Econometric Theory, **16**, 2000, 927–997. Printed in the United States of America.

ESTIMATION OF AUTOREGRESSIVE ROOTS NEAR UNITY USING PANEL DATA

Hyungsik R. Moon
University of California, Santa Barbara
and
University of Southern California

Peter C.B. Phillips
Yale University,
University of Auckland, and University of York

Time series data are often well modeled by using the device of an autoregressive root that is local to unity. Unfortunately, the localizing parameter (c) is not consistently estimable using existing time series econometric techniques and the lack of a consistent estimator complicates inference. This paper develops procedures for the estimation of a common localizing parameter using panel data. Pooling information across individuals in a panel aids the identification and estimation of the localizing parameter and leads to consistent estimation in simple panel models. However, in the important case of models with concomitant deterministic trends, it is shown that pooled panel estimators of the localizing parameter are asymptotically biased. Some techniques are developed to overcome this difficulty, and consistent estimators of c in the region $c < 0$ are developed for panel models with deterministic and stochastic trends. A limit distribution theory is also established, and test statistics are constructed for exploring interesting hypotheses, such as the equivalence of local to unity parameters across subgroups of the population. The methods are applied to the empirically important problem of the efficient extraction of deterministic trends. They are also shown to deliver consistent estimates of distancing parameters in nonstationary panel models where the initial conditions are in the distant past. In the development of the asymptotic theory this paper makes use of both sequential and joint limit approaches. An important limitation in the operation of the joint asymptotics that is sometimes needed in our development is the rate condition $n/T \to 0$. So the results in the paper are likely to be most relevant in panels where T is large and n is moderately large.

The authors thank the co-editor, Bruce Hansen, and four anonymous referees for comments on the earlier version of the paper and Donald Andrews for helpful discussions. Phillips thanks the NSF for research support under Grants SBR 94-22922 and SBR 97-30295, and Moon gratefully acknowledges financial support from a C.A. Anderson Prize Fellowship. Address correspondence to: Hyungsik R. Moon, Department of Economics, University of Southern California, KAP 300, University Park Campus, Los Angeles, CA 90089, USA.

 0266-4666/00 $9.50

1. INTRODUCTION

Time series models with roots near unity are extremely common in econometric applications, and this feature of the data is often modeled by using the device of an autoregressive root that is local to unity, so that the time series has the property of being near integrated. Such time series are more general than integrated processes, and they allow more flexibility in the econometric modeling of nonstationary series. Whereas the local to unity parameter cannot be consistently estimated using existing time series methods,[1] it is useful in many different econometric contexts. A few examples are as follows: the analysis of power properties of unit root tests (Phillips, 1987a); the construction of confidence intervals for the long run autoregressive coefficient (Stock, 1991); the development of efficient detrending methods (Phillips and Lee, 1996; Canjels and Watson, 1997); and the construction of point optimal invariant tests for a unit root (Elliott, Rothenberg, and Stock, 1996) and cointegrating rank (Xiao and Phillips, 1999).

This paper develops procedures for the estimation of the local to unity parameter in panel data models. When there is a common time series local to unity parameter across independent individuals in a panel, it is apparent that the cross section data carry additional information that can be used to assist in estimating a common localizing parameter (c). By simple pooling of time series estimates, we might expect that a common local to unity parameter could be consistently estimated with panel data that combined independent observations across individuals. In the case where the data generating process involves only a near-integrated stochastic trend process, we show that a simple pooled least-squares estimator does produce a consistent estimator for the local to unity parameter. However, the simple data-pooling heuristic does not hold in situations where there are both deterministic and near-integrated stochastic trends in the model. In such cases, it is shown that the pooled least-squares estimator of the localizing coefficient c generates an inconsistency that depends upon the true unknown localizing parameter. To resolve this problem, we develop a consistent estimator for c in the important case where $c < 0$. Asymptotic normality of these consistent local to unity parameter estimators is established, and the limit theory is used to develop an inferential framework for local to unity modeling in panel data. In particular, test statistics are constructed for exploring interesting hypotheses, such as the equivalence of the local to unity parameter across subgroups of the population.

Local to unity parameter estimation is useful in many empirical applications. We illustrate the usefulness of panel estimation of the localizing coefficient with an application to efficient deterministic trend extraction and the construction of confidence intervals for models with roots near unity. According to Phillips and Lee (1996), when the regression errors are near integrated, efficiency gains in the estimation of deterministic trends can be obtained by quasi-differencing the data. However, to implement this procedure in practice, the localizing parameter in the near-integrated error process must be known or be consistently estimable, neither of which normally applies. If inconsistent estimates of the localizing parameter are used instead, then the resulting trend coefficient estimator has a highly nonstandard limit distribution, which gives rise to new difficulties, for example, in setting up confidence intervals for the trend coefficient. Because of this problem, Cavanagh, Elliott, and Stock (1995) and Canjels and Watson (1997) suggest the use of Bonferroni-type confidence intervals, which are often very conservative. In panel data models, our consistent estimate of the local to unity parameter can be used to overcome these difficulties. In fact, our feasible efficient estimator based on consistent panel data estimates of the local to unity parameter has a standard limit distribution, and this limit theory leads to a conventional form of confidence interval for the trend.

