{B}_x \underline{B}_x' H_\perp\right)^{-1} H_\perp' \otimes \Omega_{yy.x}\right), \tag{22}$$

$$\sqrt{n\kappa_n}\left(\hat{A}^+ - A\right)H(n) \Rightarrow MN\left(0, \Omega_{yy.x}\frac{\int_0^1 V(s)^2 ds}{B_0(1)'B_0(1)}\right), \tag{23}$$

*where $B$ and $B_0$ are independent $K$-vector Brownian motions with variance matrix $\Omega_{xx}$, $H_\perp$ is a $K \times (K-1)$ random orthogonal complement to $B_0(1)$ satisfying (A.2), $\underline{B}_x(s) = B_x(s) - \int_0^1 B_x(s)ds$ is demeaned $B_x$, $B_{y.x} = B_y - \Omega_{yx}\Omega_{xx}^{-1}B_x$ is Brownian motion with covariance matrix $\Omega_{yy.x}$ independent of $B_x$ and $B_0$, and*

$$V(s) = 1 - \underline{B}_x(s)' H_\perp \left( H_\perp' \int_0^1 \underline{B}_x \underline{B}_x' H_\perp \right)^{-1} H_\perp' \left( \int_0^1 B_x \right).$$

**Remark D.**

(i) The limit distribution of $\hat{A}^+$ is mixed Gaussian, just as in the case of recent and distant initial conditions, and the dominating rate of convergence is order $n$ as usual. The dominating limit theory (22) is invariant to the infinitely distant initialization. Nonetheless, the initialization does affect the limit theory because the limit distribution (22) is singular and a faster convergence rate $\sqrt{n\kappa_n}$ applies in the direction of the infinitely distant initial condition. In that direction, the limit theory is also mixed Gaussian, and the mixing variate depends on the squared norm $\|B_0(1)\|^2 = B_0(1)'B_0(1)$ of the standardized limiting initialization. Thus, although the initialization does have an effect on the limit theory, it is of secondary importance. It should be noted, however, that this effect is present in the construction of test statistics. Mixed normality in Theorem 4 can be used in order to construct asymptotically chi-squared Wald test statistics for the linear restrictions $H_0 : M\text{vec}(A) = m$, where $M$ is a known $r \times mK$ matrix of rank $r$ and $m$ is a known $mK$-vector. Then, in the notation of (15) and Lemma A2, an asymptotically $\chi^2(r)$ Wald test statistic would take the form

$$W_n = \left[M\text{vec}\left(\hat{A}^+\right) - m\right]' \left\{M \left[H_\perp(n)\left(Z_2'Q_1Z_2\right)^{-1} H_\perp(n)' \otimes \hat{\Omega}_{yy.x}\right] M'\right\}^+ \times \left[M\text{vec}\left(\hat{A}^+\right) - m\right],$$

where $\hat{\Omega}_{yy.x}$ is a consistent estimator of $\Omega_{yy.x}$ and $Q^+$ denotes the Moore-Penrose inverse of a matrix $Q$. Similar considerations apply in testing non-linear restrictions.

(ii) As in Theorem 3, the limit theory (22) involves the demeaned process $\underline{B}_x(s)$ corresponding to the regressor $x_t$. Again, the demeaning is caused by the fact that in the direction of the initial condition, the time series $x_t$ is dominated by a component that behaves like a "constant" (in this case, $H(n)'x_t \sim B_0(1)'B_0(1)$) that acts like an intercept in the limit theory, see Remark C(ii). Therefore, one material impact of the infinitely distant initialization is that the regression equation behaves as if there is a fitted intercept.

## 4. EXTENSIONS TO MODELS WITH DRIFT

The above discussion has considered unit root and cointegration regression models without intercept and trend. Introducing drift to these models provides a practical extension that produces some further new results. It will be sufficient to use the cointegrating regression model to illustrate the effects of drift in both the sample observations and the initial conditions. One aspect of the results—an increase in the degeneracy of the limit theory stemming from a drifted initialization—is not immediate.

We take model (20), assume $K > 3$, and replace the generating mechanism of the regressors by

$$x_t = \beta t + x_t^*, \qquad t = 1, \ldots, n, \tag{24}$$

$$x_t^* = \sum_{j=1}^{t} u_{xj} + x_0^*, \qquad x_0^*(n) = \sum_{j=0}^{\kappa_n} u_{x,-j} + \beta\kappa_n, \tag{25}$$

in which case $x_0 = x_0^*(n)$ is the outcome of a random wandering process with drift so that its stochastic order is $O_p(\kappa_n)$, which is analogous to that of $x_t$. In this event, the sample data $X' = [x_1, \ldots, x_n]$ satisfy

$$X = \tau_n \beta' + \iota_n x_0^{*\prime} + S,$$

where $S' = [S_1, \ldots S_n]$ with $S_t = \sum_{j=1}^{t} u_{xj}$, $\tau_n = (1, \ldots, n)'$, and $\iota_n = (1, \ldots, 1)'$. As usual, unit root regression with a fitted trend and intercept removes the effects of the initialization $x_0^*$ and the trend coefficient $\beta$, and conventional theory applies with appropriate effects of the detrending being manifest in the limit theory, as shown in Park and Phillips (1988) across a variety of models. Similar considerations apply in the present case but with an additional complication arising from the form of the initialization (25).

We illustrate by taking the case of FM regression applied to (20) with $x_t$ generated as in (24). Here, the limit theory is given by

$$\operatorname{vec}\left\{n\left(\hat{A}^{+} - A\right)\right\} \Rightarrow MN\left(0, H_\perp\left(H_\perp' \int_0^1 \underline{B}_x \underline{B}_x' H_\perp\right)^{-1} H_\perp' \otimes \Omega_{yy.x}\right), \tag{26}$$

where $\underline{B}_x$ is the detrended process

$$\underline{B}_x(r) = B_x(r) - \left(\int_0^1 B_x Z'\right)\left(\int_0^1 ZZ'\right)^{-1} Z'(r), \qquad Z(r) = (1, r)', \tag{27}$$

so that the limit theory is entirely analogous in form to that given in (22). However, in the present case, the additional complication stems from the fact that the directional matrix $H_\perp$ has structure and rank that reflect the presence of the time trend and the space spanning the infinitely distant initialization. The latter is affected by the rate at which $\kappa_n \to \infty$ in relation to $n$ and the various components of the initialization, which we now briefly discuss.

Observe that under (25) the drift in the initialization determines the primary limit so that $\kappa_n^{-1} x_0^*(n) \to_p \beta$. Expanding the probability space as needed for the strong invariance principle $B_{0n} := \kappa_n^{-1/2} \sum_{j=0}^{\lfloor \kappa_n \cdot \rfloor} u_{x,-j} \to_{a.s} B_0(\cdot)$ to hold, with $B_0 \equiv \text{BM}(\Omega_{xx})$, the large sample behavior of the initial condition has the form

$$x_0^*(n) = \sum_{j=0}^{\kappa_n} u_{x,-j} + \beta \kappa_n = \beta \kappa_n + B_{0n} \sqrt{\kappa_n}$$

$$= [\beta, B_0(1)] \begin{bmatrix} \kappa_n \\ \sqrt{\kappa_n} \end{bmatrix} \{1 + o_{a.s}(1)\},$$

so that $x_0^*(n)$ is spanned by the two columns of the matrix $C_n = [\beta, B_{0n}]$ and in the limit by the matrix $C = [\beta, B_0(1)]$, where $\beta$ and $B_0(1)$ are a.s. linearly independent vectors. The components $\beta$ and $B_{0n}$ of the initialization vector $x_0^*(n)$ have divergence rates $\kappa_n$ and $\kappa_n^{1/2}$ corresponding to the two components in (25). So, because of (24) there will be a time trend in the regression, and because of the effect of the initial condition there is effectively a (random) intercept in the regression, since $\kappa_n$ is large. When $\kappa_n$ is very large relative to the trend, in particular if $\frac{\kappa_n^{1/2}}{n} \to \infty$, then $x_0^*(n)$ is the dominating force in the asymptotics and both components of $x_0^*(n)$ figure in the limit theory. To resolve the limit, we transform coordinates using the matrix $[C_n, C_{n\perp}]$, where $C_{n\perp}$ is a complementary matrix of vectors orthogonal to $C$, giving

$$\begin{bmatrix} C_n' \\ C_{n\perp}' \end{bmatrix} x_{\lfloor n \cdot \rfloor} = \begin{bmatrix} C_n' \beta \lfloor n \cdot \rfloor + C_n' C_n \begin{bmatrix} \kappa_n \\ \sqrt{\kappa_n} \end{bmatrix} + C_n' S_{\lfloor n \cdot \rfloor} \\ C_{n\perp}' \beta \lfloor n \cdot \rfloor + C_{n\perp}' S_{\lfloor n \cdot \rfloor} \end{bmatrix}.$$

If $\kappa_n$ is such that $\frac{\sqrt{\kappa_n}}{n} \to \infty$, the largest effect is in the direction $C_n$, so that both components $\beta$ and $B_{0n}$ are relevant. The next largest effect comes in the direction $C_{n\perp}' \beta$ and then finally the dominating effect on the limit theory for $\hat{A}^+$ with slowest asymptotics comes in the direction orthogonal to $\left[C_n, C_{n\perp}' \beta\right]$. That rate is $O_p(n)$, and the limit theory for $\hat{A}^+$ is just as given in Theorem 4 by (22) or (26) above. However, in this case, $H_\perp$ is of reduced dimension $K \times (K-3)$ and is a random orthogonal matrix spanning the orthogonal complement of the limit matrix $\left[C, C_\perp' \beta\right]$. The dimension reduction to $K-3$ in the columns of $H_\perp$ comes about because of the effect of the linear trend in $x_t$ and the initialization $x_0^*(n)$, which lies in the two-dimensional space spanned by $C$ in the limit. The process $\underline{B}_x$ in (26) is the detrended process (27). Again, inference proceeds as usual in the presence of initializations such as (25).

