HOW TO ESTIMATE AUTOREGRESSIVE ROOTS NEAR UNITY

BY

PETER C.B. PHILLIPS, HYUNGSIK ROGER MOON and ZHIJIE XIAO

COWLES FOUNDATION PAPER NO. 1028

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2001

http://cowles.econ.yale.edu/

Econometric Theory, **17**, 2001, 29–69. Printed in the United States of America.

HOW TO ESTIMATE AUTOREGRESSIVE ROOTS NEAR UNITY

PETER C.B. PHILLIPS
Yale University, University of Auckland, and University of York

HYUNGSIK ROGER MOON
University of Southern California

ZHIJIE XIAO
University of Illinois at Urbana-Champaign

A new model of near integration is formulated in which the local to unity parameter is identifiable and consistently estimable with time series data. The properties of the model are investigated, new functional laws for near integrated time series are obtained that lead to mixed diffusion processes, and consistent estimators of the localizing parameter are constructed. The model provides a more complete interface between $I(0)$ and $I(1)$ models than the traditional local to unity model and leads to autoregressive coefficient estimates with rates of convergence that vary continuously between the $O(\sqrt{n})$ rate of stationary autoregression, the $O(n)$ rate of unit root regression, and the power rate of explosive autoregression. Models with deterministic trends are also considered, least squares trend regression is shown to be efficient, and consistent estimates of the localizing parameter are obtained for this case also. Conventional unit root tests are shown to be consistent against local alternatives in the new class.

1. INTRODUCTION

Models with near unit roots have attracted much attention in recent years. These models lead to a class of near integrated time series that offer some additional flexibility over integrated processes in the modeling of nonstationary time series. They were developed originally to provide a mechanism for studying local alternatives to unit root specifications, giving limit diffusion processes in place of Brownian motion (Bobkoski, 1983; Phillips, 1987a), unifying asymptotics for stationary and nonstationary autoregressions (Chan and Wei, 1987;

This paper was partly written while Phillips was living on Waiheke Island and visiting the University of Auckland. It was presented to the New Zealand Econometric Study Group Meeting at the University of Auckland in July 1998. We thank Katsuto Tanaka and a referee for helpful comments on earlier versions of this paper. Phillips thanks the NSF for research support under grants SBR 94-22922 and 97-30295, and Xiao thanks the Cowles Foundation for hospitality during a visit in May–June 1998. Address correspondence to: Peter Phillips, Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520-8281, USA.

Phillips, 1987a), having natural extensions to vector time series (Phillips, 1988), and delivering power functions and power envelopes for unit root tests (Cavanagh, 1985; Phillips, 1987a; Johansen, 1991). They have also been used in empirical econometric work to construct confidence bands that allow for autoregressive coefficients and roots in the neighborhood of unity (Cavanagh, 1985; Stock, 1991).

The simplest local to unity model is a triangular array for a time series y_t of the form

$$y_t = ay_{t-1} + u_t, \qquad a = 1 + \frac{c}{n}, \qquad t = 1, \ldots, n \tag{1}$$

with independent and identically distributed (i.i.d.) $(0, \sigma^2)$ innovations u_t. Whereas the autoregressive coefficient $a \to 1$ as $n \to \infty$, it is apparent that for any given sample size n in (1), the model accommodates a much wider range of autoregressive coefficients as the localizing parameter c varies, including both stationary $(c < 0)$, explosive $(c > 0)$, and unit root $(c = 0)$ possibilities. This flexibility has helped to make the model popular in studying economic time series for which roots near unity are considered highly plausible but roots at unity are considered too restrictive. A feature of the local to unity model is that the localizing parameter is identifiable (c can be deduced from the conditional mean ay_{t-1} and the sample size n) but is not consistently estimable. In particular, standardized observations from the model (1) satisfy the invariance principle

$$n^{-1/2} y_{[nr]} \Rightarrow J_c(r), \tag{2}$$

a linear diffusion process (Phillips, 1987a) that depends on c. So, writing the model in the form $\Delta y_t = c(y_{t-1}/n) + u_t$, it is apparent that the sample second moment of the regressor $x_t = y_{t-1}/n$ of c satisfies the weak convergence

$$\sum_{t=1}^{n} \left(\frac{y_{t-1}}{n} \right)^2 \Rightarrow \int_0^1 J_c(r)^2 dr$$

and does not diverge as $n \to \infty$, thereby failing to satisfy the excitation condition for least squares regression consistency. Put another way, the signal to noise ratio measured by

$$\frac{\mathrm{Var}(x_t)}{\mathrm{var}(u_t)} \sim \frac{\frac{1}{n} \sum_{t=1}^{n} \left(\frac{y_{t-1}}{n} \right)^2}{\sigma^2} \to_p 0,$$

and so the signal from x_t is too weak relative to the error variation to produce a consistent estimator of the localizing coefficient c.

Although methods have been developed to utilize the way in which the limit distribution depends on the localizing coefficient (by virtue of the dependence of the limit process $J_c(r)$ on c), the failure of consistent estimation has been an impediment to inference in models of this type. The dependence of the limit distribution on c also affects resampling procedures such as the bootstrap, which are known to be inconsistent in models of this type because of this very dependence (Basawa, Lallik, McCormick, Reeves, and Taylor, 1991). One way in which the signal can be strengthened is through the use of additional data. In fact, recent work by Moon and Phillips (2000) shows how panel data with independent cross section observations are helpful in resolving the failure of consistency in time series models such as (1). This approach relies on the fact that the model (1) continues to apply with the same localizing coefficient across a section of N individual observations while $N \to \infty$. Then, $\sqrt{N}$ consistent estimation of c is possible. However, panel data for which the assumptions underlying this approach are plausible, particularly that of cross section homogeneity of the localizing parameter, seem likely to be uncommon. So, these panel data results seem at present to be of more theoretical than empirical import.

This paper offers a fresh approach to the problem of modeling time series with roots near unity. Our idea is to develop a new formulation of local to unity models that offers more flexibility than the traditional model (1). The new model leads to a class of different limit processes beyond simple diffusions, and it has the interesting property that the local coefficient is identifiable and consistently estimable with time series data, unlike (1). Consistent estimation opens up some new possibilities with respect to efficient estimation, trend elimination, and the construction of confidence intervals. The new model also provides a more complete interface between $I(0)$ and $I(1)$ models and between $O(\sqrt{n})$ and $O(n)$ asymptotics. In the traditional model (1), the rate of convergence in autoregressive coefficient estimation is $O(n)$, just as in the unit root case $c = 0$, and there continues to be a discontinuity in the asymptotics between the stationary and nonstationary cases. Only as $c \to -\infty, +\infty$ in the traditional model do we find results that correspond to the stationary and explosive autoregressions (Phillips, 1987a; Chan and Wei, 1987). By contrast, in our new model, the rate of convergence to the autoregressive coefficient is $O(n^{\alpha})$ for $\alpha \in [\frac{1}{2}, 1]$ and varies in a continuous way between that of stationary and nonstationary asymptotics. The new model also captures the power law asymptotics of explosive autoregressions and shows that, in a well defined local region greater than unity, it is possible to obtain invariance principles, in contrast to standard results for the explosive autoregression.

The paper is organized as follows. Notation is given in Section 2. The new model is laid out and some of its properties are analyzed in Section 3. A consistent estimator of the local to unity coefficient is constructed in Section 4 and cases of near stationarity, unit roots, and near explosive behavior are separately analyzed. Estimation of the local parameter in models with linear trends is discussed in Section 5. Section 6 studies issues of efficient estimation of trend coefficients and trend extraction. Section 7 concludes and describes some useful extensions of the present model. Proofs are collected in the Appendix.

2. NOTATION

$\to_{a.s.}$	almost sure convergence	$\Rightarrow, \to_d$	weak convergence
$=_d$	distributional equivalence	$[\cdot]$	integer part of
$:=$	definitional equality	$r \wedge s$	$\min(r,s)$
$o_{a.s.}(1)$	tends to zero almost surely	$\equiv$	equivalence
$\to_p$	convergence in probability	$o_p(1)$	tends to zero in probability
$W_k(r)$	standard Brownian motion $\forall k$	QD	quasi-difference
$BM(\omega^2)$	Brownian motion with variance ω^2		

3. A BLOCK LOCAL TO UNITY MODEL

The time series model we propose is a block local to unity system defined as follows:

$$\begin{aligned} &y_{k,t} = a y_{k,t-1} + u_{k,t}, \qquad t \in \mathbb{T}_m; k \in \mathbb{K}_K, \\ &y_{k,0} = y_{k-1,m}, \\ &\quad a = e^{c/m} \sim 1 + \frac{c}{m}, \end{aligned} \tag{3}$$

where $\mathbb{T}_m = \{1,\ldots,m\}$, $\mathbb{K}_K = \{-K, -K+1, \ldots, 0, 1, \ldots, M\}$ with $K \geq 0$. This system defines a sequence of blocks with m observations of the time series $\{y_{k,t} : t \in \mathbb{T}_m\}$ in each block, and the observable blocks are taken to be $k = 1,\ldots,M$. The initial conditions in each block are set so that they correspond to the final observation in the previous block. In this sense, the model is articulated to capture the evolution of a single time series. The observable series is $\{y_{k,t} : t \in \mathbb{T}_m; k = 1,\ldots,M\}$.

The coefficient in the autoregression in each block of (3) is local to unity with localizing parameter c, which is the same in each block. In later sections of the paper, depending on the sign of c, we will allow for various initial conditions, and the index set $\mathbb{K}_K$ for the blocks is introduced to provide this extra flexibility. Our initial conditions are described in the following assumption.

Assumption 1 (Initial Conditions).

(i) Infinite past initialization: $K = \infty$ with index set $\mathbb{K}_\infty$.

(ii) Distant past initialization: $K = 0$ with index set $\mathbb{K}_0$ and

$$y_{0,0} = \sum_{j=0}^{K} a^j u_{-1,-j} \qquad K = [m\kappa], \tag{4}$$

where the $u_{-1,-j}$ are independent of $u_{k,t}$ in (3), and

$m^{-1/2} y_{0,0} \Rightarrow J_{-1,c}(-\kappa)$,

where $J_{-1,c}(-\kappa) = \int_{-\kappa}^{0} e^{-(s+\kappa)c} dB_{-1}(s)$ is a reverse diffusion process and B_{-1} is a Brownian motion.

We use a general linear process generating mechanism for the errors $u_{k,t}$ in each block of (3). The idea is that there is an underlying sequence of innovations ε_j from whose present and past history the errors in each block are formed.

We further allow for the specific generating mechanism to change between blocks, thereby permitting some structural change across blocks in the short memory component of the model. The specific structure is laid out in the following assumption.

Assumption 2 (Linear Process Errors).

(i) $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is a sequence of i.i.d. (0,1) variates with $E|\varepsilon_t|^p < \infty$ for some $p > 4$.
(ii) $u_{k,t} = \sum_{j=0}^{\infty} b_{k,j}\varepsilon_{k,t-j}$, where $\varepsilon_{k,t} = \varepsilon_{mk+t}$.
(iii) $\sum_{j=0}^{\infty} j^a b_j < \infty$, for some $a \geq 1$, where $b_j := \sup_k |b_{k,j}|$.
(iv) Let $\omega_k^2 = (\sum_{j=0}^{\infty} b_{k,j})^2$, and assume that $\inf_k \omega_k^2 > 0$.
(v) $\mu_2 = \lim_{M\to\infty}(1/M)\sum_{k=1}^{M}\omega_k^2$ and $\mu_4 = \lim_{M\to\infty}(1/M)\sum_{k=1}^{M}\omega_k^4$ exist.

Remarks.

(a) When $b_{k,j} = b_j$ for all k, the time series $u_{k,t}$ have homogeneous (over k) generating mechanisms as measurable functions of the primitive innovations ε_t and differ only in terms of the timing of the shocks with each new block k bringing in a new block of primitive innovations $\varepsilon_{k,t}$. This special framework will apply, for example, when a single parametric model such as an AR(p) governs the formation of the shocks $u_{k,t}$ in every block k, so that the parameters in this model are the same for all k.
(b) Condition (iii) on the majorizing sequence b_j for the linear process coefficients $b_{k,j}$ ensures the validity of a BN decomposition for $u_{k,t}$ for each k, as in Phillips and Solo (1992) (see the discussion in the Appendix of the current paper). It also ensures that $\sup_k \omega_k^2 < \infty$.
(c) The moment condition in (i) and the summability condition (iii) ensure that fourth moments of u_t are finite.
(d) The parameters μ_2 and μ_4 in (v) are average long run variance parameter and square of long run variance parameter over the blocks in (3), respectively.

We write the data from a particular block as $y^k = (y_{k,1},\ldots,y_{k,m})'$ and then combine data from M blocks to write $y = (y^{1\prime}, y^{2\prime},\ldots,y^{M\prime})'$. In this case, the total sample size is $n = mM$.

By recursive substitution we have the representation

$$\begin{aligned}
m^{-1/2}y_{k,[mr]} &= m^{-1/2}\sum_{j=0}^{[mr]-1} e^{jc/m}u_{k,[mr]-j} + m^{-1/2}e^{[mr]c/m}y_{k,0} \\
&= m^{-1/2}\sum_{j=0}^{[mr]-1} e^{jc/m}u_{k,[mr]-j} + m^{-1/2}e^{[mr]c/m}y_{k-1,m} \\
&= m^{-1/2}\sum_{j=0}^{[mr]-1} e^{jc/m}u_{k,[mr]-j} + e^{[mr]c/m} \\
&\quad \times \sum_{f=0}^{k-1} e^{(k-1-f)c}\left[m^{-1/2}\sum_{j=0}^{m-1} e^{jc/m}u_{f,m-j}\right] \\
&\quad + m^{-1/2}e^{[mr]c/m}e^{kc}y_{0,0}.
\end{aligned}$$

If $c < 0$ and the initial conditions are in the infinite past, then we can write

$$m^{-1/2} y_{k,[mr]} = m^{-1/2} \sum_{j=0}^{[mr]-1} e^{jc/m} u_{k,[mr]-j} + e^{[mr]c/m} \sum_{f=-\infty}^{k-1} e^{(k-1-f)c} \left[m^{-1/2} \sum_{j=0}^{m-1} e^{jc/m} u_{f,m-j} \right], \tag{5}$$

where the second series converges in the mean square sense (see the proof of convergence in mean square of (5) in the Appendix).

LEMMA 1. *Let $U_j^M = (u_{0,j}, \ldots, u_{M,j})'$. Then, under Assumption* 2, *for any fixed M, as $m \to \infty$,*

$$\frac{1}{\sqrt{m}} \sum_{j=1}^{[mr]} U_j^M \Rightarrow B^M(r),$$

where $B^M(r) = (B_0(r), \ldots, B_M(r))' \equiv BM(\Omega_M)$, $\Omega_M = \text{diag}(\omega_0^2, \ldots, \omega_M^2)$.