Another useful application of panel data for nonstationary time series lies in the consistent estimation of the distancing parameter that arises in the formulation of distant initial conditions. The distancing parameter, which is expressed as a fraction (not necessarily less than unity) of the length of the present time series sample, measures how far into the past the initialization extends in terms of the shocks that have determined it. It is shown that consistent estimation of this parameter is possible with panel data when there is common distancing in the initialization across the panel and a common local to unity parameter in the dynamics. In effect, panel variation across individuals enables us to learn something very specific about the nature of presample data—how far its origins extend in relation to the historically observed data.

In other recent research (Phillips and Moon, 1999), the authors develop some rigorous asymptotic theory for multi-index situations in which two indices may pass to infinity. This general theory is applied to obtain a nonstationary panel data limit theory where there are large numbers of cross section (n) and time series (T) observations. The new limit theory allows for both sequential limits, where $T \to \infty$ and $n \to \infty$ sequentially, and joint limits where $T, n \to \infty$ simultaneously. The present paper makes use of those methods in the development of the asymptotic theory here. An important limitation in the operation of the joint asymptotics that is sometimes needed in our development is the rate condition $n/T \to 0$. This condition means that the results are likely to be most relevant in panels where T is large and n is moderately large (as is the case in some cross country macroeconomic panels).

The paper is organized as follows. Section 2 lays out the model and assumptions, gives some heuristic discussion, and shows how consistent estimation of the localizing parameter is possible in panel models with no deterministic components. Section 3 studies the same problem in models with deterministic trend components, shows the inconsistency of the pooled least-squares estimator, and develops several alternative approaches to dealing with the bias problem. A consistent estimator is given for the case where the common localizing parameter satisfies $c < 0$. A limit distribution theory is developed, and matters of inference are discussed. Section 4 applies these methods to testing for the localizing coefficient, to the empirically important problem of the efficient estimation of the deterministic trend coefficients, and to estimation of the distancing parameter that arises in the formulation of distant initial conditions. Section 5 concludes the paper. Proofs, technical derivations, and a brief review of some double index asymptotic theory are given in the Appendixes.

2. MODELS, ASSUMPTIONS, AND HEURISTICS

We start by assuming that the time series process for individual i, $z_{i,t}$, has a decomposition into both deterministic and stochastic elements as follows:

$$z_{i,t} = \beta_{i,0} + \beta_i' g_t + y_{i,t}, \qquad t = 1,\dots,T; \qquad i = 1,\dots,n,$$

$$y_{i,t} = a y_{i,t-1} + \varepsilon_{i,t}, \qquad a = \exp\left(\frac{c}{T}\right), \tag{1}$$

where $g_t = (t,\dots,t^p)'$ is a deterministic polynomial trend, $\beta_i = (\beta_{i,1},\dots,\beta_{i,p})'$, and $y_{i,t}$ is a near-integrated stochastic process. The initialization is at $t = 0$ with random variables $y_{i,0}$ that are independent and identically distributed (i.i.d.) across i with mean zero and finite variance $\sigma_{i,0}^2$ for all i. In this paper we assume that the deterministic trends $\beta_{i,0} + \beta_i' g_t$ in (1) are heterogeneous across i.[2] These heterogeneous trends reflect individual effects in the panel data $z_{i,t}$.

The parameter c in the $AR(1)$ coefficient a is the local to unity parameter, which is assumed here to be common to all individuals. One of the aims of this paper is to find a consistent estimation procedure for the parameter c. The common localizing parameter c can be considered a common limit of individually different sequences of local parameters. That is, we may regard the $AR(1)$ error process coefficient a as the limit of the sequence of coefficients $a_{i,T} = \exp((c + c_{i,T})/T)$, where $c_{i,T}/T \to 0$ uniformly in i. In this case the common coefficient $a = \exp(c/T)$ is an approximation of $a_{i,T} = \exp((c + c_{i,T})/T)$. In some empirical applications, it may be too restrictive to assume a common localizing coefficient in the panel regression model (1) for all individuals. Therefore, procedures that allow for some cross sectional heterogeneity in the localizing parameter and procedures for testing cross sectional heterogeneity in localizing coefficients will certainly be of interest in empirical work. As a partial solution of the latter problem, this paper develops a testing procedure designed to assess whether the localizing parameter is the same across subgroups of individuals in the sample.

With regard to the specification of the trend component in (1), it is important to note that individual intercept terms $\beta_{i,0}$ are not consistently estimable with time series data when the stochastic component $y_{i,t}$ is near integrated, as a result of the low signal to noise ratio relative to the latent stochastic trend $y_{i,t}$ in (1), namely, $1/\text{var}(y_{i,t}) = O(1/t) \to 0$ as $t \to \infty$. The $O_p(1)$ assumption for the initial conditions of $y_{i,t}$ is made for convenience and could be extended in the usual way to allow for distant initialization (Uhlig, 1994; Phillips and Lee, 1996; Canjels and Watson, 1997), at the cost of some additional complexity.