Thus, initialization with drift in a cointegrated system does not affect the practicalities of inference even when the initialization is in the infinite past. But initialization does influence the form of the asymptotic theory in a subtle manner in terms of its dimensionality and its support, whose orientation involves a random component that is determined by infinitely distant initialization effects.

## NOTE

1. For instance, the scalar case has been given in Yale time series lectures for some years and, as mentioned above, Andrews and Gugenberger (2007) recently considered a very near to unity scalar limit theory with infinite past initializations.

## REFERENCES


Abadir K.M. & J.R. Magnus (2005) *Matrix Algebra*. Econometric Exercises, vol. 1. Cambridge University Press.

Anderson, T.W. (1959) On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676–687.

Andrews, D.W.K. (1987) Asymptotic results for generalised Wald tests. *Econometric Theory* 3, 348–358.

Andrews, D.W.K. & P. Guggenberger (2008) Asymptotics for stationary very nearly unit root processes. *Journal of Time Series Analysis* 29, 203–210.

Elliott, G. (1999) Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *International Economic Review* 40, 767–783.

Elliott, G. & U.K. Müller (2006) Minimizing the impact of the initial condition on testing for unit roots. *Journal of Econometrics* 135, 285–310.

Harvey, D.I., S.J. Leybourne, & A.M.R. Taylor (2009) Unit root testing in practice: Dealing with uncertainty over the trend and initial condition. *Econometric Theory* 25, 587–636.

Kim, T.-H., S. Leybourne, & P. Newbold (2004) Behaviour of Dickey-Fuller unit root tests under trend misspecification. *Journal of Time Series Analysis* 25, 755–764.

Leybourne, S.J., T.-H. Kim, & P. Newbold (2005) Examination of some more powerful modifications of the Dickey-Fuller test. *Journal of Time Series Analysis* 26, 355–369.

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

Moon, H.R. & P.C.B. Phillips (2000) Estimation of autoregressive roots near unity using panel data. *Econometric Theory* 16, 927–998.

Müller, U. & G. Elliott (2003) Tests for unit roots and the initial condition. *Econometrica* 71, 1269–1286.

Park, J.Y. & P.C.B. Phillips (1988) Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4, 468–497.

Perron, P. (1991) A continuous time approximation to the unstable first order autoregressive process: The case without an intercept. *Econometrica* 59, 211–236.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–302.

Phillips, P.C.B. (1988a) Multiple regression with integrated processes. In N.U. Prabhu (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics* 80, 79–106.

Phillips, P.C.B. (1988b) Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. *Econometric Theory* 4, 528–533.

Phillips, P.C.B. (1988c) Weak convergence to the matrix stochastic integral $\int BdB'$. *Journal of Multivariate Analysis* 24(2), 252–264.

Phillips, P.C.B. (1989) Partially identified econometric models. *Econometric Theory* 5, 181–240.

Phillips, P.C.B. (2006) When the Tail Wags the Unit Root Limit Theory. Mimeo, Yale University.

Phillips, P.C.B. & S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 4, 473–495.

Phillips, P.C.B. & 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. & C.C. Lee (1996) Efficiency gains from quasi-differencing under nonstationarity. In P.M. Robinson and M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series: Essays in Memory of E.J. Hannan*. Springer–Verlag.

Phillips, P.C.B. & T. Magdalinos (2007) Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130.
Phillips, P.C.B. & T. Magdalinos (2008) Limit theory for explosively cointegrated systems. *Econometric Theory* 24, 865–887.
Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.
Uhlig, H. (1995) On Jeffreys' prior when using the exact likelihood function. *Econometric Theory* 10, 633–644.
White, J.S. (1958) The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* 29, 1188–1197.

# APPENDIX

The Appendix provides proofs of theorems in the text together with some auxiliary results. We start with preliminary results. The notation is the same as that used in the text.

LEMMA A.1. *Joint convergence in distribution of*

$$\left(\xi_n^0(1), \xi_n(1), n^{-3/2}\sum_{t=1}^{n} Y_{t-1}, n^{-2}\sum_{t=1}^{n} Y_{t-1}Y_{t-1}'\right)$$

*as* $n \to \infty$ *is equivalent to convergence in distribution of each component, where* $\xi_n^0(s) := F(1)\kappa_n^{-1/2}\sum_{j=0}^{\lfloor \kappa_n s\rfloor} \varepsilon_{-j}$, $\xi_n(s) := F(1)n^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor} \varepsilon_t$, *and* $Y_t := \sum_{j=1}^{t} u_j$.

**Proof.** Joint convergence of $\xi_n^0(1)$ and $\xi_n(1)$ holds trivially by independence. We will show that the last two components are asymptotically equivalent to continuous functionals of the partial sum process $\xi_n(\cdot)$ on the Skorohod space $D[0,1]^K$. The lemma will then follow by the continuous mapping theorem and independence of $\xi_n(\cdot)$ and $\xi_n^0(\cdot)$.

The Beveridge-Nelson (BN) decomposition yields, for each $s \in [0,1]$,

$$U_n(s) := \frac{1}{n^{1/2}}\sum_{t=1}^{\lfloor ns\rfloor} u_t = \xi_n(s) - \frac{1}{n^{1/2}}\left(\tilde{\varepsilon}_{\lfloor ns\rfloor} - \tilde{\varepsilon}_0\right), \tag{A.1}$$

where $\tilde{\varepsilon}_t = \sum_{j=0}^{\infty} \tilde{F}_j \varepsilon_{t-j}$ with $\tilde{F}_j = \sum_{k=j+1}^{\infty} F_k$. Using (A.1) and the fact that $Y_0 = 0$, $n^{-3/2}\sum_{t=1}^{n} Y_{t-1}$ can be written as

$$n^{-3/2}\sum_{t=1}^{n} Y_{t-1} = \int_0^1 U_n(s)\,ds = \int_0^1 \xi_n(s)\,ds - \frac{1}{n^{1/2}}\int_0^1 \tilde{\varepsilon}_{\lfloor ns\rfloor}\,ds + o_p(1)$$

$$= \int_0^1 \xi_n(s)\,ds + o_p(1),$$

since $n^{-1/2}\int_0^1 \tilde{\varepsilon}_{\lfloor ns\rfloor}\,ds \to_{L_1} 0$. Similarly, by (A.1),

$$n^{-2}\sum_{t=1}^{n} Y_{t-1}Y_{t-1}' = \int_0^1 U_n(s)\,U_n(s)'ds = \int_0^1 \xi_n(s)\,\xi_n(s)'ds + o_p(1),$$

since $n^{-1/2}\int_0^1 \tilde{\varepsilon}_{\lfloor ns\rfloor}\tilde{\varepsilon}'_{\lfloor ns\rfloor}ds \to_{L_1} 0$ and

$$\begin{aligned}
\mathrm{E}\left\|\frac{1}{n^{1/2}}\int_0^1 \xi_n(s)\tilde{\varepsilon}'_{\lfloor ns\rfloor}ds\right\| &\le \frac{1}{n^{1/2}}\int_0^1 \mathrm{E}\left\|\xi_n(s)\tilde{\varepsilon}'_{\lfloor ns\rfloor}\right\|ds \\
&\le \frac{1}{n^{1/2}}\int_0^1 \mathrm{E}\left(\|\xi_n(s)\|\left\|\tilde{\varepsilon}_{\lfloor ns\rfloor}\right\|\right)ds \\
&\le \frac{1}{n^{1/2}}\int_0^1 \mathrm{E}\left(\|\xi_n(s)\|^2\right)^{1/2}\mathrm{E}\left(\left\|\tilde{\varepsilon}_{\lfloor ns\rfloor}\right\|^2\right)^{1/2}ds \\
&\le \mathrm{E}\left(\|\tilde{\varepsilon}_1\|^2\right)^{1/2}\mathrm{E}\left(\|\varepsilon_1\|^2\right)^{1/2}\frac{1}{n^{1/2}}\int_0^1 \frac{\lfloor ns\rfloor}{n}ds \\
&= O\left(n^{-1/2}\right). \qquad \blacksquare
\end{aligned}$$

LEMMA A.2. *In the setup of Sections 2.4 and 3, there exists a* $K\times(K-1)$ *random orthogonal complement,* $H_\perp$*, to* $B_0(1)$ *satisfying*

$$H'_\perp B_0(1) = 0 \quad and \quad H_\perp H'_\perp = I_K - \left[B_0(1)'B_0(1)\right]^{-1} B_0(1)B_0(1)' \quad a.s. \tag{A.2}$$