As in Phillips (1987a), we have the weak convergence

$$m^{-1/2} \sum_{j=0}^{[mr]-1} e^{jc/m} U_{[mr]-j}^M \Rightarrow J_c^M(r) = \int_0^r e^{(r-s)c} dB^M(s), \tag{6}$$

where $J_c^M(r) = (J_{0,c}(r), \ldots, J_{M,c}(r))'$ and $J_{k,c}(r) = \int_0^r e^{(r-s)c} dB_k(s)$ is a linear diffusion. It follows by the continuous mapping theorem that if the initial conditions are in the distant past at $y_{0,0}$ and satisfy Assumption 1(ii), then we have

$$\begin{aligned} m^{-1/2} y_{k,[mr]} &\Rightarrow J_{k,c}(r) + e^{rc} \sum_{f=0}^{k-1} e^{(k-1-f)c} J_{f,c}(1) + e^{(r+k)c} J_{-1,c}(-\kappa) \\ &:= H_{k,c}^{\kappa}(r). \end{aligned} \tag{7}$$

If $c < 0$ and the initial conditions are in the infinite past, it follows that

$$m^{-1/2} y_{k,[mr]} \Rightarrow J_{k,c}(r) + e^{rc} \sum_{f=-\infty}^{k-1} e^{(k-1-f)c} J_{f,c}(1) := H_{k,c}(r) \tag{8}$$

(see the proof of equation (8) in the Appendix). Note that $\sum_{f=-\infty}^{k-1} \times e^{(k-1-f)c} J_{f,c}(1)$ converges because $\sum_{j=0}^{\infty} e^{2jc} < \infty$ and $E(J_{f,c}(1)^2) < \infty$ and because $\{J_{f,c}(1)\}_{f=-\infty}^{\infty}$ is a sequence of independent diffusion processes.

Note that the limit processes $H_{k,c}(r)$ and $H_{k,c}^{\kappa}(r)$ involve linear combinations of independent (across f) diffusion processes $J_{f,c}$ and are therefore both Gaussian. They may be called mixed diffusion processes. The expression $H_{k,c}(r)$ is defined for $c < 0$. However, $H_{k,c}^{\kappa}(r)$ involves only a finite linear combination of terms when k is finite, so it is also well defined when $c \geq 0$. Both these limit processes differ from the usual diffusion limit (2) that applies for the traditional local to unity model. The block structure of the model (3) ensures that the traditional diffusion limits apply within each block to linear combinations of the shocks in each block, as in (6). But the observable data cover M blocks with progressive reinitializations of the process to assure the compatibility of the block structure with the observed time series. The new limit processes $H_{k,c}(r)$ and $H_{k,c}^{\kappa}(r)$ of the normalized observed data take these progressive reinitializations into account.

The device of a block local to unity system facilitates the sequential asymptotic analysis that is used later in the paper, and it also provides a statistical model for what may be described as "isolated regions of persistent behavior" for macroeconomic time series. Many macroeconomic time series are now well known to display a form of persistence whereby economic shocks have long run effects. However, it is possible that shocks may affect an economy for a long period of time but not forever. In other words, the effects of a shock may be highly persistent over a certain range (the region of persistent behavior) but then may begin to disappear outside this range. The region of persistent behavior may constitute a "little infinity" relative to the full sample. Consider a time series, $\{z_s\}$, which evolves over blocks of time in such a way that there is persistency inside each block but only short memory across blocks, i.e.,

$$\underbrace{z_1, z_2, \ldots z_m}_{\text{Block } 1}, \underbrace{z_{m+1}, \ldots, z_{2m}}_{\text{Block } 2}, \ldots, \underbrace{z_{km+1}, \ldots, z_{(k+1)m}}_{\text{Block } k+1}, \ldots .$$

The number of observations in each block is m, and the number of blocks is M. The block local to unity system (3) (when $c < 0$) is a simple model that has this property. Because there is persistent memory inside each block but short memory across blocks, we call this type of memory "regional persistence." As a result, the partial sums inside each block have nonstationary asymptotic behavior, whereas partial sums over blocks behave like a stationary system.

4. ESTIMATION OF THE LOCAL PARAMETER

4.1. The Near Stationary Case: $c < 0$

In this section we assume that the initial conditions are in the infinite past. We propose to estimate the autoregressive coefficient by the usual least squares estimator, which we write here in pooled form as

$$\hat{a} = \frac{y_{-1}'y}{y_{-1}'y_{-1}} = \frac{\sum_{k=1}^{M} y_{-1}^{k\prime} y^k}{\sum_{k=1}^{M} y_{-1}^{k\prime} y_{-1}^k}.$$

From this estimator, we are able to extract a corresponding estimate of the localizing coefficient c. Using the model formulation $y^k = a y_{-1}^k + u^k$, we get

$$m(\hat{a} - a) = \frac{\sum_{k=1}^{M} \frac{1}{m} y_{-1}^{k\prime} u^k}{\sum_{k=1}^{M} \frac{1}{m^2} y_{-1}^{k\prime} y_{-1}^k}.$$

Asymptotic results for this estimator can be obtained most conveniently by using sequential asymptotics in which $m \to \infty$ first, followed by $M \to \infty$, which we denote by $(m, M \to \infty)_{seq}$. This type of asymptotic analysis will be used throughout the paper. Sequential asymptotics are discussed in Phillips and Moon (1999), which also explores the connections between this type of asymptotic analysis and joint limit theory in which two indices such as (m, M) may pass to infinity simultaneously. Although less general than joint limit theory, sequential asymptotics are easy to obtain and will serve our purpose in this paper of revealing the main features of the block local to unity system. As the analysis in Phillips and Moon (1999) indicates, we can expect the main results obtained here under sequential asymptotics to hold for joint limits under somewhat stronger conditions.

For fixed M, we have, as in Phillips (1987a), that as $m \to \infty$,

$$\frac{1}{m} y_{-1}^{k\prime} u^k \Rightarrow \int_0^1 H_{k,c}(r) dB_k(r) + \lambda_k$$

and

$$\frac{1}{m^2} y_{-1}^{k\prime} y_{-1}^k \Rightarrow \int_0^1 H_{k,c}(r)^2 dr,$$

where $\lambda_k = \sum_{j=1}^{\infty} b_{k,0} b_{k,j}$. It follows that as $m \to \infty$,

$$m(\hat{a} - a) = \frac{\sum_{k=1}^{M} \frac{1}{m} y_{-1}^{k\prime} u^k}{\sum_{k=1}^{M} \frac{1}{m^2} y_{-1}^{k\prime} y_{-1}^k} \Rightarrow \frac{\sum_{k=1}^{M} \left(\int_0^1 H_{k,c} dB_k + \lambda_k \right)}{\sum_{k=1}^{M} \int_0^1 H_{k,c}(r)^2 dr}.$$

We may now employ the usual nonparametric corrections (Phillips, 1987b) to $\hat{a}$ that use consistent estimates $\hat{\lambda}_k$ of λ_k giving the following modified estimator:

$$\hat{a}^+ = \frac{\sum_{k=1}^{M} (y_{-1}^{k\prime} y^k - m\hat{\lambda}_k)}{\sum_{k=1}^{M} y_{-1}^{k\prime} y_{-1}^k}.$$

It will be convenient in what follows to make the following high level assumption about the nonparametric estimates such as $\hat{\lambda}_k$ that we use in our development.

Assumption 3 (Nonparametric Estimation of λ_k and ω_k). Use δ to represent both λ and ω in (i)–(iii). Then

(i) $\hat{\delta}_k \to_p \delta_k$ as $m \to \infty$, $\forall k$.
(ii) $\sqrt{mh}(\hat{\delta}_k - \delta_k) \to_d N(0, V_k)$ as $m \to \infty$, $\forall k$, where h is the bandwidth used in the construction of the estimate $\hat{\delta}_k$.
(iii) $\sup_k V_k < \infty$.

Parts (i) and (ii) of this assumption will be satisfied by a wide class of nonparametric estimates of δ_k under Assumption 2 (see Hannan, 1970; Park and Phillips, 1988; Andrews, 1991). Part (ii) will typically be satisfied when there is undersmoothing of the estimate $\hat{\delta}_k$ through the choice of bandwidth h, to ensure the absence of bias in the limiting normal distribution. Part (iii) simply bounds the limiting variances V_k over k.

The error in the estimator $\hat{a}^+$ is

$$\sqrt{M}m(\hat{a}^+ - a) = \frac{\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \left(\frac{1}{m} y_{-1}^{k\prime} u^k - \hat{\lambda}_k \right)}{\frac{1}{M} \sum_{k=1}^{M} \frac{1}{m^2} y_{-1}^{k\prime} y_{-1}^k}$$

$$\sim \frac{\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \left[\int_0^1 H_{k,c}\, dB_k + (\lambda_k - \hat{\lambda}_k) \right]}{\frac{1}{M} \sum_{k=1}^{M} \int_0^1 H_{k,c}(r)^2 dr} \tag{9}$$

$$= \frac{\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \int_0^1 H_{k,c}\, dB_k}{\frac{1}{M} \sum_{k=1}^{M} \int_0^1 H_{k,c}(r)^2 dr} + o_p(1), \tag{10}$$

provided $M^{-1/2} \sum_{k=1}^{M} (\lambda_k - \hat{\lambda}_k) = o_p(1)$, which holds under Assumption 3, as shown in the Appendix. Now we consider taking limits as $M \to \infty$. By applying a suitable strong law of large numbers (SLLN) to $(1/M) \sum_{k=1}^{M} \int_0^1 H_{k,c}(r)^2 dr$ and a suitable central limit theorem (CLT) to $(1/\sqrt{M}) \sum_{k=1}^{M} [\int_0^1 H_{k,c}\, dB_k]$ as $M \to \infty$, it can be verified that $\hat{a}^+$ converges to a at the rate $\sqrt{M}m$ and, further, that $\sqrt{M}m(\hat{a}^+ - a)$ has an asymptotic normal distribution. In particular, we have the following result, the proof of which is in the Appendix.

THEOREM 2. *Let Assumptions* 1(i), 2, *and* 3, *hold and let* $c < 0$. *Then*, *as* $(m, M \to \infty)_{seq}$

$$\sqrt{M}m(\hat{a}^{+} - a) \Rightarrow N(0, V_a),$$

where

$$V_a = V_H^{-1}\left(\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M}\omega_k^2 E\left[\int_0^1 H_{k,c}^2\right]\right)V_H^{-1} \tag{11}$$

and

$$V_H = \lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M} E\left[\int_0^1 H_{k,c}^2\right].$$

It is shown in the proof of Theorem 2 that

$$\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M} E\left[\int_0^1 H_{k,c}^2\right] = -\frac{1}{2c}\left(\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M}\omega_k^2\right) := -\frac{1}{2c}\mu_2$$

and

$$\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M}\omega_k^2 E\left[\int_0^1 H_{k,c}^2\right] = -\frac{1}{2c}\left(\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M}\omega_k^4\right) := -\frac{1}{2c}\mu_4.$$

It follows that $V_a = (-2c)(\mu_4/\mu_2^2)$. When the errors are homogeneous across k, we get $\omega_k^2 = \omega^2$ for all k and then $\mu_4 = \mu_2^2 = \omega^4$ and $V_a = -2c$. Because $\mu_4 \geq \mu_2^2$, $V_a = -2c$ is a lower bound for the limiting variance in the general case where the long run variances vary across blocks.

The weighting in the limit variance V_a in the general case (11) indicates that we can improve the efficiency of the estimator $\hat{a}^{+}$ by means of a weighted regression. Let $\hat{\omega}_k^2$ be a nonparametric estimate of ω_k^2 satisfying Assumption 3. Define the semiparametric weighted regression estimator

$$\hat{a}_{\omega}^{+} = \frac{\sum_{k=1}^{M}\frac{1}{\hat{\omega}_k^2}\left(y_{-1}^{k\prime}y^k - m\hat{\lambda}_k\right)}{\sum_{k=1}^{M}\frac{1}{\hat{\omega}_k^2}y_{-1}^{k\prime}y_{-1}^k}.$$

The following result shows that the asymptotic theory of $\hat{a}_{\omega}^{+}$ is very simple.

THEOREM 3. *Let the conditions of Theorem* 2 *hold*. *Then*, *for* $c < 0$ *and as* $(m, M \to \infty)_{seq}$,

$$\sqrt{M}m(\hat{a}_{\omega}^{+} - a) \Rightarrow N(0, -2c). \tag{12}$$

The limiting variance formula $-2c$ in Theorem 3 has an interesting relationship to that of a stationary autoregression. In particular, the formula is identical to the limiting variance of the autoregressive coefficient $\hat{a}_m$ in a stationary autoregression with m observations, which is $(1-a^2) \sim 1-(1+(c/m))^2 \sim -2c/m$. This suggests the approximation $m(\hat{a}_m - a) \sim N(0,-2c)$, which corresponds to (12).

Observe that in Theorems 2 and 3, we still get the unit root/near integrated process result of consistent estimation of a by $\hat{a}$, $\hat{a}^+$, and $\hat{a}_\omega^+$ in spite of serial dependence (Phillips, 1987b), provided the second order bias terms are not too large and satisfy Assumption 3.

It follows from these asymptotics that

$$m(\hat{a}^+ - a) = m(\hat{a}^+ - 1) - c + O\left(\frac{1}{m}\right) \to_p 0$$

and therefore

$$\hat{c} = m(\hat{a}^+ - 1) \to_p c,$$

giving us an $O(\sqrt{M})$ consistent estimator of c. Of course, we have a corresponding estimator $\hat{c}_\omega = m(\hat{a}_\omega^+ - 1)$ in the case of the weighted regression estimator $\hat{a}_\omega^+$. In short, we have the following limit theory.

COROLLARY 4. *Let the conditions of Theorem* 2 *hold. Then, for* $c < 0$ *and as* $(m, M \to \infty)_{seq}$, $\hat{c}, \hat{c}_\omega \to_p c$ *and*

$$\sqrt{M}(\hat{c} - c) \Rightarrow N(0, V_a), \qquad \sqrt{M}(\hat{c}_\omega - c) \Rightarrow N(0,-2c).$$

The rate of convergence of $\hat{c}$ depends on the number of blocks M and is therefore determined by the number of separate blocks of information about the localizing parameter c. So, the success of this estimator relies on the homogeneity of the localizing parameter across blocks and the number of blocks in total. The form of the limit distribution of $\hat{c}_\omega$ makes inference about c particularly easy in the case where $c < 0$.