To develop some quick results, we first consider the simple case where the trend coefficient vectors β_i are known (but intercept terms $\beta_{i,0}$ are unknown) and the error processes $\varepsilon_{i,t}$ are i.i.d. $(0, \sigma_\varepsilon^2)$ across i and over t. This covers the case where there is no deterministic trend in (1) and $\beta_i = 0$. In this case, the variables $\hat{z}_{i,t} = z_{i,t} - \beta_i' g_t$ are observable. In time series regression, taking into account the relation $a \simeq 1 + c/T$, the natural estimator for c is $\tilde{c} = T(\tilde{a} - 1)$ where

$$\tilde{a} = \left(\sum_{t=1}^{T} \hat{z}_{i,t-1}^2 \right)^{-1} \left(\sum_{t=1}^{T} \hat{z}_{i,t-1} \hat{z}_{i,t} \right).$$

Then, as $T \to \infty$

$$\begin{aligned} T(\tilde{a} - a) &= \left[\frac{1}{T^2} \sum_{t=1}^{T} \hat{z}_{i,t-1}^2 \right]^{-1} \left[\frac{1}{T} \sum_{t=1}^{T} \hat{z}_{i,t-1}((1-a)\beta_{i,0} + \varepsilon_{i,t}) \right] \\ &= \left[\frac{1}{T^2} \sum_{t=1}^{T} (y_{i,t-1} + \beta_{i,0})^2 \right]^{-1} \\ &\quad \times \left[\frac{1}{T} \sum_{t=1}^{T} (y_{i,t-1} + \beta_{i,0})((1-a)\beta_{i,0} + \varepsilon_{i,t}) \right] \\ &\Rightarrow \left(\int_0^1 J_{c,i}(r)^2 dr \right)^{-1} \int_0^1 J_{c,i}(r) dW_i(r), \end{aligned}$$

where $J_{c,i}(r) = \int_0^r e^{(r-s)c} dW_i(s)$ and $W_i(r)$ is a standard Brownian motion (e.g., see Phillips, 1987b). From

$$a = \exp\left(\frac{c}{T}\right) = 1 + \frac{c}{T} + O\left(\frac{1}{T^2}\right), \tag{2}$$

we have

$$\begin{aligned} \tilde{c} - c &= T(\tilde{a} - 1) - c = T(\tilde{a} - a) + O\left(\frac{1}{T}\right) \\ &\Rightarrow \left(\int_0^1 J_{c,i}(r)^2 dr \right)^{-1} \int_0^1 J_{c,i}(r) dW_i(r). \end{aligned}$$

Thus, as is well known, $\tilde{c}$ is not a consistent estimator for c and has a nondegenerate limit distribution.

Now suppose that panel data for $y_{i,t}$ are available. Again, one of the natural ways to estimate the common $AR(1)$ coefficient a is to pool the data and run a least-squares regression. Then, we would have

$$\hat{a} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{z}_{i,t-1}^2 \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{z}_{i,t-1} \hat{z}_{i,t} \right), \tag{3}$$

and, again, in view of (2) we define

$$\hat{c} = T(\hat{a} - 1). \tag{4}$$

To take a quick look at the asymptotic behavior of $\hat{a}$ (or, equivalently, $\hat{c}$), we consider the sequential weak limit of $T(\hat{a}-a)$ by letting $T\to\infty$ first, followed by $n\to\infty$, which we denote by $(T,n\to\infty)_{seq}$ (see Phillips and Moon, 1999; and the remark that follows). Now we have

$$T(\hat{a}-a)=\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}(y_{i,t-1}+\beta_{i,0})^2\right]^{-1} \times\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}(y_{i,t-1}+\beta_{i,0})((1-a)\beta_{i,0}+\varepsilon_{i,t})\right]. \quad \mathbf{(5)}$$

As T goes to infinity while n is fixed, we have, as earlier,

$$T(\hat{a}-a)\Rightarrow\left(\frac{1}{n}\sum_{i=1}^{n}\int J_{c,i}(r)^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\int J_{c,i}(r)dW_i(r)\right). \quad \mathbf{(6)}$$

Note that $E(J_{c,i}(r)dW_i(r))=0$ and $E(\int J_{c,i}(r)^2)=\int_0^1\int_0^r e^{2c(r-s)}dsdr>0$. By the weak law of large numbers, as $n\to\infty$, $1/n\sum_{i=1}^n\int J_{c,i}(r)dW_i(r)\to_p 0$. Therefore, in sequential limits as $(T,n\to\infty)_{seq}$, we find that $T(\hat{a}-a)\to_p 0$ and

$$\hat{c}-c=T(\hat{a}-a)+o(1)\to_p 0. \quad \mathbf{(7)}$$

That is, $\hat{c}$ is a consistent estimator for the local to unity parameter c in sequential limits as $(T,n\to\infty)_{seq}$.

Remarks.