*Define* $\underline{B}(s) = B(s) - \int_0^1 B(s)\,ds$, $Z_1 = \left[z_{10}, z_{11}, \ldots, z_{1n-1}\right]'$, *the* $n\times(K-1)$ *matrix* $Z_2 = \left[z'_{20}, z'_{21}, \ldots, z'_{2n-1}\right]'$, *and*

$$\Pi_{1n} = \left(Z'_1 Z_1\right)^{-1} Z'_1 Z_2, \qquad Q_1 = I_n - Z_1\left(Z'_1 Z_1\right)^{-1} Z'_1.$$

*The following hold as* $n\to\infty$ *and* $\frac{n}{\kappa_n}\to 0$:

*(i)* $\Pi_{1n} = O_p\left(\sqrt{\frac{n}{\kappa_n}}\right) = o_p(1)$,

*(ii)* $n^{-1}\sum_{t=1}^n u_t z'_{1t-1}\Pi_{1n} = \xi_n(1)\left(\frac{1}{n^{3/2}}\sum_{t=1}^n Y_{t-1}\right)' H_\perp(n) + o_p(1)$,

*(iii)* $H_\perp(n)\left(n^{-2}Z'_2 Q_1 Z_2\right)^{-1} H_\perp(n)' \Rightarrow H_\perp\left(H'_\perp \int_0^1 \underline{B}\,\underline{B}' H_\perp\right)^{-1} H'_\perp$.

**Proof of (A.2).** We begin by establishing the existence of an orthogonal complement satisfying (A.2) in the setup of Section 2.4. In view of Assumption LP, the asymptotic behavior of the initial condition $x_0(n)$ follows by standard methods (Phillips and Solo, 1992). In particular, letting $B_0 \equiv BM(\Omega)$, we have the functional law

$$\xi_n^0(s) = F(1)\frac{1}{\kappa_n^{1/2}}\sum_{j=0}^{\lfloor\kappa_n s\rfloor}\varepsilon_{-j} \Rightarrow B_0(s), \quad \text{as } n\to\infty, \tag{A.3}$$

which, together with the BN decomposition, yields

$$\kappa_n^{-1/2}x_0(n) = \xi_n^0(1) + o_p(1) \Rightarrow B_0(1). \tag{A.4}$$

By (13) and (A.4), we obtain

$$H(n) = \frac{\xi_n^0(1)}{\left[\xi_n^0(1)'\xi_n^0(1)\right]^{1/2}} + o_p(1) \Rightarrow H := \frac{B_0(1)}{\left[B_0(1)'B_0(1)\right]^{1/2}}. \tag{A.5}$$

Since $\|H\| = 1$, the random matrix $I_K - HH'$ is positive semidefinite with rank $K-1$. Therefore, by a standard decomposition result for positive semidefinite matrices (cf. 8.21 in Abadir and Magnus, 2005), there exists a $K \times (K-1)$ random matrix $H_\perp$ such that, a.s.,

$$H_\perp H_\perp' = I_K - HH' = I_K - \left[B_0(1)' B_0(1)\right]^{-1} B_0(1)B_0(1)'$$

and $H_\perp' H_\perp$ is a diagonal matrix of rank $K-1$ containing the positive eigenvalues of $I_K - HH'$. Since $I_K - HH'$ is idempotent, all its positive eigenvalues are equal to 1, implying that $H_\perp' H_\perp = I_{K-1}$ a.s. Combining the latter with $H_\perp H_\perp' = I_K - HH'$ implies that $H_\perp' H = 0$, so the matrix $H_\perp$ is an orthogonal complement to $H$ (and hence to $B_0(1)$).

Having established the existence of an orthogonal complement $H_\perp$ satisfying (A.2), we can use (14) to write the limiting distribution of the outer product $H_\perp(n)H_\perp(n)'$ as

$$H_\perp(n)H_\perp(n)' = I_K - \frac{\xi_n^0(1)\xi_n^0(1)'}{\xi_n^0(1)'\xi_n^0(1)} + o_p(1) \Rightarrow H_\perp H_\perp'. \tag{A.6}$$

For the setup of Section 3, we can use an identical argument, replacing $\xi_n^0(s)$ by $\xi_{xn}^0(s)$ (defined in (A.29)) and $\Omega$ by $\Omega_{xx}$. ■

**Proof of Lemma A2(i).** First, note that, by (5),

$$\begin{aligned}\frac{1}{\kappa_n^{1/2}n^{3/2}}\sum_{t=1}^n x_{t-1}Y_{t-1}' &= \frac{x_0(n)}{\kappa_n^{1/2}}\frac{1}{n^{3/2}}\sum_{t=1}^n Y_{t-1}' + \frac{1}{\kappa_n^{1/2}n^{3/2}}\sum_{t=1}^n Y_{t-1}Y_{t-1}' \\ &= \frac{x_0(n)}{\kappa_n^{1/2}}\frac{1}{n^{3/2}}\sum_{t=1}^n Y_{t-1}' + O_p\left(\sqrt{\frac{n}{\kappa_n}}\right).\end{aligned} \tag{A.7}$$

Thus, since, by (A.5) and (A.6), $H(n)$ and $H_\perp(n)$ are $O_p(1)$, (A.4) yields

$$Z_1'Z_2 = H(n)'\sum_{t=1}^n x_{t-1}Y_{t-1}'H_\perp(n) = O_p\left(\kappa_n^{1/2}n^{3/2}\right).$$

The result for $\Pi_{1n}$ follows, since $\left(Z_1'Z_1\right)^{-1} = O_p\left(\kappa_n^{-1}n^{-1}\right)$. ■

**Proof of Lemma A2(ii).** By (A.19) and (A.5),

$$\frac{1}{n\kappa_n}Z_1'Z_1 = H(n)'\frac{1}{\kappa_n n}\sum_{t=1}^n x_{t-1}x_{t-1}'H(n) = \xi_n^0(1)'\xi_n^0(1) + o_p(1), \tag{A.8}$$

which, together with (A.7), (A.4), and (A.5), yields

$$\Pi_{1n} = \left(\frac{Z_1'Z_1}{n\kappa_n}\right)^{-1} H(n)' \frac{1}{n\kappa_n}\sum_{t=1}^{n} x_{t-1}Y_{t-1}' H_\perp(n)$$

$$= \left(\frac{Z_1'Z_1}{n\kappa_n}\right)^{-1} H(n)' \left[\sqrt{\frac{n}{\kappa_n}}\frac{x_0(n)}{\sqrt{\kappa_n}}\frac{1}{n^{3/2}}\sum_{t=1}^{n} Y_{t-1}' + O_p\left(\frac{n}{\kappa_n}\right)\right] H_\perp(n)$$

$$= \sqrt{\frac{n}{\kappa_n}}\left[\xi_n^0(1)'\xi_n^0(1)\right]^{-1/2}\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} Y_{t-1}\right)' H_\perp(n) + o_p\left(\sqrt{\frac{n}{\kappa_n}}\right). \tag{A.9}$$

Thus, by (A.17) and (A.5), we obtain

$$\frac{1}{n}\sum_{t=1}^{n} u_t z_{1t-1}' \Pi_{1n}$$

$$= \frac{\left[\xi_n^0(1)'\xi_n^0(1)\right]^{-1/2}}{\sqrt{n\kappa_n}}\sum_{t=1}^{n} u_t x_{t-1}' H(n)\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} Y_{t-1}\right)' H_\perp(n)$$

$$+ o_p\left(\sqrt{\frac{n}{\kappa_n}}\frac{1}{n}\sum_{t=1}^{n} u_t x_{t-1}'\right)$$

$$= \left[\xi_n^0(1)'\xi_n^0(1)\right]^{-1/2}\xi_n(1)\xi_n^0(1)' H(n)\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} Y_{t-1}\right)' H_\perp(n) + o_p(1)$$

$$= \xi_n(1)\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} Y_{t-1}\right)' H_\perp(n) + o_p(1). \tag{A.10}$$

■

**Proof of Lemma A2(iii).** We first show that

$$n^{-2}Z_2'Q_1Z_2 = H_\perp(n)'T_n H_\perp(n) + o_p(1), \tag{A.11}$$

where $T_n$ denotes the random matrix

$$T_n = \frac{1}{n^2}\sum_{t=1}^{n} Y_{t-1}Y_{t-1}' - \left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} Y_{t-1}\right)\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} Y_{t-1}\right)'.$$