The estimator of the autoregressive coefficient a pools information within and across blocks and has a rate of convergence that depends on both m and M. The rate of convergence of $\hat{a}^+$ and $\hat{a}_\omega^+$ is $\sqrt{M}m$, and this rate is intermediate between the $O(\sqrt{n})$ rate of a stationary autoregression and the $O(n)$ rate of unit root regression. For example, we may functionalize m and M on the sample size n, as in $m = n^\gamma$, and $M = n^{1-\gamma}$, with $0 \le \gamma \le 1$. Then $\sqrt{M}m = n^\alpha$ with $\alpha = \frac{1}{2} + (\gamma/2)$, and the rate of convergence, n^α, of $\hat{a}^+$ then varies continuously from $\sqrt{n}$ to n. In effect, the block to unity system (3) is a family of models that constitute an intermediate class between stationary and unit root autoregressions.

When M is fixed, it is apparent from (9) that we have a class of nonnormal asymptotics, which reduce to the traditional case (Phillips, 1987a, 1987b) only when $M = 1$ and the initial conditions are in the near or distant past (then $H_{k,c}$ is replaced by $H_{k,c}^\kappa$ in (9) and $\kappa = 0$ or $\kappa > 0$ in (7)). When m is fixed, then the model has autoregressive parameter $a \sim 1 + (c/m) < 1$ and is stationary.

In the general case where $m \to \infty$ as $n \to \infty$, the autoregressive parameter $a \sim 1 + (c/m) \to 1$. However, because $a \sim 1 + (c/m) < 1 + (c/mM) = 1 + (c/n)$, the block autoregressive system with coefficient a and $M \to \infty$ is "closer" to stationarity when $c < 0$ than a conventional near integrated model with autoregressive coefficient $1 + (c/n)$ and the same localizing coefficient c. This explains why the asymptotic distributions of $\hat{c}$ and $\hat{a}^+$ are normal and why there is enough discriminatory information in data from the block autoregressive system to consistently estimate the localizing parameter c. In effect, the model across blocks has a stationary autoregressive structure with coefficient $e^c < 1$, as is apparent in the definition of $H_{k,c}(r)$ in (8).

However, when M is fixed, we have $m = O(n)$, and the autoregressive parameter $a \sim 1 + (c/m)$ is in the same locality of unity as the conventional local to unity model. In this event, $\hat{c}$ is not consistent, and the situation is analogous to that of the conventional local to unity model. Nonetheless, the preceding analysis allows for a wider class of limit theory in this case, as indicated in (10) earlier, where the number of blocks M plays a role in the limit and the limit process $H_{k,c}$ is a diffusion average, rather than a simple linear diffusion.

In light of these remarks, it would appear that there are substantial advantages in modeling to working with the general case where both m and $M \to \infty$. This is the situation that we will pursue in what follows and in our empirical application.

4.2. The Unit Root Case: $c = 0$ and $\omega_k^2 = \omega^2\ \forall k$

Let the initialization of the process be in the distant past, rather than the infinite past, and let Assumption 1(ii) hold. We will consider the homogeneous case where $b_{k,j} = b_j$. Homogeneity in the linear process coefficients across blocks ensures that $\omega_k^2 = \omega^2$, $\forall k$, so that the model is then comparable with a conventional unit root system that has a single long run variance parameter ω^2 and a single one sided long run covariance parameter λ.

From the analysis in Section 3, we have $m^{-1/2} y_{k,[mr]} \Rightarrow H_{k,c}^{\kappa}(r)$, as defined in (7). When $c = 0$, this limit process has the form

$$H_{k,0}^{\kappa}(r) = B_k(r) + \sum_{f=0}^{k-1} B_f(1) + B_0(-\kappa),$$

a linear combination of independent Brownian motions, all with variance ω^2. Our limit theory for $\hat{a}^+$ in this case is given in the following result.

THEOREM 5. *Let Assumptions* 1(ii) *and* 2 *hold. Then, in sequential limits as* $(m, M \to \infty)_{seq}$,

$$mM(\hat{a}^+ - a) \Rightarrow \left(\int_0^1 U(s)^2 ds \right)^{-1} \int_0^1 U(s)\, dU(s) := \xi_U,$$

where $U(s) \equiv BM(\omega^2)$.

Hence, in the $c = 0$ case we get $m(\hat{a}^+ - 1) \to_p 0$, as required for $\hat{c} = m(\hat{a}^+ - 1)$ to be a consistent estimator of $c = 0$. However, $M\hat{c} \Rightarrow \xi_U$, and therefore the estimate of c has a limit distribution in the unit root class in this case. Furthermore, we revert to an $O(n = mM)$ rate of convergence for $\hat{a}^+$ and move to an $O(M)$ rate of convergence for $\hat{c}$ when $c = 0$.

4.3. The Near Explosive Case: $c > 0$ and $\omega_k^2 = \omega^2\ \forall k$

In the case where $c > 0$, it turns out that $\hat{a} \to a$ at the rate $e^{cM}m$, comparable to the power rate of convergence in an explosive autoregression. Again, we work with distant past initial conditions at $y_{0,0}$ and homogeneity across blocks so that $\omega_k^2 = \omega^2$, $\forall k$. The latter helps us to relate our results to those already well known in the literature for explosive autoregressions. The limit theory for this case is as follows.

THEOREM 6. *Let Assumptions* 1(ii) *and* 2 *hold. Then, in sequential limits as* $(m, M \to \infty)_{seq}$,

$$\frac{e^{c(M+1)}m}{e^{2c}-1}(\hat{a}-a), \frac{e^{c(M+1)}m}{e^{2c}-1}(\hat{a}^+-a) \Rightarrow \frac{Z(c)}{Y(c)+J_{0,c}(-\kappa)}, \tag{13}$$

where $Z(c) \equiv N(0,(\omega^2/2c))$, $Y(c) \equiv N(0,(\omega^2/2c))$, *and* $Z(c), Y(c)$, *and* $J_{0,c}(-\kappa)$ *are independent.*

Remarks.

(1) It is apparent from (13) that the second order bias term that arises in traditional unit root regression disappears in the near explosive case. A similar result was obtained in Phillips (1987a, Theorem 2(c)) using the traditional local to unity model (3) and sequential limits involving the localizing coefficient $c \to \infty$. The reason is that the signal from the regressor is strong enough in the explosive case to eliminate the bias effects as $M \to \infty$.

(2) The limit variate (13) is a ratio of independent normals, each with zero mean, and is therefore proportional to a Cauchy variate. Note that the initial condition distribution $J_{0,c}(-\kappa)$ plays precisely the same role in the limit distribution here as it does in the well known explosive case (e.g., see Anderson, 1959, Theorem 2.5). However, unlike the conventional explosive model, the initial condition distribution in our case is always normal as it arises from a preliminary limiting process within the initial block.

(3) When the initial condition is at the origin and $\kappa = 0$, then $J_{0,c}(-\kappa) = 0$ and $e^{c(M+1)}m/(e^{2c}-1)(\hat{a}-1)$ has a limiting distribution that is standard Cauchy. This Cauchy limit (13) corresponds to the well known result from White (1958) and Anderson (1959, Theorem 2.7)[1] about the limiting distribution of the least squares regression coefficient in an explosive model with Gaussian errors and zero initialization. However, unlike these standard results, the limit result here does not rely on Gaussian errors. The difference is a major one and can be explained as follows. What happens in the block local model, in effect, is that as $m \to \infty$ within each block we get normality in the data from the first stage as-

ymptotics. The model across blocks then mirrors the structure in a Gaussian explosive autoregression. The outcome is that an invariance principle operates in the block local model in the explosive vicinity of the unit root case.

Theorem 6 implies that

$$\hat{c} = m(\hat{a}^{+} - 1) \to_p c,$$

giving us, in this case, an $O(e^{cM})$ consistent estimator of $c > 0$. In particular, we have the following result.

COROLLARY 7. *Let the conditions of Theorem* 6 *hold. Then, if* $\kappa = 0$ *and* $c > 0$, *and as* $(m, M \to \infty)_{seq}$, $\hat{c} \to_p c$ *and*

$$\frac{e^{c(M+1)}}{e^{2c} - 1}(\hat{c} - c) \Rightarrow \xi,$$

where ξ *is a standard Cauchy variable.*

5. ESTIMATION WITH TRENDING DATA

Our results in previous sections can be extended to more general models that allow for the presence of a deterministic trend in the original data. Such an extension is important because many macroeconomic time series, such as real GNP, consumption, money, and prices, are often characterized as integrated or near integrated processes with drifts. Our treatment here will deal with the case of a linear trend but it is easy to see how the approach applies for general polynomial trends. We also assume homogeneity in the linear process coefficients across blocks so that $\omega_k^2 = \omega^2$, $\forall k$. Again, this is easily generalized using the results of the previous section.

It is convenient to write the model in component form as follows:

$$y_{k,t} = d_{k,t} + y_{k,t}^{*}, \qquad t \in \mathbb{T}_m; k \in \mathbb{K}_K, \tag{14}$$

$$d_{k,t} = \gamma_0 + \gamma_1(km + t) = \gamma' x_{k,t}, \qquad x_{k,t} = (1, km + t)', \tag{15}$$

$$y_{k,t}^{*} = a y_{k,t-1}^{*} + u_{k,t}, \qquad y_{k,0}^{*} = y_{k-1,m}^{*}, \qquad a = e^{c/m} \sim 1 + \frac{c}{m}, \qquad c \le 0. \tag{16}$$

In (14) and (15), the deterministic component, $d_{k,t}$, contains both a linear time trend t and a block specific component km that assures the continuity of the trend across blocks. The stochastic part, $y_{k,t}^{*}$, in (14) corresponds to (3) in Section 3 and is a stochastic block local to unity process of the form

$$y_{k,t}^{*} = \sum_{j=0}^{t-1} e^{jc/m} u_{k,t-j} + e^{tc/m} \sum_{f=-K}^{k-1} e^{[k-1-f]c} \left[\sum_{j=0}^{m-1} e^{jc/m} u_{f,m-j} \right] + e^{tc/m} e^{[k-1+K]c} y_{-K,0}.$$

As in the simple case with no trend, the process $y^*_{k,t}$ has both a block index, k, and a within-block temporal index, t. However, by virtue of the sequence of block initializations $y^*_{k,0} = y^*_{k-1,m}$, the representation is consistent with a well defined evolution of a single time series sequence, this time with a linear drift. To simplify the analysis, it is sometimes useful to recognize this alternative representation by reindexing $y_{k,t}$ in the following way:

$$z_s = y_{[s/m],s-m[s/m]}, \qquad s = 1,2,\ldots,n = mM.$$

We use this single indexed representation and also the block representation in what follows. It can be easily verified that

$$z_s = \gamma_0 + \gamma_1 s + z^*_s = \gamma' x_s + z^*_s,$$

where $x_s = (1,s)'$ is a single indexed linear trend and $z^*_s = y^*_{[s/m],s-m[s/m]}$. For any k and t, $y_{k,t}$ corresponds to z_{km+t}.

Our purpose is to construct a consistent estimator of the local to unity parameter c in this model, and, to do so, appropriate detrending of $y_{k,t}$ is required. The most natural procedure, as in the traditional model (1) with trend, is to apply linear least squares detrending by means of the regression

$$y_{k,t} = \hat{\gamma}_0 + \hat{\gamma}_1(km+t) + \hat{y}^*_{k,t} = \hat{\gamma}' x_{k,t} + \hat{y}^*_{k,t}. \tag{17}$$

Here, the estimate of the trend coefficient is given by the following pooled regression formula:

$$\hat{\gamma} = \left[\sum_{k=1}^{M}\sum_{t=1}^{m} x_{k,t}x'_{k,t}\right]^{-1}\left[\sum_{k=1}^{M}\sum_{t=1}^{m} x_{k,t}y_{k,t}\right]$$
$$= \gamma + \left[\sum_{k=1}^{M}\sum_{t=1}^{m} x_{k,t}x'_{k,t}\right]^{-1}\left[\sum_{k=1}^{M}\sum_{t=1}^{m} x_{k,t}y^*_{k,t}\right].$$

The cases of primary interest are those where $c < 0$ and $c = 0$. As in the analysis of the model without trend, it is convenient to separate the analysis of these cases. We shall also consider the efficiency of this type of detrending by simple regression.

5.1. The Near Stationary Case: $c < 0$

To develop the limit theory, start by defining some scaling matrices for the deterministic trends. Let $D = \mathrm{diag}[1,n], F = \mathrm{diag}[1,m]$, and $G = \mathrm{diag}[1,M]$. Then, $D = FG$, and the deterministic components have the limits

$$D^{-1}x_{[nr]} \xrightarrow{n\to\infty} X(r) = (1,r)', \qquad F^{-1}x_{k,[mr]} \xrightarrow{m\to\infty} X_k(r) = (1,k+r)'.$$

The following theorem gives the limit theory for the least squares trend coefficient estimator $\hat{\gamma}$.

THEOREM 8. *Let Assumptions* 1(i) *and* 2 *hold and suppose* $c < 0$. *Then, as* $(m, M \to \infty)_{seq}$

$$\frac{\sqrt{n}}{m} D(\hat{\gamma} - \gamma) \Rightarrow \left(-\frac{1}{c}\right)\left[\int_0^1 X(r)X(r)'dr\right]^{-1}\left[\int_0^1 X(r)dU(r)\right], \tag{18}$$

where $U(r) \equiv BM(\omega^2)$.

The scaling matrix $\sqrt{n}m^{-1}D = \text{diag}[n^{1/2}m^{-1}, n^{1/2}M]$ in (18) indicates that consistent estimation of the intercept γ_0 and also the slope γ_1 in (15) is possible when $c < 0$ provided that $n^{1/2}m^{-1} \to \infty$ or, equivalently, $M/m \to \infty$. This is in contrast to the traditional local to unity model, where the intercept or any slowly evolving components in the deterministic trend are not consistently estimable. The reason why γ_0 can be consistently estimable in the block local to unity model can be explained as follows. From (14)–(16), the regression equation can be written in the form

$$\Delta_c y_{k,t} = \gamma' \Delta_c x_{k,t} + \Delta_c y^*_{k,t} = \gamma_0\left(-\frac{c}{m}\right) + \gamma_1\left(1 - ck - c\,\frac{(t-1)}{m}\right) + u_{k,t}, \tag{19}$$

where $\Delta_c = 1 - (1 + c/m)L$ is the quasi-differencing (QD) operator and L is the lag operator. The excitation condition for least squares regression consistency for the parameter γ_0 holds when

$$\sum_{k=1}^{M}\sum_{t=1}^{m}\left(-\frac{c}{m}\right)^2 = \frac{c^2 M}{m} \to \infty.$$

For this to hold, we must have $c \neq 0$ and $M/m \to \infty$. In effect, γ_0 is consistently estimable when the stationary element of the model (M blocks with autoregressive coefficient $e^c < 1$ for $c < 0$) dominates the nonstationary element (blocks of m observations with autoregressive coefficient $1 + (c/m)$) in the sense that $M/m \to \infty$.