(a) The preceding asymptotic theory employs a sequential approach in which the index T passes to infinity first and then the index n passes to infinity later, which is denoted as $(T,n\to\infty)_{seq}$. In general, depending on how the two indices, n and T, are treated, it is possible to have a variety of limit results for double indexed random sequences. Recently, Phillips and Moon (1999) have studied this matter and suggested various limit concepts for multi-indexed sequences, classifying the main concepts into the following three cases: a sequential approach, a diagonal path approach, and a joint approach. The sequential approach passes the indices to infinity sequentially. In the present case, depending on which index tends to infinity first, we may have two different sequential limits according as $(T,n\to\infty)_{seq}$ or $(n,T\to\infty)_{seq}$, where the order of appearance of the index in the notation gives the order of the passage to infinity. The diagonal path approach allows the two indices, n and T, to pass to infinity along a specific diagonal path, say, $(n,T(n))$, in the two dimensional array. This approach simplifies the asymptotic theory by replacing the double indexed process with a single indexed process. The joint approach allows both indices, n and T, to pass to infinity simultaneously without placing specific diagonal path restrictions on the divergence. On the other hand, to obtain some joint limit results, we often need to exercise control over the relative rate of expansion of the two indices. One such requirement that is used in the present paper is $n/T\to 0$, and in such cases there will be a presumption that T is large relative to n in the limit. Although this requirement is not unreasonable for some recent macroeconomic panels, it is much less relevant in traditional dynamic panels, where n is often very large and T is quite small. In such cases, fixed T with large n asymptotics or joint asymptotics with $T/n \to 0$ will be more relevant. The present paper focuses mainly on sequential asymptotics with $(T, n \to \infty)_{seq}$ and joint asymptotics under $n/T \to 0$.

(b) We emphasize that the different approaches may yield different limits. Apostol (1974, p. 200) gives examples of real number sequences with this property, and Phillips and Moon (2000) give examples for double sequences of random variables. In light of such examples, it is natural to ask whether there are cases where the different approaches yield the same limit. The paper by Phillips and Moon (1999) provides a partial answer to this question, focusing on the relation between sequential limits and joint limits. Appendix B of the paper summarizes some important details about these relations.

(c) As the preceding analysis indicates, sequential limits are often easy to derive. Indeed, they are usually much easier to derive than joint limits. As a device for obtaining quick asymptotic results, we will proceed in this paper with $(T, n \to \infty)_{seq}$ sequential limits and then, in the Appendix, demonstrate the results under the more general environment of joint limits. There are two main reasons for dealing with $(T, n \to \infty)_{seq}$ limits instead of $(n, T \to \infty)_{seq}$ limits. The first is simply convenience. In many of the cases investigated in this paper, deriving $(T, n \to \infty)_{seq}$ limits is relatively straightforward and is especially advantageous when the nonstationary time series $y_{i,t}$ in model (1) are generated from weakly dependent processes such as those in Assumption 1, which follows. Second, $(T, n \to \infty)_{seq}$ limits seem appropriate for some recent cross country macroeconomic panels such as those of the Penn World Tables. Later in the paper and as relevant matters arise, some further discussion of these issues will be provided.

The consistency of $\hat{c}$ in (4) depends upon two unrealistic assumptions: (i) the $\varepsilon_{i,t}$ are i.i.d. $(0, \sigma_\varepsilon^2)$; and (ii) the trend coefficient vectors β_i are known. When the $\varepsilon_{i,t}$ are serially dependent, as in Assumption 1, which follows, the limit of $T(\hat{a} - a)$ in (6) involves a bias term that depends on the one-sided long-run covariance of $\varepsilon_{i,t}$. In this case, we can correct the bias easily, for example, by estimating the one-sided long-run variance nonparametrically as in Phillips (1987a) or by using parametric autoregressions in which the order of the autoregression expands with the sample size, as in Said and Dickey (1984).

When the β_i are unknown, the problem becomes much more complicated. The obvious point of departure is to remove the deterministic trends by preliminary regression and then to define $\hat{c} = T(\hat{a} - 1)$, where the estimator $\hat{a}$ is obtained by autoregression with the detrended data. Thus, suppose $\underset{\sim}{z}_{i,t}$ and $\underset{\sim}{z}_{i,t-1}$ are the detrended data, obtained as regression residuals of $z_{i,t}$ and $z_{i,t-1}$ on g_t. Then, we have $\hat{c} = T(\hat{a} - 1)$, where

$$\hat{a} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1} \underset{\sim}{z}_{i,t} \right).$$

This estimator of c is a simple extension of that used in the case where the trends were known. Interestingly, however, $\hat{c}$ is not consistent in this case. Intuitively, the reason for the inconsistency is that preliminary detrending filters the stochastic trend $\underset{\sim}{y}_{i,t-1}$ and the filtered process is correlated with the stationary error process $\tilde{\varepsilon}_{i,t}$ in (1). These matters will be explored in the next section.

We close this section with two assumptions on the error process $\varepsilon_{i,t}$.