By an application of (14) we can write

$$\frac{Z_2'Z_1}{\kappa_n^{1/2}n^{3/2}}\frac{Z_1'Z_2}{\kappa_n^{1/2}n^{3/2}} = H_\perp(n)'\frac{\sum_{t=1}^{n} Y_{t-1}x_{t-1}'}{\kappa_n^{1/2}n^{3/2}} H(n)H(n)'\frac{\sum_{t=1}^{n} x_{t-1}Y_{t-1}'}{\kappa_n^{1/2}n^{3/2}} H_\perp(n)$$

$$= H_\perp(n)'\frac{\sum_{t=1}^{n} Y_{t-1}x_{t-1}'}{\kappa_n^{1/2}n^{3/2}}\frac{\sum_{t=1}^{n} x_{t-1}Y_{t-1}'}{\kappa_n^{1/2}n^{3/2}} H_\perp(n)$$

$$- H_\perp(n)' \frac{\sum_{t=1}^n Y_{t-1} x_{t-1}'}{\kappa_n^{1/2} n^{3/2}} H_\perp(n) H_\perp(n)' \frac{\sum_{t=1}^n x_{t-1} Y_{t-1}'}{\kappa_n^{1/2} n^{3/2}} H_\perp(n)$$

$$= H_\perp(n)' \frac{\sum_{t=1}^n Y_{t-1} x_{t-1}'}{\kappa_n^{1/2} n^{3/2}} \frac{\sum_{t=1}^n x_{t-1} Y_{t-1}'}{\kappa_n^{1/2} n^{3/2}} H_\perp(n)$$

$$- H_\perp(n)' \frac{\sum_{t=1}^n Y_{t-1} Y_{t-1}'}{\kappa_n^{1/2} n^{3/2}} H_\perp(n) H_\perp(n)' \frac{\sum_{t=1}^n Y_{t-1} Y_{t-1}'}{\kappa_n^{1/2} n^{3/2}} H_\perp(n)$$

$$= H_\perp(n)' \frac{\sum_{t=1}^n Y_{t-1} x_{t-1}'}{\kappa_n^{1/2} n^{3/2}} \frac{\sum_{t=1}^n x_{t-1} Y_{t-1}'}{\kappa_n^{1/2} n^{3/2}} H_\perp(n) + O_p\left(\frac{n}{\kappa_n}\right),$$

where $H_\perp(n)' \sum_{t=1}^n x_{t-1} Y_{t-1}' = H_\perp(n)' \sum_{t=1}^n Y_{t-1} Y_{t-1}'$ because of (5) and the fact that $H_\perp(n)' x_0(n) = 0$. Thus, (A.7) and (A.4) yield

$$\frac{Z_2' Z_1}{\kappa_n^{1/2} n^{3/2}} \frac{Z_1' Z_2}{\kappa_n^{1/2} n^{3/2}} = \xi_n^0(1)' \xi_n^0(1) H_\perp(n) \frac{1}{n^{3/2}} \sum_{t=1}^n Y_{t-1} \frac{1}{n^{3/2}} \sum_{j=1}^n Y_{j-1}' H_\perp(n)' + o_p(1),$$

and (A.11) follows by (A.8) and the identity $Z_2' Z_2 = H_\perp(n)' \sum_{t=1}^n Y_{t-1} Y_{t-1}' H_\perp(n)$ since

$$\frac{1}{n^2} Z_2' Q_1 Z_2 = \frac{1}{n^2} Z_2' Z_2 - \left(\frac{Z_1' Z_1}{n\kappa_n}\right)^{-1} \frac{Z_2' Z_1}{\kappa_n^{1/2} n^{3/2}} \frac{Z_1' Z_2}{\kappa_n^{1/2} n^{3/2}}$$

$$= \frac{1}{n^2} Z_2' Z_2 - H_\perp(n)' \frac{1}{n^{3/2}} \sum_{t=1}^n Y_{t-1} \frac{1}{n^{3/2}} \sum_{j=1}^n Y_{j-1}' H_\perp(n) + o_p(1)$$

$$= H_\perp(n)' T_n H_\perp(n) + o_p(1).$$

Having established (A.11), the limiting distribution of

$$H_\perp(n) \left(n^{-2} Z_2' Q_1 Z_2\right)^{-1} H_\perp(n)' = H_\perp(n) \left[H_\perp(n)' T_n H_\perp(n)\right]^{-1} H_\perp(n)' + o_p(1)$$

is derived as follows: By Lemma A1, $\left(\xi_n^0(1), T_n\right) \Rightarrow (B_0(1), T)$, where $T = \int_0^1 \underline{BB'}$. So (A.6) implies that

$$\left(H_\perp(n) H_\perp(n)', T_n\right) \Rightarrow \left(H_\perp H_\perp', T\right). \tag{A.12}$$

Thus, the Skorohod representation theorem implies that there exist random matrices $\left(L_n, \tilde{T}_n\right)$ and $\left(L, \tilde{T}\right)$ defined on the same probability space for all $n \in \mathbb{N}$ such that $\left(L_n, \tilde{T}_n\right) =_d \left(H_\perp(n) H_\perp(n)', T_n\right)$ and $\left(L_n, \tilde{T}_n\right) \to_{\text{a.s.}} \left(L, \tilde{T}\right)$ as $n \to \infty$. By (A.12), $\left(L, \tilde{T}\right) =_d \left(H_\perp H_\perp', T\right)$. Denote by $M^+$ the Moore-Penrose inverse of a matrix $M$. Since the rank of both $L_n\tilde{T}_nL_n$ and $L\tilde{Q}L$ is $K-1$ a.s., Theorem 2 of Andrews (1987) yields

$$\begin{aligned} H_\perp(n)\left[H_\perp(n)'T_nH_\perp(n)\right]^{-1}H_\perp(n)' &= \left[H_\perp(n)H_\perp(n)'T_nH_\perp(n)H_\perp(n)'\right]^+ \\ &=_d \left(L_n\tilde{T}_nL_n\right)^+ \to_{\text{a.s.}} \left(L\tilde{T}L\right)^+ \\ &=_d \left(H_\perp H_\perp' T H_\perp H_\perp'\right)^+ \\ &= H_\perp\left(H_\perp' T H_\perp\right)^{-1}H_\perp', \end{aligned}$$

which proves part (iii) of the lemma. ∎

**Proofs of Theorems 1 and 2.** The limit theory for sample moments involving trajectories of $x_t$ may incorporate elements from both initial conditions and sample period observations depending on the behavior of $\kappa_n$ as $n\to\infty$. Decompose $x_t$ as $x_t = x_0(n)+Y_t$, as in (5). Recalling that $Y_0=0$, the limit behavior of $Y_t$ and its sample moments is standard (Phillips and Durlauf, 1986; Phillips, 1988b), viz.,

$$n^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor} u_t = \xi_n(s)+o_p(1), \qquad \xi_n(s) := F(1)n^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor}\varepsilon_t \Rightarrow F(1)\Sigma^{1/2}W(s), \tag{A.13}$$

and

$$n^{-3/2}\sum_{t=1}^{n} Y_{t-1} \Rightarrow \int_0^1 B,$$

$$n^{-2}\sum_{t=1}^{n} Y_{t-1}Y_{t-1}' \Rightarrow \int_0^1 BB',\ n^{-1}\sum_{t=1}^{n} u_tY_{t-1}' \Rightarrow \int_0^1 dBB'+\Lambda, \tag{A.14}$$

where $B = F(1)\Sigma^{1/2}W \equiv BM(\Omega)$, $\Omega = F(1)\Sigma F(1)'$, $\Lambda = \sum_{h=1}^{\infty}\mathrm{E}\left(u_tu_{t-h}'\right)$, and $W$ is standard $K$-vector Brownian motion. By virtue of the independence of the $\varepsilon_t$, the processes

$$\kappa_n^{-1/2}\sum_{j=0}^{\lfloor \kappa_n s\rfloor} u_{-j} = \xi_n^0(s)+o_p(1) \quad \text{and} \quad n^{-1/2}\sum_{t=1}^{\lfloor ns\rfloor} u_t = \xi_n(s)+o_p(1) \tag{A.15}$$

are asymptotically independent for all $s\in[0,1]$, and so the Brownian motions $B_0$ and $B$ are also independent. The asymptotic equivalences in (A.15) follow by employing the BN decomposition and partial summation as in Phillips and Solo (1992), in view of the summability assumption in (4).

The effect of the initial condition on the asymptotic behavior of the sample moments of $x_t$ can be obtained by comparing the convergence rate of $x_0(n)$ with that of the sample moments of $Y_t$. First,

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n} u_tx_{t-1}' &= \sqrt{\frac{\kappa_n}{n}}\frac{1}{n^{1/2}}\sum_{t=1}^{n} u_t\left(\frac{x_0(n)}{\kappa_n^{1/2}}\right)' + \frac{1}{n}\sum_{t=1}^{n} u_tY_{t-1}' \\ &= \sqrt{\tau}\xi_n(1)\xi_n^0(1)' + \frac{1}{n}\sum_{t=1}^{n} u_tY_{t-1}' + o_p(1) \end{aligned}$$

$$\Rightarrow \sqrt{\tau}B(1)B_0(1)' + \int_0^1 dBB' + \Lambda$$

$$= \int_0^1 dBB_\tau^{+\prime} + \Lambda, \tag{A.16}$$

where $B_\tau^+(s) = B(s) + \sqrt{\tau}B_0(1)$, giving the limit result for recent ($\tau = 0$) and distant ($0 < \tau < \infty$) past initializations. For infinite ($\tau = \infty$) past initializations, (A.16) requires rescaling so that

$$\frac{1}{\sqrt{\kappa_n n}} \sum_{t=1}^{n} u_t x_{t-1}' = \xi_n(1)\xi_n^0(1)' + O_p\left(\sqrt{\frac{n}{\kappa_n}}\right) \Rightarrow B(1)B_0(1)', \tag{A.17}$$

and the sample moments involving $Y_t$ are asymptotically negligible under the revised standardization, thereby eliminating the components that produce the usual unit root limit theory. Instead, the asymptotic behavior of the sample covariance $\sum_{t=1}^n u_t x_{t-1}'$ is determined exclusively by the infinite past initialization $x_0(n)$ and partial sums of $u_t$.