The detrended time series is obtained from the residuals

$$\hat{y}^*_{k,t} = y_{k,t} - \hat{\gamma}' x_{k,t},$$

whose asymptotic behavior is shown in the following lemma to be the same as that of the stochastic component of the series, $y^*_{k,t}$.

LEMMA 9. *Under Assumptions* 1(i) *and* 2, *and when* $c < 0$,

$$m^{-1/2}\hat{y}^*_{k,[mr]} \Rightarrow H_{k,c}(r).$$

We now estimate the autoregressive coefficient in (16) by least squares regression on the detrended time series $\hat{y}^*_{k,t}$, giving

$$\tilde{a} = \frac{\sum_k \sum_t \hat{y}^*_{k,t-1}\hat{y}^*_{k,t}}{\sum_k \sum_t (\hat{y}^*_{k,t-1})^2},$$

and construct the modified estimator of a as in Section 4.1, i.e.,

$$\tilde{a}^{+} = \frac{\sum_{k=1}^{M}\left(\sum_{t}\hat{y}^{*}_{k,t-1}\hat{y}^{*}_{k,t} - m\hat{\lambda}_{k}\right)}{\sum_{k=1}^{M}\sum_{t}(\hat{y}^{*}_{k,t-1})^{2}}.$$

THEOREM 10. *Suppose* $c < 0$, *Assumptions* 2 *and* 3, *and the distant past initialization condition* 1(i) *hold. Then, in sequential limits as* $(m, M \to \infty)_{seq}$,

$$\sqrt{M}m(\tilde{a}^{+} - a) \Rightarrow N(0, -2c).$$

It therefore turns out that the estimation errors that arise from detrending are negligible in the limit and do not influence the asymptotic distribution of the coefficient estimator when $c < 0$. As a result, the limiting distribution of $\sqrt{M}m(\tilde{a}^{+} - a)$ is the same as that of $\sqrt{M}m(\hat{a}^{+} - a)$ in Theorem 2. This is entirely analogous to the situation of a stationary autoregression about a deterministic trend.

Furthermore, in the same way as before, we may construct the localizing parameter estimates

$$\hat{c} = m(\tilde{a}^{+} - 1) \to_{p} c,$$

giving us $O(\sqrt{M})$ consistent estimators of c. Corollary 4 continues to hold for $\hat{c}$.

5.2. The Unit Root Case: $c = 0$

When $c = 0$, we find that

$$\frac{1}{\sqrt{n}} D(\hat{\gamma} - \gamma) \Rightarrow \left[\int X(r)X(r)'\right]^{-1}\left[\int X(r)U(r)\right],$$

where $U(r) \equiv BM(\omega^{2})$. The detrended time series are constructed as

$$\hat{y}^{*}_{k,t} = y_{k,t} - \hat{\gamma}' x_{k,t},$$
$$\hat{z}^{*}_{s} = z_{s} - \hat{\gamma}' x_{s},$$

and, as is usual in unit root theory, the detrending process influences the asymptotic behavior of the filtered data. In particular, we have the following conventional result.

LEMMA 11. *For* $c = 0$ *and under Assumptions* 1(ii) *and* 2,

$$n^{-1/2}\hat{z}^{*}_{[nr]} \Rightarrow U(r) - \left[\int UX'\right]\left[\int XX'\right]^{-1} X(r) := \underline{U}(r).$$

Again, we estimate the autoregressive coefficient by least squares regression on the detrended time series $\hat{y}^*_{k,t}$, giving the pooled estimator

$$\tilde{a} = \frac{\sum_k \sum_t \hat{y}^*_{k,t-1} \hat{y}^*_{k,t}}{\sum_k \sum_t (\hat{y}^*_{k,t-1})^2},$$

and construct $\tilde{a}^+$ as before. Then we have the following asymptotics.

THEOREM 12. *When* $c = 0$ *and under Assumptions* 2, 3, *and* 1(ii),

$$n(\tilde{a}^+ - 1) \Rightarrow \frac{\int \underline{U} dU}{\int \underline{U}^2}. \tag{20}$$

An $O(M)$ consistent estimator of $c = 0$ can be obtained immediately from this result because $\hat{c} = m(\tilde{a}^+ - 1) \to_p 0$, and then

$$M\hat{c} \Rightarrow \frac{\int \underline{U} dU}{\int \underline{U}^2}. \tag{21}$$

Thus, when $c = 0$, we revert back to unit root asymptotics, and the distribution (20) is identical to that of the traditional model. In particular, $\tilde{a}^+$ converges to a at rate $O_p(n)$, and the limit distribution is a function of a detrended Brownian motion that depends on the limiting deterministic trend function just as in Phillips and Perron (1988) and Park and Phillips (1988). Moreover, because the limit distribution of $M\hat{c}$, (21), is identical to that of the Z_a unit test in the traditional model, it turns out that a significance test of the null hypothesis $c = 0$ against $c < 0$ that is based on the statistic $Z_a = M\hat{c}$ is identical to that of a conventional unit root test against a trend stationary alternative. As is apparent from Theorem 10, $Z_a = O_p(M)$ when $c < 0$, so our theory shows that the Z_a test is, in fact, consistent against local alternatives in the block local system with $c < 0$. Similar results can be shown to apply to other unit root tests.

6. EFFECT OF QUASI-DIFFERENCING IN TREND ELIMINATION

In the block local model (14)–(16), the residual process in the ordinary least squares regression (17) is near integrated, and it might appear at first blush that least squares estimation of the linear trend coefficient is not efficient, as is the case in the traditional local to unity model (Phillips and Lee, 1996). In the traditional model, an efficient estimator of the trend coefficients can be constructed by first quasi-differencing the regression equation. If we apply the same QD procedure to (14), we get, as in (19),

$$\Delta_c y_{k,t} = \gamma' \Delta_c x_{k,t} + u_{k,t}, \tag{22}$$

where $\Delta_c = 1 - (1 + c/m)L$ is the QD operator. Then, the trend coefficient can be fitted by regression on (22), giving

$$\begin{aligned}\tilde{\gamma} &= \left[\sum_{k=1}^{M}\sum_{t=1}^{m} \Delta_c x_{k,t} \Delta_c x_{k,t}'\right]^{-1}\left[\sum_{k=1}^{M}\sum_{t=1}^{m} \Delta_c x_{k,t} \Delta_c y_{k,t}\right] \\ &= \gamma + \left[\sum_{k=1}^{M}\sum_{t=1}^{m} \Delta_c x_{k,t} \Delta_c x_{k,t}'\right]^{-1}\left[\sum_{k=1}^{M}\sum_{t=1}^{m} \Delta_c x_{k,t} u_{k,t}\right].\end{aligned}$$

Analogous estimates of the trend coefficient in the traditional model were used by Elliot, Rothenberg, and Stock (1996) to construct modified unit root tests (with a prespecified value of the localizing parameter $\bar{c}$).

In practical work, the local parameter is not known, and so the QD operation in (22) is not feasible, thereby explaining the use of prespecified values such as $\bar{c}$. However, in block local models such as those considered here, c can be consistently estimated and used in a second stage QD detrending procedure. Thus, it might appear that there would be an advantage to QD detrending with an estimated operator. However, this turns out not to be the case.

Suppose that $c < 0$ and we estimate c by $\hat{c} = m(\tilde{a} - 1)$, as in Section 5. Then $\hat{c} = c + O_p(M^{-1/2})$. If we apply QD detrending with the operator $\Delta_{\hat{c}}$ to model (14)–(16), we get

$$\Delta_{\hat{c}} y_{k,t} = \gamma' \Delta_{\hat{c}} x_{k,t} + \Delta_{\hat{c}} y_{k,t}^*. \tag{23}$$

The ordinary least squares (OLS) estimator of γ from (23) is

$$\begin{aligned}\tilde{\gamma}_f &= \left[\sum_{k=1}^{M}\sum_{t=1}^{m} \Delta_{\hat{c}} x_{k,t} \Delta_{\hat{c}} x_{k,t}'\right]^{-1}\left[\sum_{k=1}^{M}\sum_{t=1}^{m} \Delta_{\hat{c}} x_{k,t} \Delta_{\hat{c}} y_{k,t}\right] \\ &= \gamma + \left[\sum_{k=1}^{M}\sum_{t=1}^{m} \Delta_{\hat{c}} x_{k,t} \Delta_{\hat{c}} x_{k,t}'\right]^{-1}\left[\sum_{k=1}^{M}\sum_{t=1}^{m} \Delta_{\hat{c}} x_{k,t} \Delta_{\hat{c}} y_{k,t}^*\right].\end{aligned}$$

The limiting distributions of $\tilde{\gamma}$ and $\tilde{\gamma}_f$ are given in the following theorem.

THEOREM 13. *Suppose* $c < 0$, *Assumption* 2, *and the distant past initialization condition* 1(i) *hold. Then, in sequential limits as* $(m, M \to \infty)_{seq}$,

$$\frac{\sqrt{n}}{m} D(\tilde{\gamma} - \gamma),\ \frac{\sqrt{n}}{m} D(\tilde{\gamma}_f - \gamma) \Rightarrow -\frac{1}{c}\left[\int X(r)X(r)'\right]^{-1}\int X(r)\,dU(r), \tag{24}$$

where $U(r) \equiv BM(\omega^2)$.

It follows from Theorem 13 that the errors arising from preliminary estimation of c are asymptotically negligible in the estimation of the trend coefficients, and the limiting distribution of the feasible trend coefficient vector $\tilde{\gamma}_f$ is the same as that of $\tilde{\gamma}$, the infeasible estimator that uses the true local parameter. Moreover, both these estimates are asymptotically equivalent to the least squares trend estimator $\hat{\gamma}$ that uses no information about the localizing parameter c. Hence, the simple trend estimator $\hat{\gamma}$ is efficient in the sense that it is asymptotically equivalent to the generalized least squares (GLS) estimator, were we to know c. Thus, in the block local to unity model there is no need to apply QD procedures in fitting the trend coefficient, at least asymptotically. The explanation for this phenomenon is that when $M \to \infty$, the deterministic trend becomes a dominating characteristic across blocks (because of the continuity of the trend) and when $c < 0$ the behavior of the model across blocks is, as we have seen, essentially stationary. This produces a stochastic environment that validates the Grenander and Rosenblatt (1957) theory of efficient trend elimination by least squares regression.

7. CONCLUSIONS

This paper introduces a new statistical model to capture the notion of near integration. It has the advantage over the traditional model developed in earlier work (Phillips, 1987a; Chan and Wei, 1987) that the local parameter can be consistently estimated. The model also provides a more complete interface between $I(0)$ and $I(1)$ models and between $O(\sqrt{n})$ and $O(n)$ asymptotics. In fact, the rate of convergence to the autoregressive coefficient in the new model is $O(n^{\alpha})$ for $\alpha \in [\frac{1}{2},1]$ and varies in a continuous way between that of stationary and nonstationary asymptotics. The model also captures the power law asymptotics of explosive autoregressions and shows that, in a well defined local region greater than unity, it is possible to obtain invariance principles, in contrast to standard results for the explosive autoregression.

Some additional features of the model stand out. First, semiparametric estimation of the autoregressive parameter is possible using the methods of earlier work on unit root estimation, giving a robust estimator in models that are closer to stationarity than unit root models and traditional local to unity models. In other words, specification of the short memory component of the model is not necessary for consistent estimation, in contrast to stationary autoregression, where short memory error serial dependence induces inconsistency. Second, conventional unit root tests are seen to be consistent against alternatives that are local to unity in the new sense. Third, least squares regression estimates of deterministic trend components are asymptotically efficient, and it is not necessary to quasi-difference the data or to use GLS techniques to improve efficiency in trend elimination procedures.

Implementation of the procedures given here requires the selection of the index parameters m and M. A serious study of this matter is likely to be complex, as a result of the interactive role of the localizing parameter c and the block size m. Ideally, we would like to obtain data based rules, and, as in kernel estimation, this will require the use of a suitable criterion function and some more refined asymptotics than we have presented here. A further matter of interest is the extension of the present model to allow for heterogeneous deterministic trends across blocks. The blocking mechanism in the present model provides a natural structure for introducing such breaking trend functions. Of course, allowance for endogenously determined breaks would require the further extension of variable block sizes. Moreover, because the model allows for the number of blocks to pass to infinity, this extension effectively introduces an infinite number of nuisance parameters as $M \to \infty$. Although these and other interesting considerations extend beyond the limitations of this initial study, they serve to give an idea of the potential of block nonstationary systems in modeling time series economic data.

NOTE

1. The normalization factor in Theorem 2.7 of Anderson (1959) is $a^T/(a^2 - 1)$, corresponding to a sample of size T. The normalization in (13) is $e^{c(M+1)}m/(e^{2c} - 1)$, which corresponds to $a^{M+1}m/(a^2 - 1)$. The reason for the exponent $M + 1$, rather than simply M, is that we have M blocks in the data but $M + 1$ blocks in the process from the initialization at $y_{0,0}$. In an explosive model, a change in the initial conditions does affect the limit theory, and it figures here in the normalization factor.

REFERENCES

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. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Basawa, I.V. & P.J. Brockwell (1984) Asymptotic conditional inference for regular nonergodic models with an application to autoregressive processes. *Annals of Statistics* 12, 161–171.

Basawa, I.V., A.K. Lallik, W.P. McCormick, J.H. Reeves, & R.L. Taylor (1991) Bootstrapping unstable first order autoregressive process. *Annals of Statistics* 19, 1098–1101.

Bobkoski, M.J. (1983) *Hypothesis Testing in Nonstationary Time Series*. Ph.D. Thesis, University of Wisconsin, Department of Statistics.

Cavanagh, C. (1985) Roots local to unity. Manuscript, Harvard University, Department of Economics.

Chan, N.H. & C.Z. Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.

Chan, N.H. & C.Z. Wei (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367–401.

Elliot, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests of an autoregressive unit root. *Econometrica* 64, 813–836.

Grenander, U. & M. Rosenblatt (1957) *Statistical Analysis of Stationary Time Series*. New York: Wiley.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. New York: Academic Press.

Hannan, E.J. (1970) *Multiple Time Series*. New York: Wiley.

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.

Moon H.R. & P.C.B. Phillips (2000) Estimation of autoregressive roots near unity using panel data. *Econometric Theory* 16, 927–997.

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

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

Phillips, P.C.B. (1987b) Time series regression with a unit root. *Econometrica* 55, 277–301.

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

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

Phillips, P.C.B. & H.R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.

Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.

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

Stock, J.H. (1991) Confidence intervals for the largest autoregressive root in US macroeconomic time series. *Journal of Monetary Economics* 28, 435–460.

van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

White, J.S. (1958) The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* 29, 1188–1197.