Assumption 1. $\varepsilon_{i,t}$ are linear processes satisfying the following conditions:

(a) $\varepsilon_{i,t} = C_i(L)u_{i,t} = \sum_{j=0}^{\infty} C_{i,j} u_{i,t-j}$.
(b) $u_{i,t}$ are i.i.d. across i and over t with $Eu_{i,t} = 0$, $Eu_{i,t}^2 = 1$, and $Eu_{i,t}^4 = \sigma_4$.
(c) $C_{i,j}$ are a sequence of real numbers with $\bar{C}_j \equiv \sup_i |C_{i,j}| < \infty$ and $\sum_{j=0}^{\infty} j^b \bar{C}_j < \infty$ for some $b \geq 1$.
(d) $\sup_i \sigma_{i,0}^2 < \infty$, where $\sigma_{i,0}^2 = E(y_{i,0}^2)$.

Let $C_i = C_i(1)$, $\Omega_i = C_i^2$, and $\Lambda_i = \sum_{j=1}^{\infty} C_{i,0} C_{i,j}$. The terms Ω_i and Λ_i are the long-run variance and the one-sided long-run covariance of the error process $\varepsilon_{i,t}$, respectively. The next assumption is about the limits of the averages of the individual long-run variances and covariances.

Assumption 2.

(a) $\Omega = \lim_n (1/n) \sum_{i=1}^n \Omega_i$ is finite.
(b) $\Lambda = \lim_n (1/n) \sum_{i=1}^n \Lambda_i$ is finite.
(c) $\Phi = \lim_n (1/n) \sum_{i=1}^n \Omega_i^2$ is finite.

Remark. Let $\Omega_{\varepsilon_i} = E\varepsilon_{i,t}^2$. Under Assumption 2, there exist $\Omega_\varepsilon = \lim_n (1/n) \times \sum_{i=1}^n \Omega_{\varepsilon_i}$ and $\Omega_\varepsilon = \Omega - 2\Lambda$.

3. ESTIMATION OF THE LOCALIZING COEFFICIENT IN PANEL MODELS WITH DETERMINISTIC TRENDS

First, rewrite the panel model (1) in augmented regression format as

$$z_{i,t} = az_{i,t-1} + \gamma_{i,0} + \gamma_i' g_t + \varepsilon_{i,t}, \tag{8}$$

where $\gamma_{i,0} = \beta_{i,0}(1-a) + a\beta_i' \iota_p$, the deterministic trend component $\gamma_i' g_t$ is constructed as

$$\gamma_i' g_t = \beta_i'(g_t - ag_{t-1}) - a\beta_i' \iota_p = \beta_i' A_T(c) g_t,$$

$A_T(c)$ is a $p \times p$ matrix that depends upon c and T, and $\iota_p = (-1, (-1)^2, \ldots, (-1)^p)'$.

As is well known, the formulation (8) has the drawback that the regression leads to inefficient trend elimination, but it has the advantage that the detrended data are invariant to the trend parameters in (1). It will be convenient for us to work with both formulations (1) and (8), depending on the context.

3.1. Iterative Ordinary Least Squares: Biased Estimation

We start by introducing some definitions. Let

$$\tilde{g}_t = (1, g_t')', \qquad g(r) = (r, \ldots, r^p)', \qquad \tilde{g}(r) = (1, g(r)')',$$

$$D_T = \mathrm{diag}(T^{-1}, \ldots, T^{-p}), \qquad \tilde{D}_T = \mathrm{diag}(1, D_T)$$

and define

$$h_T(t,s) = (D_T g_t)' \left(\frac{1}{T} \sum_{t=1}^{T} D_T g_t g_t' D_T \right)^{-1} D_T g_s,$$

$$\tilde{h}_T(t,s) = (\tilde{D}_T \tilde{g}_t)' \left(\frac{1}{T} \sum_{t=1}^{T} \tilde{D}_T \tilde{g}_t \tilde{g}_t' \tilde{D}_T \right)^{-1} \tilde{D}_T \tilde{g}_s,$$

$$h(r,s) = g(r)' \left(\int_0^1 g(r) g(r)' dr \right)^{-1} g(s),$$

$$\tilde{h}(r,s) = \tilde{g}(r)' \left(\int_0^1 \tilde{g}(r) \tilde{g}(r)' dr \right)^{-1} \tilde{g}(s).$$

When $t = [Tr]$ and $s = [Tp]$, it is easy to see that as $T \to \infty$,

$$D_T \tilde{g}_t \to \tilde{g}(r) \quad \text{uniformly in } r \in [0,1]$$

and

$$\tilde{h}_T(t,s) \to \tilde{h}(r,p) \quad \text{uniformly in } (r,p) \in [0,1] \times [0,1].$$