For $\tau \in [0, \infty)$, the sample moment matrix of $x_t$ has the expanded form

$$\frac{1}{n^2} \sum_{t=1}^{n} x_{t-1}' x_{t-1}' = \frac{\kappa_n}{n} \frac{x_0(n)x_0(n)'}{\kappa_n} + \sqrt{\frac{\kappa_n}{n}} \frac{1}{n^{3/2}} \sum_{t=1}^{n} Y_{t-1} \frac{x_0(n)'}{\kappa_n^{1/2}}$$

$$+ \sqrt{\frac{\kappa_n}{n}} \frac{x_0(n)}{\kappa_n^{1/2}} \frac{1}{n^{3/2}} \sum_{t=1}^{n} Y_{t-1}' + \frac{1}{n^2} \sum_{t=1}^{n} Y_{t-1} Y_{t-1}'$$

$$= \int_0^1 B_\tau^+ B_\tau^{+\prime}, \tag{A.18}$$

giving the limit result for recent and distant past initializations. Under infinite ($\tau = \infty$) past initializations, the sample moment matrix has a faster rate of convergence that is driven by the behavior of $x_0(n)$. In particular,

$$\frac{1}{\kappa_n n} \sum_{t=1}^{n} x_{t-1} x_{t-1}' = \xi_n^0(1)\xi_n^0(1)' + O_p\left(\sqrt{\frac{n}{\kappa_n}}\right) \Rightarrow B_0(1)B_0(1)', \tag{A.19}$$

producing a singular limit for the sample moment matrix unless (3) is a scalar autoregression ($K = 1$). In the scalar case, (A.17) and (A.19) yield

$$\sqrt{\kappa_n n}\left(\hat{R}_n - 1\right) = \frac{\frac{1}{\sqrt{\kappa_n n}} \sum_{t=1}^n x_{t-1} u_t}{\frac{1}{\kappa_n n} \sum_{t=1}^n x_{t-1}^2} = \frac{\frac{F(1)}{n^{1/2}} \sum_{t=1}^n \varepsilon_t}{\frac{F(1)}{\kappa_n^{1/2}} \sum_{j=0}^{\kappa_n} \varepsilon_{-j}} + O_p\left(\sqrt{\frac{n}{\kappa_n}}\right) \Rightarrow \frac{B(1)}{B_0(1)} =_d \mathcal{C},$$

where $\mathcal{C}$ is a standard Cauchy variate, giving the scalar result of Theorem 2. In this case where $\tau = \infty$, the tail of the process from the origination of $x_t$ wags the dog in the limit theory of estimator. The distribution depends on the past through $B_0$ and the sample through $B$.

Combining (A.16) and (A.18), we have the least squares regression limit theory for (3) under recent or distant past initializations

$$n\left(\hat{R}_n - I_K\right) \Rightarrow \left(\int_0^1 dBB_\tau^{+\prime} + \Lambda\right) \left(\int_0^1 B_\tau^+ B_\tau^{+\prime}\right)^{-1},$$

as stated in Theorem 1. ■

**Proof of Theorem 3.** Set $Z = [Z_1, Z_2]$. In the new coordinates given by (15), we have

$$n\left(\hat{R}_n - I_K\right) = \left(\frac{1}{n}\sum_{t=1}^{n} u_t z_{t-1}'\right)\left(\frac{1}{n^2}\sum_{t=1}^{n} z_{t-1}z_{t-1}'\right)^{-1} M(n)'$$

$$= \left(\frac{1}{n}\sum_{t=1}^{n} u_t z_{t-1}'\right)\begin{bmatrix} n^{-2}Z_1'Z_1 & n^{-2}Z_1'Z_2 \\ n^{-2}Z_2'Z_1 & n^{-2}Z_2'Z_2 \end{bmatrix}^{-1}\begin{bmatrix} H(n)' \\ H_\perp(n)' \end{bmatrix}. \quad \textbf{(A.20)}$$

Now standard partitioned inversion gives

$$\left(n^{-2}Z'Z\right)^{-1} = \begin{bmatrix} \left(\frac{Z_1'Z_1}{n^2}\right)^{-1} + \Pi_{1n}\left(\frac{Z_2'Q_1Z_2}{n^2}\right)^{-1}\Pi_{1n}' & -\Pi_{1n}\left(\frac{Z_2'Q_1Z_2}{n^2}\right)^{-1} \\ -\left(\frac{Z_2'Q_1Z_2}{n^2}\right)^{-1}\Pi_{1n}' & \left(\frac{Z_2'Q_1Z_2}{n^2}\right)^{-1} \end{bmatrix}, \quad \textbf{(A.21)}$$

where

$$\Pi_{1n} = \left(Z_1'Z_1\right)^{-1} Z_1'Z_2 \quad \text{and} \quad Q_1 = I_n - Z_1\left(Z_1'Z_1\right)^{-1} Z_1'.$$

Combining (A.20) and (A.21), the least squares estimator becomes

$$\begin{aligned} n\left(\hat{R}_n - I_K\right) &= \frac{1}{n}\sum_{t=1}^{n} u_t z_{1t-1}'\left(\frac{Z_1'Z_1}{n^2}\right)^{-1} H(n)' \\ &\quad + \frac{1}{n}\sum_{t=1}^{n} u_t z_{1t-1}'\Pi_{1n}\left(\frac{Z_2'Q_1Z_2}{n^2}\right)^{-1}\left[\Pi_{1n}'H(n)' - H_\perp(n)'\right] \\ &\quad + \frac{1}{n}\sum_{t=1}^{n} u_t z_{2t-1}'\left(\frac{Z_2'Q_1Z_2}{n^2}\right)^{-1}\left[H_\perp(n)' - \Pi_{1n}'H(n)'\right]. \end{aligned} \quad \textbf{(A.22)}$$

By (A.5) and (A.6) we know that both $H(n)$ and $H_\perp(n)$ are bounded in probability. Thus, recalling that the effect of the initial condition is present only in $z_{1t-1}$, we have

$$Z_1'Z_1 = O_p(n\kappa_n), \sum_{t=1}^{n} u_t z_{1t-1}' = O_p\left(\sqrt{n\kappa_n}\right) \quad \text{and} \quad \sum_{t=1}^{n} u_t z_{2t-1}' = O_p(n). \quad \textbf{(A.23)}$$

The asymptotic behavior of the remaining terms of (A.22) is given in Lemma A2 above. Consideration of these terms leads to the simplification

$$\begin{aligned} n\left(\hat{R}_n - I_K\right) &= \left[\frac{1}{n}\sum_{t=1}^{n} u_t z_{2t-1}' - \frac{1}{n}\sum_{t=1}^{n} u_t z_{1t-1}'\Pi_{1n}\right] \\ &\quad \times \left(\frac{Z_2'Q_1Z_2}{n^2}\right)^{-1} H_\perp(n)' + O_p\left(\sqrt{\frac{n}{\kappa_n}}\right) \end{aligned}$$

$$= \left[ \frac{1}{n} \sum_{t=1}^{n} u_t Y_{t-1}' - \xi_n(1) \left( \frac{1}{n^{3/2}} \sum_{t=1}^{n} Y_{t-1} \right)' \right]$$

$$\times H_{\perp}(n) \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} H_{\perp}(n)' + o_p(1). \tag{A.24}$$

Joint convergence in distribution of the various elements in (A.24) needs to be proved. The proof of Lemma A2 (iii) yields

$$H_{\perp}(n) \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} H_{\perp}(n)' = \left[ H_{\perp}(n) H_{\perp}(n)' T_n H_{\perp}(n) H_{\perp}(n)' \right]^{+} + o_p(1),$$

which together with (A.6), implies that the right-hand side of (A.24) is a continuous function of

$$\left( \xi_n^0(1), \xi_n(1), n^{-3/2} \sum_{t=1}^{n} Y_{t-1}, n^{-2} \sum_{t=1}^{n} Y_{t-1} Y_{t-1}', \frac{1}{n} \sum_{t=1}^{n} u_t Y_{t-1}' \right).$$