TECHNICAL APPENDIX AND PROOFS

BN *decomposition of* $u_{k,t}$. Following Phillips and Solo (1992) we decompose $u_{k.t}$ as

$$u_{k,t} = b_k(1)\varepsilon_{k,t} + \tilde{\varepsilon}_{k,t-1} - \tilde{\varepsilon}_{k,t},$$

where $\tilde{\varepsilon}_{k,t} = \sum_{j=0}^{\infty} \tilde{b}_{k,j}\varepsilon_{k,t-j}$, $\tilde{b}_j = \sum_{l=j+1}^{\infty} b_{k,l}$, and $b_k(1) = \sum_{j=0}^{\infty} b_{k,j}$. Under the summability condition in Assumption 2, it is apparent that there exist finite constants M_1 and M_2 such that

$$E\tilde{\varepsilon}_{k,t}^2 \le M_1 \tag{A.1}$$

and

$$Eu_{k,t}^2 \le M_2$$

uniformly in k and t (see Moon and Phillips, 1998).

Proof of Convergence in Mean Square of (5). Let $X_f = m^{-1/2}\sum_{j=0}^{m-1} e^{jc/m} \times u_{k-1-f,m-j}$. Write $x_k = m^{-1/2}\sum_{j=0}^{m-1} e^{jc/m}\varepsilon_{k,m-j}$ and $R_k = m^{-1/2}(e^{[m-1/m]c}\tilde{\varepsilon}_{k,0} - \tilde{\varepsilon}_{k,m}) + m^{-1/2}\sum_{j=0}^{m-1} e^{[(j-1)c]/m}\tilde{\varepsilon}_{k,m-j}(1 - e^{c/m})$. Then, using the BN decomposition of $u_{k,t}$, under Assumption 2, there exists a constant M such that

$$
\begin{aligned}
EX_f^2 &= E\left(\frac{1}{\sqrt{m}}\sum_{j=0}^{m-1} e^{jc/m} u_{k-1-f,m-j}\right)^2 \\
&= E(b_{k-1-f}(1)x_{k-1-f} + R_{k-1-f})^2 \le M, \qquad \textbf{(A.2)}
\end{aligned}
$$

where the last inequality is proved in Moon and Phillips (2000) and holds uniformly in f. To finish the proof, we need only show that

$$
\lim_{n\to\infty} E\left(\sum_{f=n}^{\infty} e^{fc} X_f\right)^2 = 0,
$$

which holds because

$$
\begin{aligned}
E\left(\sum_{f=n}^{\infty} e^{fc} X_f\right)^2 &= E\left(\sum_{f=n}^{\infty} e^{(1/2)fc} e^{(1/2)fc} X_f\right)^2 \\
&\le \left(\sum_{f=n}^{\infty} e^{fc}\right) E\left(\sum_{f=n}^{\infty} e^{fc} X_f^2\right) \le M\left(\sum_{f=n}^{\infty} e^{fc}\right)^2 \\
&\to 0 \qquad \text{as } n \to \infty,
\end{aligned}
$$

where the first inequality holds by the Cauchy–Schwarz inequality and the last convergence holds because $c < 0$ and $\sum_{f=1}^{\infty} e^{fc} < \infty$. ■

Proof of Lemma 1. From the BN decompositions of $u_{k,t}$, we have

$$
\begin{aligned}
\frac{1}{\sqrt{m}}\sum_{j=1}^{[mr]} \begin{pmatrix} u_{0,j} \\ \vdots \\ u_{M,j} \end{pmatrix} &= \frac{1}{\sqrt{m}}\sum_{j=1}^{[mr]} \begin{pmatrix} b_0(1)\varepsilon_{0,j} \\ \vdots \\ b_M(1)\varepsilon_{M,j} \end{pmatrix} + \frac{1}{\sqrt{m}} \begin{pmatrix} \tilde{\varepsilon}_{0,0} - \tilde{\varepsilon}_{0,[mr]} \\ \vdots \\ \tilde{\varepsilon}_{M,0} - \tilde{\varepsilon}_{M,[mr]} \end{pmatrix} \\
&= \Omega_M^{1/2} \frac{1}{\sqrt{m}}\sum_{j=1}^{[mr]} \begin{pmatrix} \varepsilon_{0,j} \\ \vdots \\ \varepsilon_{M,j} \end{pmatrix} + o_p(1),
\end{aligned}
$$

where the second line holds by the same argument as that in Phillips and Solo (1992, p. 978). Because

$$
\frac{1}{\sqrt{m}}\sum_{j=1}^{[mr]} \begin{pmatrix} \varepsilon_{0,j} \\ \vdots \\ \varepsilon_{M,j} \end{pmatrix} \Rightarrow W^M(r) \equiv \begin{pmatrix} W_0(r) \\ \vdots \\ W_M(r) \end{pmatrix},
$$

a $(M+1)$ vector standard Brownian motion, we have

$$
\frac{1}{\sqrt{m}}\sum_{j=1}^{[mr]} \begin{pmatrix} u_{0,j} \\ \vdots \\ u_{M,j} \end{pmatrix} \Rightarrow \begin{pmatrix} B_0(r) \\ \vdots \\ B_M(r) \end{pmatrix} \equiv BM(\Omega_M)
$$

as required. ■

Proof of (8). We start by introducing some notation and definitions. Suppose that the i.i.d. sequence $\{\varepsilon_t\}_t$ in Assumption $(2)(i)$ is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define

$$X_{m,k} = \frac{1}{\sqrt{m}} \sum_{j=1}^{m-1} e^{jc/m} u_{k,m-j}, \qquad Y_{m,k} = e^{ck/2} X_{m,k},$$

$$X_k = J_{k,c}(1) \quad \text{and} \quad Y_k = e^{ck/2} X_k,$$

$$Y_m = (Y_{m,0}, \ldots, Y_{m,k}, \ldots)' \quad \text{and} \quad Y = (Y_0, \ldots, Y_k, \ldots)'.$$

The terms Y_m and Y are $\mathbb{R}^\infty (= \times_1^\infty \mathbb{R})$-valued random elements. We use the following distance metric between two elements of $\mathbb{R}^\infty$:

$$d(x,y) = \sup_{k \geq 0} |x_k - y_k|, \tag{A.3}$$

where $x = (x_1, x_2, \ldots)$, $y = (y_1, y_2, \ldots) \in \mathbb{R}^\infty$. Let $\mathbb{N}$ be the set of non-negative integers, $\{0,1,2,\ldots\}$. The space $l^\infty(\mathbb{N})$ is defined as the set of all uniformly bounded, real functions on $\mathbb{N}$, i.e., all functions $x : \mathbb{N} \to \mathbb{R}$ such that $d(x,0) < \infty$.

By virtue of (A.2)

$$E\left(\sup_{k\geq 0} |Y_{m,k}|\right) \leq \sum_{k=0}^{\infty} E|Y_{m,k}| = \sum_{k=0}^{\infty} e^{ck/2} E|X_{m,k}|$$
$$\leq \sum_{k=0}^{\infty} e^{ck/2} \max\{1, EX_{m,k}^2\} \leq \sum_{k=0}^{\infty} e^{ck/2} \max\{1, M\} < \infty,$$

and it follows that Y_m is a sequence of $l^\infty(\mathbb{N})$-valued random elements with probability one. Similarly, it is easy to verify that Y is also an $l^\infty(\mathbb{N})$-valued random element with probability one. Thus, we may restrict attention to the case where Y_m takes values in $l^\infty(\mathbb{N})$.

For weak convergence in $l^\infty(\mathbb{N})$, we need only establish the following two conditions: (i) finite dimensional convergence and (ii) asymptotic tightness. In fact, according to Theorem 1.5.4 of van der Vaart and Wellner (1996), Y_m converges weakly to Y if (i)

$$\begin{pmatrix} Y_{m,k_1} \\ \vdots \\ Y_{m,k_n} \end{pmatrix} \Rightarrow \begin{pmatrix} Y_{k_1} \\ \vdots \\ Y_{k_n} \end{pmatrix} \tag{A.4}$$

as $m \to \infty$ for an arbitrary subset $\{k_1, \ldots, k_n\}$ of $\mathbb{N}$ and (ii) Y_m is asymptotically tight. We already know that the finite dimensional convergence (A.4) holds by Lemma 1. For asymptotic tightness of Y_m, we appeal to part of Theorem 1.5.6 of van der Vaart and Wellner (1996): specifically, the sequence $Y_m : \Omega \to l^\infty(\mathbb{N})$ is asymptotically tight if $Y_{m,k}$ is asymptotically tight in $\mathbb{R}$ for every k and, for all $\varepsilon, \eta > 0$, there exists a finite partition $\mathbb{N} = \cup_{i=1}^I \mathbb{N}_i$ such that

$$\limsup_m \mathbb{P}\left\{\sup_i \sup_{s,t \in \mathbb{N}_i} |Y_{m,s} - Y_{m,t}| > \varepsilon\right\} < \eta. \tag{A.5}$$

(In van der Vaart and Wellner, 1996, to define (A.5), an outer probability measure of $\mathbb{P}$ is used. However, because the index set $\mathbb{N}$ of $l^\infty(\mathbb{N})$ is a countable set and there is no measurability problem on the sup operator on the set $\mathbb{N}$, we use the underlying probability measure $\mathbb{P}$ in defining (A.5).)

Because the individual sequence of random variables $Y_{m,k}$ converges in distribution to $Y_k = e^{kc/2}J_{k,c}(1)$ for all k, $Y_{m,k}$ is asymptotically tight in $\mathbb{R}$ for every k. Next, condition (A.5) is satisfied if we show that for all $\varepsilon, \eta > 0$, there exists a constant k_0 such that

$$\limsup_m \mathbb{P}\left\{ \sup_{s,t\geq k_0} |Y_{m,s} - Y_{m,t}| > \varepsilon \right\} < \eta. \tag{A.6}$$

For, if (A.6) holds, we can choose $\mathbb{N}_i = \{i\}$ for $i < k_0$ and $\mathbb{N}_{k_0} = \{t : t \geq k_0\}$. Then, $\mathbb{N} = \cup_{i=1}^{k_0}\mathbb{N}_i$ is a finite partition and (A.5) is satisfied. Note that

$$\begin{aligned}
&\mathbb{P}\left\{ \sup_{s,t\geq k_0} |Y_{m,s} - Y_{m,t}| > \varepsilon \right\} \\
&\quad \leq \mathbb{P}\left\{ \sum_{k=k_0}^{\infty} |Y_{m,k+1} - Y_{m,k}| > \varepsilon \right\} \\
&\quad \leq \frac{1}{\varepsilon^2} E\left(\sum_{k=k_0}^{\infty} |Y_{m,k+1} - Y_{m,k}| \right)^2 = \frac{1}{\varepsilon^2} E\left(\sum_{k=k_0}^{\infty} e^{ck/2} |e^{c/2}X_{m,k+1} - X_{m,k}| \right)^2 \\
&\quad \leq \frac{1}{\varepsilon^2} \left(\sum_{k=k_0}^{\infty} e^{ck/2} \right) \left(\sum_{k=k_0}^{\infty} e^{ck/2} E(e^{c/2}X_{m,k+1} - X_{m,k})^2 \right),
\end{aligned} \tag{A.7}$$

where the last inequality holds by the Cauchy–Schwarz inequality. In view of (A.2), by choosing k_0 large enough, the right hand side of (A.7) can be made less than η. Thus, (A.5) is satisfied, and we have

$$Y_m \Rightarrow Y$$

as $m \to \infty$.

Next consider the functional $\nu : l^\infty(\mathbb{N}) \to \mathbb{R}$ defined by

$$\nu(x) = \sum_{k=0}^{\infty} e^{kc/2} x_k.$$

Then, it is easy to see that $\nu(x)$ is continuous with respect to d in (A.3), and by the continuous mapping theorem, we have the required result. ■

Proof of Theorem 2. For fixed M, as $m \to \infty$, under the assumption that the initial conditions are in the infinite past, we have as in (9)

$$\sqrt{M}m(\hat{a}^+ - a) = \frac{\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \left[\int_0^1 H_{k,c}\, dB_k + (\lambda_k - \hat{\lambda}_k) \right]}{\frac{1}{M} \sum_{k=1}^{M} \int_0^1 H_{k,c}(r)^2 dr} + o_p(1). \tag{A.8}$$

First, under Assumption 3 we have $\sqrt{mh}(\hat{\lambda}_k - \lambda_k) \to_d \xi_k \equiv N(0, V_k)$ as $m \to \infty$, $\forall k$, and $\{\xi_k\}$ is an independent sequence of normal variates with zero mean and variance that are bounded uniformly in k, because $\sup_k V_k < \infty$. It follows that as $m \to \infty$,

$$\frac{1}{\sqrt{M}} \sum_{k=1}^{M} (\lambda_k - \hat{\lambda}_k) \sim \frac{1}{\sqrt{M}} \sum_{k=1}^{M} \frac{1}{\sqrt{mh}} \xi_k = o_p(1), \tag{A.9}$$

as $(m, M \to \infty)_{seq}$.