Let $\underset{\sim}{z}_{i,t-1}$ and $\Delta \underset{\sim}{z}_{i,t}$ denote the ordinary least squares (OLS) detrended processes of $z_{i,t-1}$ and $\Delta z_{i,t}$, respectively, that is, for $t \geq 2$[3]

$$\underset{\sim}{z}_{i,t-1} = z_{i,t-1} - \frac{1}{T} \sum_{s=1}^{T} \tilde{h}_T(t,s) z_{i,s-1},$$

$$\Delta \underset{\sim}{z}_{i,t} = \Delta z_{i,t} - \frac{1}{T} \sum_{s=1}^{T} \tilde{h}_T(t,s) \Delta z_{i,s}.$$

Then, from model (1), we have

$$\underset{\sim}{z}_{i,t-1} = \underset{\sim}{y}_{i,t-1} - \tilde{\beta}_i' \left(\Delta \tilde{g}_t - \left(\sum_{s=1}^{T} \Delta \tilde{g}_s \tilde{g}_s' \right) \left(\sum_{t=1}^{T} \tilde{g}_t \tilde{g}_t' \right)^{-1} \tilde{g}_t \right) = \underset{\sim}{y}_{i,t-1} \quad \text{for } t \geq 2. \tag{9}$$

Also, let $\underset{\sim}{z}_{i,0} = z_{i,0} = \underset{\sim}{y}_{i,0} = y_{i,0}$.

It is well known that under Assumption 1, as $T \to \infty$,

$$\frac{1}{\sqrt{T}} y_{i,[Tr]} \Rightarrow C_i J_{c,i}(r) \tag{10}$$

(Phillips, 1987b). Using standard manipulations, it is not difficult to show that when $[Tr] = t$, as $T \to \infty$,

$$\frac{\underset{\sim}{z}_{i,t-1}}{\sqrt{T}} = \frac{\underset{\sim}{y}_{i,t-1}}{\sqrt{T}} \Rightarrow C_i \underset{\sim}{J}_{c,i}(r), \tag{11}$$

where $\underset{\sim}{J}_{c,i}(r) = J_{c,i}(r) - \int_0^1 \tilde{h}(r,s) J_{c,i}(s)\,ds$.

We now discuss an estimation procedure for the local to unity parameter with panel data when the trend coefficients β_i are unknown. Suppose that $\hat{\Lambda}_i$ are consistent estimators for Λ_i as $T \to \infty$. Consider a simple estimator, $\hat{c}^+$, defined as a serial correlation bias corrected (if required) pooled least-squares estimator $\hat{a}^+$ of a,

$$\hat{a}^+ = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 \right)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (\underset{\sim}{z}_{i,t-1} \underset{\sim}{z}_{i,t} - T\hat{\Lambda}_i), \tag{12}$$

and

$$\hat{c}^+ = T(\hat{a}^+ - 1).$$

The estimator $\hat{a}^+$ is a bias corrected[4] pooled least-squares estimator with OLS detrended data. We define $\hat{c}^+$ from $\hat{a}^+$ in view of the relation $a \simeq 1 + (c/T)$. Hereafter, we call $\hat{c}^+$ an iterative OLS estimator.

In view of (11) we have

$$T(\hat{a}^+ - a) = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T} \sum_{t=1}^{T} (\underset{\sim}{z}_{i,t-1} \underset{\sim}{\varepsilon}_{i,t} - \tilde{\Lambda}_i) \tag{13}$$

and, from the limit theory in Phillips (1987b), as $T \to \infty$ for fixed n,

$$T(\hat{a}^+ - a) \Rightarrow \left(\frac{1}{n} \sum_{i=1}^{n} \int \underset{\sim}{J}_{c,i}^2(r)\,dr \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \int \underset{\sim}{J}_{c,i}(r)\,dW_i(r) \right).$$

Note that

$$\begin{aligned} E&\left(\int_0^1 \underset{\sim}{J}_{c,i}^2(r)\,dr \right) \\ &= E\left(\int_0^1 J_{c,i}^2(r)\,dr \right) - E\left(\int_0^1 \int_0^1 J_{c,i}(r) J_{c,i}(s) \tilde{h}(r,s)\,dr\,ds \right) \\ &= \frac{-1}{2c} \left\{ 1 + \frac{1}{2c}(1 - e^{2c}) \right\} - \int_0^1 \int_0^1 e^{c(r+s)} \frac{1}{2c} (1 - e^{-2c(r \wedge s)}) \tilde{h}(r,s)\,dr\,ds \\ &= \omega_1(c), \text{ say,} \end{aligned} \tag{14}$$

and

$$E\left(\int_0^1 \underset{\sim}{J}_{c,i}(r)dW_i(r)\right)$$
$$= E\left(\int_0^1 J_{c,i}(r)dW_i(r)\right) - E\left(\int_0^1 J_{c,i}(r)dW_i(s)\tilde{h}(r,s)drds\right)$$
$$= -\int_0^1\int_0^r e^{(r-s)c}\tilde{h}(r,s)dsdr$$
$$= \omega_2(c), \text{ say.} \tag{15}$$