Joint convergence of the first four terms has been established in Lemma A1. The sample covariance $n^{-1} \sum_{t=1}^{n} u_t Y_{t-1}'$ does not admit a neat integral representation like the other two sample moments. The stochastic component of its limiting distribution is nonetheless driven by the partial sum process $U_n(\cdot)$ in (A.1), and joint convergence of $n^{-1} \sum_{t=1}^{n} u_t Y_{t-1}'$ and other sample moments of $Y_t$ is well documented (cf. Phillips, 1988a, 1988b, 1988c). Thus, it is enough to show that $n^{-1} \sum_{t=1}^{n} u_t Y_{t-1}' - \Lambda$ is asymptotically independent of $\xi_n^0(\cdot)$. To see this, note that $n^{-1} \sum_{t=1}^{n} \tilde{\varepsilon}_t u_t' \to_{\text{a.s.}} \mathrm{E}\left(\tilde{\varepsilon}_t u_t'\right) = \Lambda$ by the ergodic theorem and a simple calculation. Using the BN decomposition and summation by parts, we can write

$$\frac{1}{n} \sum_{t=1}^{n} u_t Y_{t-1}' - \Lambda = \frac{F(1)}{n} \sum_{t=1}^{n} \varepsilon_t Y_{t-1}' - \frac{1}{n} \tilde{\varepsilon}_n Y_n' + \frac{1}{n} \sum_{t=1}^{n} \tilde{\varepsilon}_t u_t' - \Lambda$$

$$= \frac{F(1)}{n} \sum_{t=1}^{n} \varepsilon_t Y_{t-1}' + o_p(1)$$

$$= \frac{F(1)}{n} \sum_{t=1}^{n} \varepsilon_t \sum_{j=1}^{t-1} \varepsilon_j' F(1)' + \frac{F(1)}{n} \sum_{t=1}^{n} \varepsilon_t \tilde{\varepsilon}_{t-1}' + o_p(1)$$

$$= \frac{F(1)}{n} \sum_{t=1}^{n} \varepsilon_t \sum_{j=1}^{t-1} \varepsilon_j' F(1)' + o_p(1), \tag{A.25}$$

since $n^{-1} \sum_{t=1}^{n} \varepsilon_t \tilde{\varepsilon}_{t-1}' \to 0$ in $L_2$ by a martingale LLN. This establishes the required asymptotic independence.

Since joint convergence of the various terms of (A.24) applies, (A.14), (A.13), and Lemma A2(iii) give

$$n\left(\hat{R}_n - I_K\right) \Rightarrow \left\{ \int_0^1 dB B' + \Lambda - B(1) \int_0^1 B' \right\} H_{\perp} \left( H_{\perp}' \int_0^1 \underline{B} \underline{B}' H_{\perp} \right)^{-1} H_{\perp}'$$

$$= \left( \int_0^1 dB\underline{B}' + \Lambda \right) H_\perp \left( H_\perp' \int_0^1 \underline{BB}' H_\perp \right)^{-1} H_\perp',$$

as stated in (17).

For (18), using the fact that

$$\frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^n u_t z_{1t-1}' = \frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^n u_t x_0(n)' H(n) + O_p\left( \sqrt{\frac{n}{\kappa_n}} \right)$$

$$= \left[ \xi_n^0(1)' \xi_n^0(1) \right]^{1/2} \xi_n(1) + o_p(1)$$

and that $H(n)' H(n) = 1$, $H_\perp(n)' H(n) = 0$ a.s., (A.22) gives

$$\sqrt{n\kappa_n} \left( \hat{R}_n - I_K \right) H(n)$$

$$= \frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^n u_t z_{1t-1}' \left( \frac{Z_1' Z_1}{n\kappa_n} \right)^{-1} - \frac{1}{n} \sum_{t=1}^n u_t z_{2t-1}' \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} \left( \sqrt{\frac{\kappa_n}{n}} \Pi_{1n} \right)'$$

$$+ \frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^n u_t z_{1t-1}' \left( \sqrt{\frac{\kappa_n}{n}} \Pi_{1n} \right) \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} \left( \sqrt{\frac{\kappa_n}{n}} \Pi_{1n} \right)'$$

$$= \left[ \xi_n^0(1)' \xi_n^0(1) \right]^{-1/2} \xi_n(1) - \left[ \xi_n^0(1)' \xi_n^0(1) \right]^{-1/2} \left\{ \sum_{t=1}^n \frac{u_t Y_{t-1}'}{n} - \xi_n(1) \left( \sum_{t=1}^n \frac{Y_{t-1}}{n^{3/2}} \right)' \right\}$$

$$\times H_\perp(n) \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} H_\perp(n)' \left( \frac{1}{n^{3/2}} \sum_{t=1}^n Y_{t-1} \right)$$

$$\Rightarrow \left[ B_0(1)' B_0(1) \right]^{-1/2} \left[ B(1) - \left\{ \int_0^1 dB\underline{B}' + \Lambda \right\} H_\perp \left( H_\perp' \int_0^1 \underline{BB}' H_\perp \right)^{-1} H_\perp' \int_0^1 \underline{B} \right],$$

where we have used (A.8), (A.9), (A.14), Lemma A2, and joint convergence developed in Lemma A1 and (A.25). ∎

**Proof of Theorem 4.** Setting $u_{y.xt} = u_{yt} - \Omega_{yx}\Omega_{xx}^{-1}\Delta x_t$ and $U_{y.x} = \left[ u_{y.x1}', \ldots, u_{y.xn}' \right]'$ as the corresponding data matrix, we have

$$\hat{y}_t^+ = y_t - \hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1}\Delta x_t = A x_t + u_{0.xt} - \left( \hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1} \right) \Delta x_t,$$

and, letting $U_x := \Delta X = \left[ u_{x1}', \ldots, u_{xn}' \right]'$, we obtain

$$\left( \hat{A}^+ - A \right) = \left( U_{y.x}' X - n\hat{\Delta}_{yx}^+ - \left( \hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1} \right) U_x' X \right) \left( X'X \right)^{-1}. \quad \textbf{(A.26)}$$

Writing (A.26) in the rotated coordinates (15) and using the inversion formula (A.21), we have

$$n\left( \hat{A}^+ - A \right)$$

$$= \left( \frac{U_{y.x}' Z}{n} - \hat{\Delta}_{yz}^+ - \left( \hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1} \right) \frac{U_x' Z}{n} \right) \left( n^{-2} Z'Z \right)^{-1} M(n)'$$

$$
= \left(\frac{1}{n}\sum_{t=1}^{n} u_{y.xt} z_{1t} - \hat{\Delta}^{+}_{yz_1} - \left(\hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1}\right)\frac{U_x' Z_1}{n}\right)\left(\frac{Z_1' Z_1}{n^2}\right)^{-1} H(n)'
$$
$$
+ \left(\frac{1}{n}\sum_{t=1}^{n} u_{y.xt} z_{1t} - \hat{\Delta}^{+}_{yz_1} - \left(\hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1}\right)\frac{U_x' Z_1}{n}\right)
$$
$$
\times \Pi_{1n}\left(\frac{Z_2' Q_1 Z_2}{n^2}\right)^{-1}\left[\Pi_{1n}' H(n)' - H_{\perp}(n)'\right]
$$
$$
+ \left(\frac{1}{n}\sum_{t=1}^{n} u_{y.xt} z_{2t}' - \hat{\Delta}^{+}_{yz_2} - \left(\hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1}\right)\frac{U_x' Z_2}{n}\right)
$$
$$
\times \left(\frac{Z_2' Q_1 Z_2}{n^2}\right)^{-1}\left[H_{\perp}(n)' - \Pi_{1n}' H(n)'\right]. \qquad \textbf{(A.27)}
$$

In order to analyze the components of (A.27), note that, by an identical argument to Lemma A2, both $\sum_{t=1}^{n} u_{y.xt} z_{1t}$ and $U_x' Z_1$ are of order $O_p\left(\sqrt{n\kappa_n}\right)$, $\sum_{t=1}^{n} u_{y.xt} z_{2t}'$ and $U_x' Z_2$ are of order $O_p(n)$, $Z_1' Z_1 = O_p(n\kappa_n)$, and $Z_2' Z_2 = O_p\left(n^2\right)$. Also, given an integrable lag kernel function $k(\cdot)$ and a lag truncation parameter $M$ satisfying $\frac{1}{M} + \frac{M}{n} \to 0$,

$$
\hat{\Delta}^{+}_{yz} = \sum_{h=0}^{M} k\left(\frac{h}{M}\right)\frac{1}{n}\sum_{t=1}^{n} \hat{u}_{y.xt}\Delta z_{t-h}' = \sum_{h=0}^{M} k\left(\frac{h}{M}\right)\frac{1}{n}\sum_{t=1}^{n} \hat{u}_{y.xt}\Delta x_{t-h}' M(n)
$$
$$
= \hat{\Delta}^{+}_{yx} M(n) = O_p(1),
$$

since $\hat{\Delta}^{+}_{yx}$ is a consistent estimator (cf. Phillips and Hansen, 1990) and $M(n) = O_p(1)$. Thus both $\hat{\Delta}^{+}_{yz_1}$ and $\hat{\Delta}^{+}_{yz_2}$ are bounded in probability. Finally, as both $\hat{\Omega}_{yx}$ and $\hat{\Omega}_{xx}$ are consistent estimators (Phillips and Hansen, 1990), $\hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1} = o_p(1)$.