Next, we apply a strong law to $(1/M)\sum_{k=1}^{M}\int_0^1 H_{k,c}(r)^2 dr$ and a CLT to $(1/\sqrt{M}) \times \sum_{k=1}^{M}[\int_0^1 H_{k,c}\, dB_k]$ as $M \to \infty$. To find the limit of $(1/M)\sum_{k=1}^{M}\int_0^1 H_{k,c}(r)^2 dr$, we write

$$\begin{aligned}
\frac{1}{M}\sum_{k=1}^{M}\int_0^1 H_{k,c}(r)^2 dr
&= \frac{1}{M}\sum_{k=1}^{M}\int_0^1 J_{k,c}^2 dr + \frac{2}{M}\sum_{k=1}^{M}\int_0^1 J_{k,c} e^{rc} dr \left(\sum_{f=-\infty}^{k-1} e^{(k-1-f)c} J_{f,c}(1)\right) \\
&\quad + \int_0^1 e^{2rc} dr \, \frac{1}{M}\sum_{k=1}^{M}\left(\sum_{f=-\infty}^{k-1} e^{(k-1-f)c} J_{f,c}(1)\right)^2 \\
&= I + II + III, \quad \text{say.}
\end{aligned}$$

Now $\{J_{k,c}\}$ is a sequence of independent normal variates, and because $\sup_k \omega_k^2 < \infty$ we have $\sup_k E(\int_0^1 J_{k,c}^2 dr)^{1+\delta} < \infty$. It follows from the Markov strong law for independent and nonidentically distributed (i.n.i.d.) sequences that

$$\begin{aligned}
I &= \frac{1}{M}\sum_{k=1}^{M}\int_0^1 J_{k,c}^2 dr \\
&\to_{a.s.} \left(\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M}\omega_k^2\right) E\left(\int_0^1 S_c^2\right),
\end{aligned} \tag{A.10}$$

where $S_c(r) = \int_0^r e^{(r-s)c} dW(s)$. Next consider term II. From the independence of $\{J_{k,c}\}_k$ and because

$$\begin{aligned}
&\sup_k E\left(\int_0^1 J_{k,c} e^{rc} dr \left(\sum_{f=-\infty}^{k-1} e^{(k-1-f)c} J_{f,c}(1)\right)\right)^2 \\
&\quad \le \sup_k E\left(\int_0^1 J_{k,c} e^{rc} dr\right)^2 \sup_k E\left(\sum_{f=-\infty}^{k-1} e^{(k-1-f)c} J_{f,c}(1)\right)^2 < \infty,
\end{aligned}$$

it follows that $\{\int_0^1 J_{k,c} e^{rc} dr (\sum_{f=-\infty}^{k-1} e^{(k-1-f)c} J_{f,c}(1))\}_k$ is a sequence of square integrable martingale differences with respect to the natural filtration, and by the strong law for martingale differences (e.g., Hall and Heyde, 1980, p. 36)

$$II \to_{a.s.} 0. \tag{A.11}$$

Before considering term III, under Assumption 2, we define

$$\bar{Z}_c = \sup_k |\omega_k||S_c(1)|,$$

where $S_c(1) = \int_0^1 e^{(1-s)c} dW(s)$. Then, $E(\bar{Z}_c)^4 < \infty$, and $\bar{Z}_c$ is a dominating random variable for the martingale difference sequence $\{J_{k,c}(1)\}$ in the sense that

$$P\{|J_{k,c}(1)| > x\} < P\{|\bar{Z}_c| > x\}.$$

Also, we have

$$\frac{1}{M}\sum_{k=1}^{M} EJ_{k,c}(1)^2 \to \left(\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M}\omega_k^2\right) ES_c(1)^2$$

and

$$\sum_{j=0}^{\infty} je^{2jc} < \infty.$$

Then, by Theorem 3.16 of Phillips and Solo (1992), we have

$$\begin{aligned} III &= \int_0^1 e^{2rc}dr\,\frac{1}{M}\sum_{k=1}^{M}\left(\sum_{j=0}^{\infty} e^{jc}J_{k-1-j,c}(1)\right)^2 \\ &\to_{a.s.} \int_0^1 e^{2rc}dr\left(\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M}\omega_k^2\right)\left(\frac{1}{1-e^{2c}}\right) ES_c(1)^2 \\ &= \left(-\frac{1}{2c}\right)\left(\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M}\omega_k^2\right) ES_c(1)^2. \end{aligned} \quad \textbf{(A.12)}$$

Combining (A.10)–(A.12), we have

$$\begin{aligned} &\frac{1}{M}\sum_{k=1}^{M}\int_0^1 H_{k,c}(r)^2 dr \\ &\quad \to_{a.s.} \left(\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M}\omega_k^2\right)\left(E\left(\int_0^1 S_c^2\right) - \frac{1}{2c} ES_c(1)^2\right) = V_H. \end{aligned} \quad \textbf{(A.13)}$$

To reduce V_H note that

$$\begin{aligned} &\left[E\left(\int_0^1 S_c^2\right) - \frac{1}{2c} ES_c(1)^2\right] \\ &\quad = \int_0^1 \frac{e^{2rc}-1}{2c}dr - \frac{1}{2c}\,\frac{e^{2c}-1}{2c} \\ &\quad = \left\{\frac{1}{2c}\left[\frac{1}{2c}(e^{2c}-1) - 1\right] - \frac{1}{2c}\,\frac{e^{2c}-1}{2c}\right\} \\ &\quad = -\frac{1}{2c}, \end{aligned} \quad \textbf{(A.14)}$$

so that $V_H = \mu_2/-2c$, where $\mu_2 = \lim_{M\to\infty}(1/M)\sum_{k=1}^{M}\omega_k^2$.

We now derive the limit distribution of $(1/\sqrt{M})\sum_{k=1}^{M}\int_0^1 H_{k,c}\,dB_k$. Let $Z_{k,c} = \int_0^1 H_{k,c}\,dB_k$. In view of the independence of B_k across k and the fact that

$$\begin{aligned}
\sup_k EZ_{k,c}^2 &= \sup_k \omega_k^2\left[E\int_0^1 H_{k,c}^2\right] \\
&= \sup_k\left[\omega_k^4 E\int_0^1 S_c^2 + \omega_k^2\left(\sum_{f=-\infty}^{k-1} e^{2[k-1-f]c}\omega_f^2 ES_c(1)^2\right)\int_0^1 e^{2rc}dr\right] \\
&\le \left(\sup_k \omega_k^4\right)\left[E\int_0^1 S_c^2 - \frac{1}{2c}ES_c(1)^2\right],
\end{aligned}$$

we know that $Z_{k,c}$ is a sequence of martingale differences with respect to the natural filtration, and we may therefore employ a CLT for martingale differences. Let $\Sigma_M^2 = \sum_{k=1}^{M} EZ_{k,c}^2$. Define

$$\tilde{Z}_{k,c} = \left(\sup_k \omega_k^2\right)\left(\left|\int_0^1 S_k\,dW_k\right| + \left|\int_0^1 e^{rc}dW_k\right|\left(\sum_{f=-\infty}^{k-1} e^{(k-1-f)c}|S_{f,c}(1)|\right)\right).$$

It is easy to verify that $\tilde{Z}_{k,c}$ is strictly stationary, $\tilde{Z}_{k,c} \ge |Z_{k,c}|\ \forall k$, and

$$\begin{aligned}
&E(\tilde{Z}_{k,c}^2) \\
&\quad\le 2\left(\sup_k \omega_k^2\right)^2 E\left(\int_0^1 S_k\,dW_k\right)^2 \\
&\qquad + 2\left(\sup_k \omega_k^2\right)^2 E\left(\int_0^1 e^{rc}dW_k\right)^2 \\
&\qquad \times\left\{E(S_{f,c}(1)^2)\left(\frac{1}{1-e^{2c}}\right) + \left(\frac{E|S_{f,c}(1)|}{1-e^c}\right)^2\right\} \\
&\quad\le C,
\end{aligned}$$

for some constant C uniformly in k. Then, for any $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{k=1}^{M} E\left[\frac{Z_{k,c}^2}{\Sigma_M^2}1\left\{\left|\frac{Z_{k,c}^2}{\Sigma_M^2}\right| > \varepsilon\right\}\right] \\
&\quad\le \sum_{k=1}^{M} E\left[\frac{\tilde{Z}_{k,c}^2}{\Sigma_M^2}1\left\{\left|\frac{\tilde{Z}_{k,c}^2}{\Sigma_M^2}\right| > \varepsilon\right\}\right] \\
&\quad= \frac{M}{\Sigma_M^2}E\left[\tilde{Z}_{1,c}^2 1\left\{\tilde{Z}_{1,c}^2 > \frac{\Sigma_M^2}{M}M\varepsilon\right\}\right] \\
&\quad\le \frac{1}{\inf_k E(Z_{k,c}^2)}E[\tilde{Z}_{1,c}^2 1\{\tilde{Z}_{1,c}^2 > (\inf_k E(Z_{k,c}^2))M\varepsilon\}] \\
&\quad\to 0 \quad \text{as } M \to \infty,
\end{aligned}$$

where the second line holds by $\tilde{Z}_{k,c} \geq |Z_{k,c}|$, the third line holds by the strict stationarity of $\tilde{Z}_{k,c}^2$, the fourth line is well defined because $\inf_k E(Z_{k,c}^2) > 0$ in view of the fact that $\inf_k \omega_k^2 > 0$, and the last line holds by virtue of the fact that $E(\tilde{Z}_{k,c}^2) < C$. Therefore, as $M \to \infty$

$$\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \int_0^1 H_{k,c} \, dB_k \Rightarrow N\left(0, \lim_{M\to\infty} \frac{1}{M} \sum_{k=1}^{M} \omega_k^2 E\left[\int_0^1 H_{k,c}^2\right]\right). \tag{A.15}$$

Combining (A.9), (A.13), and (A.15), we have the required limit distribution.

To simplify the variance formula, observe that as in (A.13) and (A.14) we have

$$\begin{aligned}
&\frac{1}{M} \sum_{k=1}^{M} \omega_k^2 E\left[\int_0^1 H_{k,c}^2\right] \\
&\quad \to \left(\lim_{M\to\infty} \frac{1}{M} \sum_{k=1}^{M} \omega_k^4\right)\left(E\left(\int_0^1 S_c^2\right) - \frac{1}{2c} ES_c(1)^2\right) \\
&\quad = -\frac{1}{2c} \mu_4.
\end{aligned}$$

■

Proof of Theorem 3. The proof follows the same lines as that of Theorem 2 given earlier, and so we simply outline the argument. For fixed M, as $m \to \infty$, we have as in (A.8)

$$\sqrt{M} m(\hat{a}_\omega^+ - a) = \frac{\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \frac{1}{\omega_k^2}\left[\int_0^1 H_{k,c} \, dB_k + (\lambda_k - \hat{\lambda}_k)\right]}{\frac{1}{M} \sum_{k=1}^{M} \frac{1}{\omega_k^2} \int_0^1 H_{k,c}(r)^2 dr} + o_p(1).$$

Then

$$\frac{1}{M} \sum_{k=1}^{M} \frac{1}{\omega_k^2} \int_0^1 H_{k,c}(r)^2 dr \to_{a.s.} \left(E\left(\int_0^1 S_c^2\right) - \frac{1}{2c} ES_c(1)^2\right) = -\frac{1}{2c},$$

as in (A.14). Further, in the same way as (A.15) we find

$$\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \frac{1}{\omega_k^2} \int_0^1 H_{k,c} \, dB_k \Rightarrow N\left(0, \lim_{M\to\infty} \frac{1}{M} \sum_{k=1}^{M} \frac{1}{\omega_k^2} E\left[\int_0^1 H_{k,c}^2\right]\right) = N\left(0, -\frac{1}{2c}\right).$$

It follows that in sequential limits as $(m, M \to \infty)_{seq}$

$$\sqrt{M} m(\hat{a}^+ - a) \Rightarrow N(0, -2c),$$

giving the stated result.

Proof of Theorem 5. If Assumptions 1(ii) and 2 hold, then, as in (9) but using the fact that $c = 0$, it follows that as $m \to \infty$ for fixed M,

$$Mm(\hat{a}^{+} - a) = \frac{\frac{1}{M}\sum_{k=1}^{M}\left(\frac{1}{m}y_{-1}^{k\prime}u^{k} - m\hat{\lambda}_{k}\right)}{\frac{1}{M^{2}}\sum_{k=1}^{M}\frac{1}{m^{2}}y_{-1}^{k\prime}y_{-1}^{k}}$$

$$\Rightarrow \frac{\frac{1}{M}\sum_{k=1}^{M}\left[\int_{0}^{1}H_{k,0}^{\kappa}dB_{k} + (\lambda_{k} - \hat{\lambda}_{k})\right]}{\frac{1}{M^{2}}\sum_{k=1}^{M}\int_{0}^{1}H_{k,0}^{\kappa}(r)^{2}dr},$$

where $H_{k,0}^{\kappa}(r) = B_{k}(r) + \sum_{f=0}^{k-1}B_{f}(1) + B_{0}(-\kappa)$. As in the proof of Theorem 2, we have

$$\frac{1}{M}\sum_{k=1}^{M}(\lambda_{k} - \hat{\lambda}_{k}) = o_{p}(1) \tag{A.16}$$

in sequential limit as $(m, M \to \infty)_{seq}$.

Before proceeding further with the proof, note that the sequence $\{B_{k}(1)\}_{k}$ is i.i.d. $N(0,\omega^{2})$. Then, by Donsker's functional law for partial sums of i.i.d. random variables we have, as $M \to \infty$,

$$\frac{1}{\sqrt{M}}\sum_{f=0}^{[Mr]}B_{f}(1) \Rightarrow U(r),$$

where $U(r) \equiv BM(\omega^{2})$. It follows that as $M \to \infty$,

$$\frac{1}{M}\sum_{k=1}^{M}\int_{0}^{1}H_{k,0}^{\kappa}(r)\,dB_{k}(r) \Rightarrow \int_{0}^{1}U(s)\,dU(s)$$

because

$$\begin{aligned}
&\frac{1}{M}\sum_{k=1}^{M}\int_{0}^{1}H_{k,0}^{\kappa}(r)\,dB_{k}(r)\\
&\quad= \frac{1}{M}\sum_{k=1}^{M}\int_{0}^{1}B_{k}(r)\,dB_{k}(r) + \frac{1}{M}\sum_{k=1}^{M}\left(\sum_{f=0}^{k-1}B_{f}(1)\right)B_{k}(1)\\
&\qquad+ B_{0}(-\kappa)\frac{1}{M}\sum_{k=1}^{M}B_{k}(1)\\
&\quad= \sum_{k=1}^{M}\left(\frac{1}{\sqrt{M}}\sum_{f=0}^{k-1}B_{f}(1)\right)\frac{1}{\sqrt{M}}B_{k}(1) + O_{p}\left(\frac{1}{\sqrt{M}}\right)\\
&\quad\Rightarrow \int_{0}^{1}U(r)\,dU(r),
\end{aligned} \tag{A.17}$$

where the final line follows by using partial summation techniques, as in Phillips (1987b). Similarly, we have

$$\frac{1}{M^2}\sum_{k=1}^{M}\int_0^1 H_{k,0}^{\kappa}(r)^2 dr \Rightarrow \int_0^1 U(r)^2 dr. \tag{A.18}$$

Thus, in view of (A.16)–(A.18),

$$mM(\hat{a}^+ - 1) \Rightarrow \frac{\int_0^1 U(r)\,dU(r)}{\int_0^1 U(r)^2 ds},$$

as required. ■

Proof of Theorem 6. For fixed M, as $m \to \infty$ we have, as in (9),

$$e^{cM} m(\hat{a} - a) = \frac{e^{-cM}\sum_{k=1}^{M}\frac{1}{m}\sum_{j=1}^{m} y_{k,j-1}u_{k,j}}{e^{-2cM}\sum_{k=1}^{M}\frac{1}{m^2}\sum_{j=1}^{m} y_{k,j-1}^2}$$

$$\Rightarrow \frac{e^{-cM}\sum_{k=1}^{M}\int_0^1 H_{k,c}^{\kappa}(r)\,dB_k(r) + \lambda M e^{-cM}}{e^{-2cM}\sum_{k=1}^{M}\int_0^1 H_{k,c}^{\kappa}(r)^2 dr}. \tag{A.19}$$

For the limit when $M \to \infty$, we follow arguments similar to those of Basawa and Brockwell (1984). First, consider the numerator of (A.19).