Because both of the i.i.d. sequences $\{\int_0^1 \underset{\sim}{J}_{c,i}^2(r)dr\}_i$ and $\{\int_0^1 \underset{\sim}{J}_{c,i}(r)dW_i(r)\}_i$ have finite second moments, it follows by the weak law of large numbers that as $n \to \infty$

$$\left(\frac{1}{n}\sum_{i=1}^n \int \underset{\sim}{J}_{c,i}^2(r)dr\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \int \underset{\sim}{J}_{c,i}(r)dW(r)\right) \to_p \frac{\omega_2(c)}{\omega_1(c)}.$$

Thus, in sequential limits as $(T, n \to \infty)_{seq}$,

$$T(\hat{a}^+ - a) \to_p \frac{\omega_2(c)}{\omega_1(c)},$$

and, in consequence,

$$\hat{c}^+ - c = T(\hat{a}^+ - a) + O\left(\frac{1}{T}\right) \to_p \frac{\omega_2(c)}{\omega_1(c)}. \tag{16}$$

Hence, the iterative OLS estimator $\hat{c}^+$ is inconsistent and has an asymptotic bias given by the ratio $[\omega_2(c)/\omega_1(c)]$ that depends on the unknown parameter c. The main reason for the inconsistency of $\hat{c}^+$ is that the detrending procedure produces a correlation between the lagged filtered regressor $\underset{\sim}{z}_{i,t-1}$ and the equation error $\underset{\sim}{\varepsilon}_{i,t}$. This correlation yields the nonvanishing limit

$$\omega_2(c) = -\int_0^1\int_0^r e^{(r-s)c}\tilde{h}(r,s)dsdr \tag{17}$$

in the numerator of $T(\hat{a}^+ - a)$.

Define

$$F(c) = c + \frac{\omega_2(c)}{\omega_1(c)}. \tag{18}$$

Because $\omega_2(c)$ is nonzero in general, $\hat{c}^+$ is not consistent. However, because the probability limit of $\hat{c}^+$, $F(c)$, depends only on c, we can expect the limit function $F(c)$ to give some information about the true parameter c, especially in regions where $F(c)$ is a monotone function. The graph of $F(c)$ is plotted for

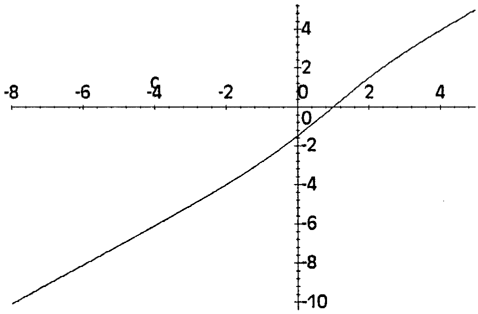

FIGURE 1. Graph of $F(c)$ when $\tilde{g}_t = 1$.

the two cases $\tilde{g}_t = 1$ (Figure 1)[5] and $\tilde{g}_t = (1, t)'$ (Figure 2), which are the most common in empirical applications.

When $\tilde{g}_t = (1, t)'$, that is, when we detrend the data to estimate c, it is apparent from Figure 2 that in the region $\{c: -0.8 \le c \le 1.2\}$ the limit function of the estimate $\hat{c}^+$ does not identify the true parameter, because $F(c)$ is not a one-to-one function in the region. Outside of this region the probability limit of the estimate $\hat{c}^+$ does identify the true value of the local to unity parameter c and can be used to construct a consistent estimate of c. Furthermore, if we assume that the true localizing parameter is nonpositive, that is, the true localizing parameter set is $\{c: c \le 0\}$, then we can identify the local to unity parameter c for all $c \le 0$ using $\hat{c}^+$ (and its probability limit) because the probability limit function $F(c)$ is monotonic with respect to c on $\{c: c \le 0\}$, the true localizing parameter set. In this case (i.e., under the assumption that $c \le 0$), there is no unidentifiable region, and $F^{-1}(\hat{c}^+)$ is a consistent estimator of c.

An analytic form of the inverse function $F^{-1}(c)$ of the probability limit function $F(c)$ is not readily available. But the function is easy to calculate numeri-

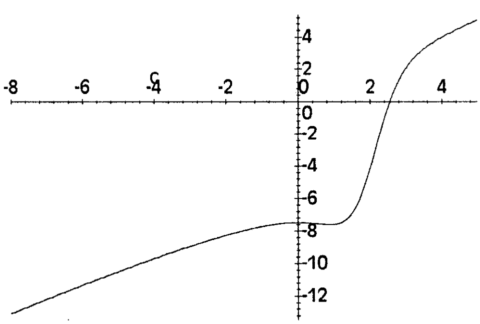

FIGURE 2. Graph of $F(c)$ when $\tilde{g}_t = (1, t)'$.

cally and is given in Table 1 for the case $\tilde{g}_t = (1, t)'$.[6] We summarize the results in the following two theorems.

THEOREM 1. *Let* $F(c) = c + [\omega_2(c)/\omega_1(c)]$. *Under Assumptions* 1–4 (*Assumptions* 3 *and* 4 *follow*), $\hat{c}^+ \to_p F(c)$ *in sequential limits as* $(T, n \to \infty)_{seq}$.