The above facts imply, for the first term of (A.27),

$$
\frac{n}{\kappa_n}\left(\frac{1}{n}\sum_{t=1}^{n} u_{y.xt} z_{1t} - \hat{\Delta}^{+}_{yz_1} - \left(\hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1}\right)\frac{U_x' Z_1}{n}\right)\left(\frac{Z_1' Z_1}{n\kappa_n}\right)^{-1}
$$
$$
= \left(\frac{1}{\kappa_n}\sum_{t=1}^{n} u_{y.xt} z_{1t} - \frac{n}{\kappa_n}\hat{\Delta}^{+}_{yz_1} - o_p(1)\frac{1}{\kappa_n} U_x' Z_1\right) O_p(1) = O_p\left(\sqrt{\frac{n}{\kappa_n}}\right).
$$

Since $n/\kappa_n \to 0$, this shows that the first term of (A.27) is $o_p(1)$ as $n \to \infty$. For the second term of (A.27), since $\Pi_{1n} = O_p\left(\sqrt{\frac{n}{\kappa_n}}\right)$ from Lemma A2, we obtain

$$
\left(\frac{1}{n}\sum_{t=1}^{n} u_{y.xt} z_{1t} - \hat{\Delta}^{+}_{yz_1} - \left(\hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1}\right)\frac{U_x' Z_1}{n}\right)\Pi_{1n}\left(\frac{Z_2' Q_1 Z_2}{n^2}\right)^{-1}\Pi_{1n}' H(n)'
$$
$$
= \left(\frac{1}{n}\sum_{t=1}^{n} u_{y.xt} z_{1t} - \hat{\Delta}^{+}_{yz_1} - \left(\hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1}\right)\frac{U_x' Z_1}{n}\right) O_p\left(\frac{n}{\kappa_n}\right)
$$

$$= \left( \frac{1}{\kappa_n} \sum_{t=1}^{n} u_{y.xt} z_{1t} - \frac{n}{\kappa_n} \hat{\Delta}^{+}_{yz_1} - o_p(1) \frac{U_x' Z_1}{\kappa_n} \right) O_p(1) = O_p\left( \sqrt{\frac{n}{\kappa_n}} \right),$$

and

$$\left[ \hat{\Delta}^{+}_{yz_1} + \left( \hat{\Omega}_{yx} \hat{\Omega}_{xx}^{-1} - \Omega_{yx} \Omega_{xx}^{-1} \right) \frac{U_x' Z_1}{n} \right] \Pi_{1n} \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} H_{\perp}(n)'$$

$$= O_p\left( \sqrt{\frac{n}{\kappa_n}} \right) + \left( \hat{\Omega}_{yx} \hat{\Omega}_{xx}^{-1} - \Omega_{yx} \Omega_{xx}^{-1} \right) \frac{U_x' Z_1}{n} O_p\left( \sqrt{\frac{n}{\kappa_n}} \right)$$

$$= O_p\left( \sqrt{\frac{n}{\kappa_n}} \right) + o_p(1) \frac{U_x' Z_1}{\sqrt{n \kappa_n}} O_p(1) = o_p(1) O_p(1) = o_p(1).$$

A similar argument on the third term of (A.27) yields

$$\left( \frac{1}{n} \sum_{t=1}^{n} u_{y.xt} z_{2t}' - \hat{\Delta}^{+}_{yz_2} - \left( \hat{\Omega}_{yx} \hat{\Omega}_{xx}^{-1} - \Omega_{yx} \Omega_{xx}^{-1} \right) \frac{U_x' Z_2}{n} \right) \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} \Pi_{1n}' H(n)'$$

$$= O_p(1) \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} \Pi_{1n}' H(n)' = O_p\left( \|\Pi_{1n}\| \right) = O_p\left( \sqrt{\frac{n}{\kappa_n}} \right),$$

and

$$\left( \hat{\Omega}_{yx} \hat{\Omega}_{xx}^{-1} - \Omega_{yx} \Omega_{xx}^{-1} \right) \frac{U_x' Z_2}{n} \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} H_{\perp}(n)' = o_p(1).$$

Thus, (A.27) yields

$$n\left( \hat{A}^{+} - A \right) = \left[ \frac{1}{n} \sum_{t=1}^{n} u_{y.xt} z_{2t}' - \hat{\Delta}^{+}_{yz_2} - \frac{1}{n} \sum_{t=1}^{n} u_{y.xt} z_{1t} \Pi_{1n} \right]$$

$$\times \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} H_{\perp}(n)' + o_p(1). \qquad \textbf{(A.28)}$$

Corresponding to the notation of Lemma A1, let $Y_{xt} := \sum_{j=1}^{t} u_{xj}$,

$$\xi^{0}_{xn}(s) := F_x(1) \kappa_n^{-1/2} \sum_{j=0}^{\lfloor \kappa_n s \rfloor} \varepsilon_{-j}, \qquad \xi_{Jn}(s) := F_J(1) n^{-1/2} \sum_{t=1}^{\lfloor ns \rfloor} \varepsilon_t, \qquad \textbf{(A.29)}$$

for $J \in \{x, y\}$, and

$$\xi^{+}_{y.xn}(s) := \xi_{yn}(s) - \Omega_{yx} \Omega_{xx}^{-1} \xi_{xn}(s) \Rightarrow B_{y.x}(s) \equiv \mathrm{BM}\left( \Omega_{y.xx} \right).$$

Then, an identical argument to that used in the derivation of (A.10) in Lemma A2(ii) yields

$$\frac{1}{n} \sum_{t=1}^{n} u_{y.xt} z_{1t}' \Pi_{1n} = \frac{1}{n} \sum_{t=1}^{n} u_{yt} z_{1t}' \Pi_{1n} - \Omega_{yx} \Omega_{xx}^{-1} \frac{1}{n} \sum_{t=1}^{n} u_{xt} z_{1t}' \Pi_{1n}$$

$$= \left[ \xi_{yn}(1) - \Omega_{yx} \Omega_{xx}^{-1} \xi_{xn}(1) \right] \left( \frac{1}{n^{3/2}} \sum_{t=1}^{n} Y_{xt} \right)' H_{\perp}(n) + o_p(1)$$

$$= \xi^{+}_{y.xn}(1)\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} Y_{xt}\right)' H_{\perp}(n) + o_p(1)$$

$$\Rightarrow B_{y.x}(1)\int_0^1 B_x. \qquad \textbf{(A.30)}$$

Also, using Lemma A2(iii) with $\underline{B}_x$ in place of $\underline{B}$ we obtain

$$H_{\perp}(n)\left(\frac{Z_2'Q_1Z_2}{n^2}\right)^{-1} H_{\perp}(n)' \Rightarrow H_{\perp}\left(H_{\perp}'\int_0^1 \underline{B}_x\underline{B}_x' H_{\perp}\right)^{-1} H_{\perp}'. \qquad \textbf{(A.31)}$$

Thus, for the final component of (A.28), the fact that $\hat{\Delta}^{+}_{yz_2} = \hat{\Delta}^{+}_{yx}H_{\perp}(n) = \hat{\Delta}^{+}_{yY}H_{\perp}(n)$ yields

$$\left(\frac{1}{n}\sum_{t=1}^{n} u_{y.xt}z_{2t}' - \hat{\Delta}^{+}_{yz_2}\right)\left(\frac{Z_2'Q_1Z_2}{n^2}\right)^{-1} H_{\perp}(n)'$$

$$= \left[\frac{1}{n}\sum_{t=1}^{n} u_{y.xt}Y_{xt}' - \hat{\Delta}^{+}_{yY}\right] H_{\perp}(n)\left(\frac{Z_2'Q_1Z_2}{n^2}\right)^{-1} H_{\perp}(n)'$$

$$\Rightarrow \left(\int_0^1 dB_{y.x}B_x'\right) H_{\perp}\left(H_{\perp}'\int_0^1 \underline{B}_x\underline{B}_x' H_{\perp}\right)^{-1} H_{\perp}', \qquad \textbf{(A.32)}$$

as in (A.25) and using Lemma A2(iii).

Substituting (A.30), (A.31), and (A.32) into (A.28) and using joint convergence of the various elements as in the proof of Theorem 3, we obtain

$$n\left(\hat{A}^{+} - A\right) \Rightarrow \left\{\left(\int_0^1 dB_{y.x}B_x'\right) - B_{y.x}(1)\int_0^1 B_x\right\} H_{\perp}\left(H_{\perp}'\int_0^1 \underline{B}_x\underline{B}_x' H_{\perp}\right)^{-1} H_{\perp}'$$

$$= \left(\int_0^1 dB_{y.x}\underline{B}_x'\right) H_{\perp}\left(H_{\perp}'\int_0^1 \underline{B}_x\underline{B}_x' H_{\perp}\right)^{-1} H_{\perp}'$$

and vectorizing producing the stated result (22). Mixed normality holds because the limit process $B_{y.x}$ is independent of both $B_x$ and $B_0$.