Note that

$$\begin{aligned} & e^{-cM}\sum_{k=1}^{M}\int_0^1 H_{k,c}^{\kappa}(r)\,dB_k(r) + \lambda M e^{-cM} \\ &= e^{-cM}\sum_{k=1}^{M}\int_0^1 J_{k,c}(r)\,dB_k(r) \\ &\quad + \sum_{k=1}^{M} e^{-cM}\left(\sum_{f=0}^{k-1} e^{(k-1-f)c} J_{f,c}(1) + e^{kc} J_{-1,c}(-\kappa)\right)\int_0^1 e^{rc} dB_k(r) + o(1) \\ &= \sum_{k=1}^{M} e^{-cM}\left(\sum_{f=0}^{k-1} e^{(k-1-f)c} J_{f,c}(1) + e^{kc} J_{-1,c}(-\kappa)\right)\int_0^1 e^{rc} dB_k(r) + o_p(1), \end{aligned} \tag{A.20}$$

where the last line holds because

$$e^{-cM}\sum_{k=1}^{M}\int_0^1 J_{k,c}(r)\,dB_k(r) = O_p\left(\frac{\sqrt{M}}{e^{cM}}\right) = o_p(1).$$

Before proceeding further, define

$$
\begin{aligned}
X_{k-1} &= \sum_{f=0}^{k-1} e^{(k-1-f)c} J_{f,c}(1) \\
&= J_{k-1,c}(1) + e^{c} J_{k-2,c}(1) + \cdots + e^{(k-1)c} J_{0,c}(1) \\
&\equiv N\left(0, \left(\sum_{f=0}^{k-1} e^{2(k-1-f)c}\right) \omega^2 \frac{(e^{2c}-1)}{2c}\right) \\
&\equiv N\left(0, \omega^2 \frac{e^{2kc}-1}{2c}\right)
\end{aligned}
$$

and

$$
Q_k = \int_0^1 e^{rc} dB_k(r) \equiv N\left(0, \omega^2 \frac{(e^{2c}-1)}{2c}\right).
$$

Then, we have

$$
X_t = e^{c} X_{t-1} + J_{t,c}(1) \qquad t = 0, \ldots, k-1,
$$

where $X_{-1} = 0$.

Also, note that

$$
\begin{aligned}
\sup_{M \geq 0} E(e^{-cM} X_M)^2 &= \sup_{M \geq 0} e^{-2cM} \omega^2 \frac{e^{2(M+1)c}-1}{2c} \\
&= \sup_{M \geq 0} \omega^2 \frac{e^{2c} - e^{-2cM}}{2c} \leq \frac{\omega^2 e^{2c}}{2c} < \infty.
\end{aligned}
$$

Then, $e^{-cM} X_M$ is a martingale with respect to the filtration $\mathcal{F}_M = \sigma(J_{M,c}(1), J_{M-1,c}(1), \ldots)$, and, by the martingale convergence theorem, we have

$$
e^{-cM} X_M \to_{a.s.} e^{c} Y(c), \text{ where } Y(c) \equiv N\left(0, \frac{\omega^2}{2c}\right). \tag{A.21}
$$

In addition, it is easy to see that

$$
\begin{aligned}
&\sum_{k=0}^{M-1} e^{-c(k+1)} Q_{M-k} \\
&\quad \equiv N\left(0, \left(\sum_{k=0}^{M-1} e^{-2c(k+1)}\right) \frac{\omega^2}{2c} (e^{2c}-1)\right) \\
&\quad \Rightarrow N\left(0, \frac{e^{-2c}}{1-e^{-2c}} \frac{\omega^2}{2c} (e^{2c}-1)\right) = N\left(0, \frac{\omega^2}{2c}\right) := Z(c).
\end{aligned}
$$

Moreover, the limit variates $Z(c)$, $Y(c)$, and $J_{-1,c}(-\kappa)$ are independent, and $Y(c)$ and $Z(c)$ have the same normal distribution.

Next consider the limit of (A.20). Proceeding as in Basawa and Brockwell (1984) we get

$$
\begin{aligned}
&\sum_{k=1}^{M} e^{-cM}\left(\sum_{f=0}^{k-1} e^{(k-1-f)c} J_{f,c}(1) + e^{kc} J_{-1,c}(-\kappa)\right)\int_0^1 e^{rc} dB_k(r) \\
&\quad = (1 \quad e^{c} J_{-1,c}(-\kappa))\begin{pmatrix} \sum_{h=0}^{M-1} e^{-c(M-h-1)} X_{M-h-1}\, e^{-c(h+1)} Q_{M-h} \\ \sum_{h=0}^{M-1} e^{-c(h+1)} Q_{M-h} \end{pmatrix} \\
&\quad \sim e^{c}(Y(c) + J_{-1,c}(-\kappa)) \sum_{h=0}^{M-1} e^{-c(h+1)} Q_{M-h} \\
&\quad \Rightarrow e^{c}(Y(c) + J_{-1,c}(-\kappa)) Z(c), \qquad \textbf{(A.22)}
\end{aligned}
$$

where the third line holds because $|e^{-c(M-k-1)} X_{M-k-1}| \leq K_c$ almost surely for some random variable K_c (cf. Basawa and Brockwell, 1984, pp. 161–171) and we can apply dominated convergence. Thus, as $M \to \infty$

$$
e^{-cM} \sum_{k=1}^{M} \int_0^1 H_{k,c}^{\kappa}(r)\, dB_k(r) + \lambda M e^{-cM} \Rightarrow e^{c}(Y(c) + J_{-1,c}(-\kappa)) Z(c). \qquad \textbf{(A.23)}
$$

Now, we proceed to the denominator of (A.19). By definition

$$
\begin{aligned}
&\frac{1}{e^{2cM}} \sum_{k=1}^{M} \int_0^1 H_{k,c}^{\kappa}(r)^2 dr \\
&\quad = \frac{1}{e^{2cM}} \sum_{k=1}^{M} \int_0^1 J_{k,c}(r)^2 dr \\
&\qquad + 2\, \frac{1}{e^{2cM}} \sum_{k=1}^{M} \int_0^1 e^{rc} J_{k,c}(r)\, dr \left(\sum_{f=0}^{k-1} e^{(k-1-f)c} J_{f,c}(1) + e^{kc} J_{-1,c}(-\kappa)\right) \\
&\qquad + \int_0^1 e^{2rc} dr\, \frac{1}{e^{2cM}} \sum_{k=1}^{M} \left(\sum_{f=0}^{k-1} e^{(k-1-f)c} J_{f,c}(1) + e^{kc} J_{-1,c}(-\kappa)\right)^2 \\
&\quad = I + 2II + III, \quad \text{say.}
\end{aligned}
$$

It is easy to verify that, by the strong law for i.i.d. variates,

$$
I = O_{a.s.}\left(\frac{M}{e^{2cM}}\right) = o_{a.s.}(1).
$$

For III, note that

$$
\begin{aligned}
&\frac{1}{e^{2cM}} \sum_{k=1}^{M} \left(\sum_{f=0}^{k-1} e^{(k-1-f)c} J_{f,c}(1) + e^{kc} J_{-1,c}(-\kappa)\right)^2 \\
&\quad = \sum_{k=0}^{M-1} \left\{ \frac{1}{e^{(M-k-1)c}} \left(X_{M-k-1} + e^{(M-k-1)c} e^{c} J_{-1,c}(-\kappa)\right) \right\}^2 \left(\frac{1}{e^{2(k+1)c}}\right) \\
&\quad = \sum_{k=0}^{\infty} \left\{ \frac{1}{e^{(M-k-1)c}} \left(X_{M-k-1} + e^{(M-k-1)c} e^{c} J_{-1,c}(-\kappa)\right) \right\}^2 \\
&\qquad \times 1\{0 \leq k \leq M-1\} \left(\frac{1}{e^{2(k+1)c}}\right). \qquad \textbf{(A.24)}
\end{aligned}
$$

From (A.21) we know that for fixed k

$$\left\{ \frac{1}{e^{(M-k-1)c}} \left(X_{M-k-1} + e^{(M-k-1)c} e^{c} J_{0,c}(-\kappa) \right) \right\}^2 1\{0 \le k \le M-1\}$$

$$\to_{a.s.} e^{2c} (Y(c) + J_{0,c}(-\kappa))^2,$$

as $M \to \infty$. Because

$$\left\{ \frac{1}{e^{(M-k-1)c}} \left(X_{M-k-1} + e^{(M-k-1)c} e^{c} J_{0,c}(-\kappa) \right) \right\}^2 1\{0 \le k \le M-1\}$$

is almost surely dominated by a random variable as in (A.22), it follows by the dominated convergence theorem that

$$\text{(A.24)} \to_{a.s.} (Y(c) + J_{0,c}(-\kappa))^2 \frac{e^{2c}}{e^{2c}-1}. \tag{A.25}$$

For II, by the Cauchy–Schwarz inequality, we have

$$II \le \left(\frac{1}{e^{2cM}} \sum_{k=1}^{M} \left(\int_0^1 e^{rc} J_{k,c}(r)\, dr \right)^2 \right)^{1/2}$$

$$\times \left(\frac{1}{e^{2cM}} \sum_{k=1}^{M} \left(\sum_{f=0}^{k-1} e^{(k-1-f)c} J_{f,c}(1) + e^{kc} J_{0,c}(-\kappa) \right)^2 \right)^{1/2}$$

$$= o_p\left(\frac{\sqrt{M}}{e^{cM}} \right) O_p(1) = o_p(1),$$

where the first equality holds by (A.25). Thus

$$\frac{1}{e^{2cM}} \sum_{k=1}^{M} \int_0^1 H_{k,c}^{\kappa}(r)^2 dr \to_{a.s.} (Y(c) + J_{0,c}(-\kappa))^2 \frac{e^{2c}}{e^{2c}-1}. \tag{A.26}$$

Finally, combining (A.23) and (A.26), we have

$$e^{cM} m(\hat{a} - a) \Rightarrow \frac{e^{c} Z(c)}{\frac{e^{2c}}{e^{2c}-1} (Y(c) + J_{0,c}(-\kappa))}.$$

It follows that

$$\frac{e^{c(M+1)} m}{e^{2c}-1} (\hat{a} - a) \Rightarrow \frac{Z(c)}{Y(c) + J_{0,c}(-\kappa)},$$

as stated. ■

Proof of Theorem 8. For each k,

$$\frac{1}{m}\sum_t F^{-1}x_{k,t}\frac{y^*_{k,t}}{\sqrt{m}} \Rightarrow \int_0^1 X_k(r)H_{k,c}(r)dr = \begin{bmatrix} \int H_{k,c}(r)dr \\ \int (k+r)H_{k,c}(r)dr \end{bmatrix}. \tag{A.27}$$

Notice that

$$\frac{1}{\sqrt{M}}\sum_{k=1}^{M}\int_0^1 H_{k,c}(r)dr$$
$$= \frac{1}{\sqrt{M}}\sum_{k=1}^{M}\int_0^1 J_{k,c}\,dr + \frac{1}{\sqrt{M}}\sum_{k=1}^{M}\int_0^1 e^{rc}dr\left(\sum_{f=-\infty}^{k-1} e^{(k-1-f)c}J_{f,c}(1)\right). \tag{A.28}$$

We start with the first member of (A.28), and use the relation (cf. Phillips, 1987a)

$$J_{k,c}(r) = B_k(r) + c\int_0^r e^{(r-s)c}B_k(s)ds, \tag{A.29}$$

where $B_k(r) \equiv BM(\omega^2)$. Because $\{B_k(r)\}_{k=1}^M$ is an independent sequence of Brownian motions, it follows that $(1/\sqrt{M})\sum_{k=1}^M B_k(r)$ is $BM(\omega^2)$ for all M and

$$\frac{1}{\sqrt{M}}\sum_{k=1}^{M} B_k(r) \Rightarrow U(r) \equiv BM(\omega^2). \tag{A.30}$$

Hence,

$$\frac{1}{\sqrt{M}}\sum_{k=1}^{M}\int_0^1 J_{k,c}\,dr \Rightarrow \int_0^1\left[U(r) + c\int_0^r e^{(r-s)c}U(s)ds\right]dr$$
$$= \int_0^1\int_0^r e^{(r-s)c}dU(s)dr$$
$$= \frac{1}{c}e^c\int_0^1 e^{-sc}dU(s) - \frac{1}{c}U(1) \tag{A.31}$$

by using partial integration.

Next, for the second member of (A.28), let

$$\eta_k = \sum_{f=-\infty}^{k-1} e^{(k-1-f)c}J_{f,c}(1).$$

Then

$$\frac{1}{\sqrt{M}}\sum_{k=1}^{M}\int_0^1 e^{rc}dr\left(\sum_{f=-\infty}^{k-1} e^{(k-1-f)c}J_{f,c}(1)\right) = \int_0^1 e^{rc}dr\,\frac{1}{\sqrt{M}}\sum_{k=1}^{M}\eta_k.$$

The time series η_k is a linear process of the form

$$\eta_k = a(L)\nu_k,$$

where $\{\nu_k = J_{k-1,c}(1)\}$ is a sequence of independent normal innovations and $a(L) = \sum_{j=0}^{\infty} a_j L^j$, with $a_j = e^{jc}$. Thus, following the approach of Phillips and Solo (1992), we can write

$$\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \eta_k = a(1) \frac{1}{\sqrt{M}} \sum_{k=1}^{M} J_{k-1,c}(1) + O_p\left(\frac{1}{\sqrt{M}}\right)$$
$$= \frac{1}{1-e^c} \frac{1}{\sqrt{M}} \sum_{k=1}^{M} J_{k-1,c}(1) + O_p\left(\frac{1}{\sqrt{M}}\right).$$

Then, as in (A.29) and (A.30), we obtain

$$\frac{1}{\sqrt{M}} \sum_{k=1}^{M} J_{k-1,c}(1) \Rightarrow U(1) + c \int_0^1 e^{(1-s)c} U(s) ds = \int_0^1 e^{(1-s)c} dU(s).$$

The limit behavior of the second member of (A.28) is therefore

$$\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \int_0^1 e^{rc} dr \left(\sum_{f=-\infty}^{k-1} e^{(k-1-f)c} J_{f,c}(1) \right)$$
$$\Rightarrow \left(\int_0^1 e^{rc} dr \right) \left(\frac{1}{1-e^c} \right) \left(\int_0^1 e^{(1-s)c} dU(s) \right)$$
$$= -\frac{1}{c} \left(\int_0^1 e^{(1-s)c} dU(s) \right). \tag{A.32}$$

Combining (A.28), (A.31), and (A.32) we obtain

$$\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \int_0^1 H_{k,c}(r) dr \Rightarrow \frac{1}{c} e^c \int_0^1 e^{-sc} dU(s) - \frac{1}{c} U(1) - \frac{1}{c} \left(\int_0^1 e^{(1-s)c} dU(s) \right)$$
$$= -\frac{1}{c} U(1) = -\frac{1}{c} \int_0^1 dU(s). \tag{A.33}$$

If we denote $\int_0^1 H_{k,c}(r) dr$ by ζ_k, under the assumption of homogeneity, ζ_k is a stationary process, and, in a similar way to (A.33), the partial sum

$$\frac{1}{\sqrt{M}} \sum_{k=1}^{[Mt]} \zeta_k \Rightarrow B_\zeta(t) = -\frac{1}{c} U(t).$$

Thus

$$\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \int \left(\frac{k+r}{M} \right) H_{k,c}(r) dr = \frac{1}{\sqrt{M}} \sum_{k=1}^{M} \frac{k}{M} \int H_{k,c}(r) dr + O_p\left(\frac{1}{M}\right)$$
$$= \sum_{k=1}^{M} \left(\frac{k}{M} \right) \frac{\zeta_k}{\sqrt{M}} + O_p\left(\frac{1}{M}\right)$$
$$\Rightarrow \int_0^1 s dB_\zeta(s) = -\frac{1}{c} \int_0^1 s dU(s). \tag{A.34}$$

Thus, from (A.27), (A.33), and (A.34) we obtain

$$\frac{1}{m\sqrt{n}}\sum_k\sum_t D^{-1}x_{k,t}y^*_{k,t} = \frac{1}{\sqrt{M}}\sum_{k=1}^{M} G^{-1}\frac{1}{m}\sum_t F^{-1}x_{k,t}\frac{y^*_{k,t}}{\sqrt{m}}$$

$$\Rightarrow \frac{1}{\sqrt{M}}\sum_{k=1}^{M} G^{-1}\int_0^1 X_k(r)H_{k,c}(r)dr$$

$$= \frac{1}{\sqrt{M}}\sum_{k=1}^{M}\begin{bmatrix}\int_0^1 H_{k,c}(r)dr \\ \int_0^1\left(\frac{k+r}{M}\right)H_{k,c}(r)dr\end{bmatrix}$$

$$\Rightarrow -\frac{1}{c}\begin{bmatrix}\int_0^1 dU(s) \\ \int_0^1 sdU(s)\end{bmatrix} = -\frac{1}{c}\int_0^1 X(s)\,dU(s)$$

in sequential asymptotics as $(m, M \to \infty)_{seq}$. Further, in a straightforward way

$$\frac{1}{n}\sum_k\sum_t D^{-1}x_{k,t}x'_{k,t}D^{-1} \to \int_0^1 X(r)X(r)'dr.$$

In consequence, we have

$$\frac{\sqrt{n}}{m}D(\hat{\gamma}-\gamma) = \left(\frac{1}{n}\sum_k\sum_t D^{-1}x_{k,t}x'_{k,t}D^{-1}\right)^{-1}\left(\frac{1}{m\sqrt{n}}\sum_k\sum_t D^{-1}x_{k,t}y^*_{k,t}\right)$$

$$\Rightarrow -\frac{1}{c}\left(\int_0^1 X(r)X(r)'dr\right)^{-1}\left(\int_0^1 X(s)\,dU(s)\right),$$

giving the stated result. ■

Proof of Lemma 9.

$$m^{-1/2}\hat{y}^*_{k,[mr]} = m^{-1/2}y^*_{k,[mr]} - m^{-1/2}(\hat{\gamma}-\gamma)'x_{k,[mr]}$$

$$= m^{-1/2}y^*_{k,[mr]} - \frac{1}{\sqrt{M}}\frac{\sqrt{n}}{m}(\hat{\gamma}-\gamma)'DD^{-1}x_{k,[mr]}$$

$$= m^{-1/2}y^*_{k,[mr]} + O_p(M^{-1/2})$$

$$\Rightarrow H_{k,c}(r).$$

■

Proof of Theorem 10.

$$\tilde{a} = a + \frac{\sum_k \sum_t \hat{y}^*_{k,t-1} u_{k,t}}{\sum_k \sum_t (\hat{y}^*_{k,t-1})^2}$$

$$+ \frac{(a-1)(\hat{\gamma}-\gamma)' \sum_k \sum_t x_{k,t-1} \hat{y}^*_{k,t-1}}{\sum_k \sum_t (\hat{y}^*_{k,t-1})^2}$$

$$- \frac{(\hat{\gamma}-\gamma)' \sum_k \sum_t (x_{k,t} - x_{k,t-1}) \hat{y}^*_{k,t-1}}{\sum_k \sum_t (\hat{y}^*_{k,t-1})^2}.$$

Thus

$$\sqrt{M} m(\tilde{a} - a)$$

$$= \frac{\frac{1}{\sqrt{M}} \sum_k \left(\frac{1}{m} \sum_t \hat{y}^*_{k,t-1} u_{k,t} \right)}{\frac{1}{M} \sum_k \left(\frac{1}{m^2} \sum_t (\hat{y}^*_{k,t-1})^2 \right)}$$

$$+ \frac{1}{\sqrt{M}} \frac{m(a-1)\left[(\hat{\gamma}-\gamma)' D \frac{\sqrt{M}}{\sqrt{m}}\right]\left[\frac{1}{\sqrt{M}} \sum_k G^{-1}\left(\frac{1}{m}\sum_t F^{-1} x_{k,t-1} \frac{\hat{y}^*_{k,t-1}}{\sqrt{m}}\right)\right]}{\frac{1}{M}\sum_k \left[\frac{1}{m^2}\sum_t (\hat{y}^*_{k,t-1})^2\right]}$$

$$- \frac{1}{M^{3/2}} \frac{\left[(\hat{\gamma}-\gamma)' D \frac{\sqrt{M}}{\sqrt{m}}\right]\left[\frac{1}{\sqrt{M}} \sum_k \left(\frac{1}{m}\sum_t nD^{-1}(x_{k,t} - x_{k,t-1}) \frac{\hat{y}^*_{k,t-1}}{\sqrt{m}}\right)\right]}{\frac{1}{M}\sum_k \left[\frac{1}{m^2}\sum_t (\hat{y}^*_{k,t-1})^2\right]}$$

$$= \frac{\frac{1}{\sqrt{M}} \sum_k \left(\frac{1}{m} \sum_t \hat{y}^*_{k,t-1} u_{k,t} \right)}{\frac{1}{M} \sum_k \left(\frac{1}{m^2} \sum_t (\hat{y}^*_{k,t-1})^2 \right)} + O_p\left(\frac{1}{\sqrt{M}}\right) + O_p\left(\frac{1}{M^{3/2}}\right).$$

As $m \to \infty$

$$\frac{\frac{1}{\sqrt{M}} \sum_k \left(\frac{1}{m} \sum_t \hat{y}^*_{k,t-1} u_{k,t} \right)}{\frac{1}{M} \sum_k \left(\frac{1}{m^2} \sum_t (\hat{y}^*_{k,t-1})^2 \right)} \Rightarrow \frac{\frac{1}{\sqrt{M}} \sum_k \left(\int H_{k,c}\, dB_k + \lambda_k \right)}{\frac{1}{M} \sum_k \int H_{k,c}(r)^2}.$$

Thus, if we consider the adjusted estimator $\hat{a}^{+}$ and take sequential limits as $(m, M \to \infty)_{seq}$, we have

$$\sqrt{M}m(\tilde{a}^{+} - a) = \frac{\frac{1}{\sqrt{M}}\sum_k \left(\frac{1}{m}\sum_t \hat{y}^*_{k,t-1}u_{k,t} - \hat{\lambda}_k\right)}{\frac{1}{M}\sum_k \left(\frac{1}{m^2}\sum_t (\hat{y}^*_{k,t-1})^2\right)} + o_p(1) \Rightarrow N(0, -2c). \quad \blacksquare$$

Proof of Lemma 11. When $c = 0$,

$$\frac{1}{\sqrt{n}} D(\hat{\gamma} - \gamma) = \left[\frac{1}{n}\sum_{s=1}^{n} D^{-1}x_s x_s' D^{-1}\right]^{-1}\left[\frac{1}{n}\sum_{s=1}^{n} D^{-1}x_s \frac{z_s^*}{\sqrt{n}}\right]$$

$$\Rightarrow \left[\int X(r)X(r)'\right]^{-1}\left[\int X(r)U(r)\right].$$

By definition,

$$\hat{y}^*_{k,t} = y_{k,t} - \hat{\gamma}' x_{k,t},$$

$$\hat{z}^*_s = z_s - \hat{\gamma}' x_s,$$

and thus

$$n^{-1/2}\hat{z}^*_{[nr]} = n^{-1/2}z^*_{[nr]} - n^{-1/2}(\hat{\gamma} - \gamma)' D D^{-1} x_{[nr]}$$

$$\Rightarrow U(r) - \left[\int UX'\right]\left[\int XX'\right]^{-1} X(r) := \underline{U}(r). \quad \blacksquare$$

Proof of Theorem 12. Because

$$\tilde{a} = 1 + \frac{\sum_k \sum_t \hat{y}^*_{k,t-1}u_{k,t}}{\sum_k \sum_t (\hat{y}^*_{k,t-1})^2} + \frac{(\hat{\gamma} - \gamma)' \sum_k \sum_t (x_{k,t} - x_{k,t-1})\hat{y}^*_{k,t-1}}{\sum_k \sum_t (\hat{y}^*_{k,t-1})^2},$$

and

$$n(\tilde{a} - 1) = \frac{\sum_s (\hat{z}^*_{s-1}/\sqrt{n})(v_s/\sqrt{n}) - n^{-1/2}(\hat{\gamma} - \gamma)' D\left[n^{-1}\sum_s n D^{-1}\Delta x_s \hat{z}^*_{s-1}/\sqrt{n}\right]}{n^{-1}\sum_s (\hat{z}^*_{s-1}/\sqrt{n})^2},$$

we get

$$n(\tilde{a}^{+} - 1) \Rightarrow \frac{\int \underline{U}dU - \int UX'\left[\int XX'\right]^{-1}\int X_0 \underline{U}}{\int \underline{U}^2},$$

where $X_0(r) = X'(r) = (0,1)' = PX(r)$ with

$$P = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Because

$$\int X_0 \underline{U} = P \int X\underline{U} = 0,$$

we obtain

$$n(\tilde{a}^+ - 1) \Rightarrow \frac{\int \underline{U} dU}{\int \underline{U}^2},$$

giving the stated result. ■

Proof of Theorem 13. By definition,

$$\tilde{\gamma} = \left[\sum_{k=1}^{M} \sum_{t=1}^{m} \Delta_c x_{k,t} \Delta_c x_{k,t}' \right]^{-1} \left[\sum_{k=1}^{M} \sum_{t=1}^{m} \Delta_c x_{k,t} \Delta_c y_{k,t} \right]$$
$$= \gamma + \left[\sum_{k=1}^{M} \sum_{t=1}^{m} \Delta_c x_{k,t} \Delta_c x_{k,t}' \right]^{-1} \left[\sum_{k=1}^{M} \sum_{t=1}^{m} \Delta_c x_{k,t} u_{k,t} \right].$$

Thus,

$$\sqrt{n} N(\tilde{\gamma} - \gamma) = \left[\frac{1}{n} \sum_s N^{-1} \Delta_c x_{k,t} \Delta_c x_{k,t}' N^{-1} \right]^{-1} \left[\frac{1}{\sqrt{n}} \sum_s N^{-1} \Delta_c x_{k,t} u_{k,t} \right]$$
$$\Rightarrow \left[\int X_c(r) X_c(r)' \right]^{-1} \int X_c(r) dU(r),$$

where

$$N = \operatorname{diag}[m^{-1}, M], \qquad X_c(r) = -c(1,r)'.$$

Simply noting that $X_c(r) = -cX(r)$ gives the stated result for $\tilde{\gamma}$.

Suppose that $c < 0$ and we estimate c by $\hat{c} = m(\tilde{a} - 1)$, as in Section 5. Then $\hat{c} = c + O_p(M^{-1/2})$. If we apply QD detrending to model (14)–(16) based on $\hat{c}$, we get

$$\Delta_{\hat{c}} y_{k,t} = \gamma' \Delta_{\hat{c}} x_{k,t} + \Delta_{\hat{c}} y_{k,t}^{*}. \tag{A.35}$$

The OLS estimator of γ from (A.35) is

$$\tilde{\gamma}_f = \left[\sum_{k=1}^{M} \sum_{t=1}^{m} \Delta_{\hat{c}} x_{k,t} \Delta_{\hat{c}} x_{k,t}' \right]^{-1} \left[\sum_{k=1}^{M} \sum_{t=1}^{m} \Delta_{\hat{c}} x_{k,t} \Delta_{\hat{c}} y_{k,t} \right]$$
$$= \gamma + \left[\sum_{k=1}^{M} \sum_{t=1}^{m} \Delta_{\hat{c}} x_{k,t} \Delta_{\hat{c}} x_{k,t}' \right]^{-1} \left[\sum_{k=1}^{M} \sum_{t=1}^{m} \Delta_{\hat{c}} x_{k,t} \Delta_{\hat{c}} y_{k,t}^{*} \right].$$

Notice that

$$\Delta_{\hat{c}} y^*_{k,t} = u_{k,t} - (\hat{c} - c) m^{-1} y^*_{k,t-1}$$

and

$$\frac{1}{\sqrt{n}} \sum_s N^{-1} \Delta_{\hat{c}} x_{k,t} \Delta_{\hat{c}} y^*_{k,t}$$

$$= \frac{1}{\sqrt{n}} \sum_s N^{-1} \Delta_{\hat{c}} x_{k,t} u_{k,t} - (\hat{c} - c) \frac{1}{\sqrt{M}} \sum_k N_1^{-1} \frac{1}{m} \sum_t N_2^{-1} \Delta_{\hat{c}} x_{k,t} \frac{y^*_{k,t-1}}{\sqrt{m}}$$

$$= \frac{1}{\sqrt{n}} \sum_s N^{-1} \Delta_{\hat{c}} x_{k,t} u_{k,t} + O_p(M^{-1/2}),$$

where

$$N_1 = \mathrm{diag}[1, M], \qquad N_2 = \mathrm{diag}[m^{-1}, 1], \qquad N = N_1 N_2.$$

It can be verified that the error terms coming from the preliminary estimation of c are of smaller order of magnitude and

$$\sqrt{n} N(\tilde{\gamma}_f - \gamma) \Rightarrow \left[\int X_c(r) X_c(r)' \right]^{-1} \int X_c(r) dU(r).$$

The limiting distribution of the trend coefficient vector $\tilde{\gamma}_f$ is then the same as that of $\tilde{\gamma}$, the estimator using the true local parameter. ■