It remains to show (23). From (A.27), we have

$$\sqrt{n\kappa_n}\left(\hat{A}^{+} - A\right) H(n)$$

$$= \left(\frac{1}{\sqrt{n\kappa_n}}\sum_{t=1}^{n} u_{y.xt}z_{1t}' - \sqrt{\frac{n}{\kappa_n}}\hat{\Delta}^{+}_{yz_1} - \left(\hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1}\right)\frac{U_x'Z_1}{\sqrt{n\kappa_n}}\right)\left(\frac{Z_1'Z_1}{n\kappa_n}\right)^{-1}$$

$$+ \sqrt{\frac{\kappa_n}{n}}\left(\frac{1}{n}\sum_{t=1}^{n} u_{y.xt}z_{1t}' - \hat{\Delta}^{+}_{yz_1} - \left(\hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1} - \Omega_{yx}\Omega_{xx}^{-1}\right)\frac{U_x'Z_1}{n}\right)$$

$$\times \Pi_{1n}\left(\frac{Z_2'Q_1Z_2}{n^2}\right)^{-1} \Pi_{1n}'$$

$$- \sqrt{\frac{\kappa_n}{n}} \left( \frac{1}{n} \sum_{t=1}^{n} u_{y.xt} z_{2t}' - \hat{\Delta}_{yz_2}^{+} - \left( \hat{\Omega}_{yx} \hat{\Omega}_{xx}^{-1} - \Omega_{yx} \Omega_{xx}^{-1} \right) \frac{U_x' Z_2}{n} \right)$$

$$\times \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} \Pi_{1n}'$$

$$= \frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^{n} u_{y.xt} z_{1t}' \left( \frac{Z_1' Z_1}{n\kappa_n} \right)^{-1} + \sqrt{\frac{\kappa_n}{n}} \frac{1}{n} \sum_{t=1}^{n} u_{y.xt} z_{1t}' \Pi_{1n} \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} \Pi_{1n}'$$

$$- \sqrt{\frac{\kappa_n}{n}} \left( \frac{1}{n} \sum_{t=1}^{n} u_{y.xt} z_{2t}' - \hat{\Delta}_{yz_2}^{+} \right) \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} \Pi_{1n}' + o_p(1), \qquad \textbf{(A.33)}$$

using similar arguments for the remainder terms as those used in the derivation of (A.28). Using (A.8) and the fact that $\sum_{t=1}^{n} u_{y.xt} Y_{xt}' = O_p(n)$, the first term of (A.33) becomes

$$\frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^{n} u_{y.xt} z_{1t}' = \frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^{n} u_{y.xt} x_t' H(n) \left( \frac{Z_1' Z_1}{n\kappa_n} \right)^{-1}$$

$$= \frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^{n} u_{y.xt} x_0(n)' H(n) \left( \frac{Z_1' Z_1}{n\kappa_n} \right)^{-1} + O_p\left( \sqrt{\frac{n}{\kappa_n}} \right)$$

$$= \left[ \frac{x_0(n)' x_0(n)}{\kappa_n} \right]^{1/2} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_{y.xt} \left( \frac{Z_1' Z_1}{n\kappa_n} \right)^{-1} + O_p\left( \sqrt{\frac{n}{\kappa_n}} \right)$$

$$= \left[ \xi_{xn}^{0}(1)' \xi_{xn}^{0}(1) \right]^{-1/2} \xi_{y.xn}^{+}(1) + o_p(1)$$

$$\Rightarrow \left[ B_0(1)' B_0(1) \right]^{-1/2} B_{y.x}(1). \qquad \textbf{(A.34)}$$

Similarly, letting

$$T_{xn} = \frac{1}{n^2} \sum_{t=1}^{n} Y_{xt} Y_{xt}' - \left( \frac{1}{n^{3/2}} \sum_{t=1}^{n} Y_{xt} \right) \left( \frac{1}{n^{3/2}} \sum_{t=1}^{n} Y_{xt} \right)'$$

and using the above, (A.9), (A.11), and Lemma A2, the second term of (A.33) can be written as

$$\left( \frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^{n} u_{y.xt} z_{1t}' \right) \left( \sqrt{\frac{\kappa_n}{n}} \Pi_{1n} \right) \left( \frac{Z_2' Q_1 Z_2}{n^2} \right)^{-1} \left( \sqrt{\frac{\kappa_n}{n}} \Pi_{1n} \right)'$$

$$= \left[ \xi_{xn}^{0}(1)' \xi_{xn}^{0}(1) \right]^{-1/2} \xi_{y.xn}^{+}(1) \left( \frac{1}{n^{3/2}} \sum_{t=1}^{n} Y_{xt} \right)'$$

$$\times H_{\perp}(n)\left[H_{\perp}(n)' T_{xn} H_{\perp}(n)\right]^{-1} H_{\perp}(n)'\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} Y_{xt}\right)+o_p(1)$$

$$\Rightarrow \left[B_0(1)' B_0(1)\right]^{-1/2} B_{y.x}(1)\int_0^1 B_x' H_{\perp}\left(H_{\perp}'\int_0^1 \underline{B}_x \underline{B}_x' H_{\perp}\right)^{-1} H_{\perp}'\int_0^1 B_x. \tag{A.35}$$

For the third term of (A.33), since $\hat{\Delta}^{+}_{yz_2} = \hat{\Delta}^{+}_{yY} H_{\perp}(n)$, we obtain

$$\sqrt{\frac{\kappa_n}{n}}\left(\frac{1}{n}\sum_{t=1}^{n} u_{y.xt} z_{2t}' - \hat{\Delta}^{+}_{yz_2}\right)\left(\frac{Z_2' Q_1 Z_2}{n^2}\right)^{-1}\Pi_{1n}'$$

$$= \left(\frac{1}{n}\sum_{t=1}^{n} u_{y.xt} Y_{xt}' - \hat{\Delta}^{+}_{yY}\right) H_{\perp}(n)\left(\frac{Z_2' Q_1 Z_2}{n^2}\right)^{-1}\left(\sqrt{\frac{\kappa_n}{n}}\Pi_{1n}\right)'$$

$$= \left[\xi_n^0(1)'\xi_n^0(1)\right]^{-1/2}\left(\frac{1}{n}\sum_{t=1}^{n} u_{y.xt} Y_{xt}' - \hat{\Delta}^{+}_{yY}\right) H_{\perp}(n)\left(\frac{Z_2' Q_1 Z_2}{n^2}\right)^{-1} H_{\perp}(n)'$$

$$\times\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} Y_{xt}\right)+o_p(1)$$

$$\Rightarrow \left[B_0(1)' B_0(1)\right]^{-1/2}\left(\int_0^1 dB_{y.x} B_x'\right) H_{\perp}\left(H_{\perp}'\int_0^1 \underline{B}_x \underline{B}_x' H_{\perp}\right)^{-1} H_{\perp}'\int_0^1 B_x. \tag{A.36}$$

Applying (A.34), (A.35), and (A.36) to (A.33) and using the joint weak convergence of the random elements of Lemma A1, we obtain

$$\sqrt{n\kappa_n}\left(\hat{A}^{+} - A\right) H(n)$$

$$\Rightarrow B_{y.x}(1)\left\{B_0(1)' B_0(1)\right\}^{-1/2}$$

$$+ B_{y.x}(1)\left\{B_0(1)' B_0(1)\right\}^{-1/2}\left(\int_0^1 B_x\right)' H_{\perp}\left(H_{\perp}'\int_0^1 \underline{B}_x \underline{B}_x' H_{\perp}\right)^{-1} H_{\perp}'\left(\int_0^1 B_x\right)$$

$$- \left(\int_0^1 dB_{y.x} B_x'\right) H_{\perp}\left(H_{\perp}'\int_0^1 \underline{B}_x \underline{B}_x' H_{\perp}\right)^{-1} H_{\perp}'\left(\int_0^1 B_x\right)\left\{B_0(1)' B_0(1)\right\}^{-1/2}$$

$$= B_{y.x}(1)\left\{B_0(1)' B_0(1)\right\}^{-1/2}$$

$$- \left(\int_0^1 dB_{y.x} \underline{B}_x'\right) H_{\perp}\left(H_{\perp}'\int_0^1 \underline{B}_x \underline{B}_x' H_{\perp}\right)^{-1} H_{\perp}'\left(\int_0^1 B_x\right)\left\{B_0(1)' B_0(1)\right\}^{-1/2}$$

$$= \frac{\int_0^1 dB_{y.x} - \left(\int_0^1 dB_{y.x}\underline{B}_x'\right) H_{\perp}\left(H_{\perp}'\int_0^1 \underline{B}_x \underline{B}_x' H_{\perp}\right)^{-1} H_{\perp}'\left(\int_0^1 B_x\right)}{\left\{B_0(1)' B_0(1)\right\}^{1/2}}$$

$$= \frac{\int_0^1 dB_{y.x}(s)\left[1 - \underline{B}_x(s)' H_\perp \left(H_\perp' \int_0^1 \underline{B}_x \underline{B}_x' H_\perp\right)^{-1} H_\perp' \left(\int_0^1 B_x\right)\right]}{\left\{B_0(1)' B_0(1)\right\}^{1/2}}$$

$$\equiv MN\left(0, \Omega_{yy.x} \int_0^1 V(s)^2 ds / B_0(1)' B_0(1)\right),$$

where

$$V(s) = 1 - \underline{B}_x(s)' H_\perp \left(H_\perp' \int_0^1 \underline{B}_x \underline{B}_x' H_\perp\right)^{-1} H_\perp' \left(\int_0^1 B_x\right),$$

as required for (23). Again, mixed normality holds because $B_{y.x}$ is independent of both $B_x$ and $B_0$. ■