THEOREM 2. *Under Assumptions* 1–4 (*Assumptions* 3 *and* 4 *follow*), *in sequential limits as* $(T, n \to \infty)_{seq}$

$$\sqrt{n}(\hat{c}^+ - F(c)) \Rightarrow N(0, \Phi V_{\hat{c}^+}(c)),$$

where

$$V_{\hat{c}^+}(c) = \begin{pmatrix} \dfrac{-\omega_2(c)}{\omega_1(c)^2} & \dfrac{1}{\omega_1(c)} \end{pmatrix} V(c) \begin{pmatrix} \dfrac{-\omega_2(c)}{\omega_1(c)^2} \\ \dfrac{1}{\omega_1(c)} \end{pmatrix},$$

$V(c)$ *is defined in Appendix* A, *and* Φ *is defined in Assumption* 2(c).

The variance $V_{\hat{c}^+}(c)$ is a complicated function of the unknown parameter c but, again, can be calculated numerically as shown in Table 2 for $\tilde{g}_t = (1, t)'$.

Remarks.

(a) The two results are stated here in terms of $(T, n \to \infty)_{seq}$ sequential limits for the indices T and n. Appendixes C and D show that these results continue to hold when joint limits $(T, n \to \infty)$ are taken. In fact, according to the results given there, joint asymptotic normality of $\sqrt{n}(\hat{c}^+ - F(c))$ continues to hold under the additional rate restriction $(n/T) \to 0$ as $(n, T \to \infty)$, whereas joint convergence in probability, $\hat{c}^+ \to_p F(c)$ as $(n, T \to \infty)$, holds without the additional rate restriction.

(b) The intuition behind the requirement $(n/T) \to 0$ for joint asymptotic normality of $\hat{c}$ is simple. Under the assumptions in the theorem, we usually have $E(\hat{c}) \neq F(c)$ for fixed T, but $E(\hat{c}) \to F(c)$ as $T \to \infty$. In this case, the restriction $(n/T) \to 0$ works to prevent an explosive bias in $\sqrt{n}(\hat{c}^+ - F(c))$.

(c) When $E(\hat{c}) \neq F(c)$ for fixed T, which is the case under the assumptions of this paper, a limit theory based on $n \to \infty$ with T fixed encounters some additional difficulties. In the case of the probability limit of $\hat{c}^+$, when $n \to \infty$ with T fixed, we obtain a different limit from $F(c)$ and one that depends on T. Additionally, as far as the limit distribution of $\hat{c}^+$ is concerned, central limit theory as $n \to \infty$ with T fixed cannot be applied to $\sqrt{n}(\hat{c}^+ - F(c))$ but rather to the recentered estimator $\sqrt{n}(\hat{c}^+ - E(\hat{c}^+))$, which is not as useful because $E(\hat{c}^+)$ depends on additional unknown parameters.

(d) In the region where $F(c)$ is one to one, we can define a consistent estimator for c by taking the inverse value of the bias function $F(c)$, and we define $\tilde{c} = F^{-1}(\hat{c}^+)$. Then, the limit distribution of $\tilde{c}$ is found easily by the delta method. Let $b = F(c)$. Because the bias function $F(c)$ is differentiable, on the region where $F(c)$ is one to one, we have

TABLE 1. Numerical values of bias function $F(c)$ in (18) when $\tilde{g}_t = (1, t)'$

c	$F(c)$	c	$F(c)$	c	$F(c)$	c	$F(c)$
−8	−13.09	−4	−9.71	0	−7.5	4	3.97
−7.9	−13	−3.9	−9.63	0.1	−7.5	4.1	4.09
−7.8	−12.91	−3.8	−9.55	0.2	−7.51	4.2	4.2
−7.7	−12.82	−3.7	−9.48	0.3	−7.52	4.3	4.31
−7.6	−12.73	−3.6	−9.4	0.4	−7.54	4.4	4.42
−7.5	−12.64	−3.5	−9.33	0.5	−7.56	4.5	4.52
−7.4	−12.56	−3.4	−9.25	0.6	−7.58	4.6	4.63
−7.3	−12.47	−3.3	−9.18	0.7	−7.6	4.7	4.73
−7.2	−12.38	−3.2	−9.1	0.8	−7.62	4.8	4.83
−7.1	−12.29	−3.1	−9.03	0.9	−7.63	4.9	4.93
−7	−12.2	−3	−8.96	1	−7.61	5	5.03
−6.9	−12.12	−2.9	−8.88	1.1	−7.58	5.1	5.13
−6.8	−12.03	−2.8	−8.81	1.2	−7.51	5.2	5.23
−6.7	−11.94	−2.7	−8.74	1.3	−7.38	5.3	5.33
−6.6	−11.86	−2.6	−8.67	1.4	−7.2	5.4	5.43
−6.5	−11.77	−2.5	−8.6	1.5	−6.93	5.5	5.52
−6.4	−11.68	−2.4	−8.54	1.6	−6.57	5.6	5.62
−6.3	−11.6	−2.3	−8.47	1.7	−6.1	5.7	5.72
−6.2	−11.51	−2.2	−8.4	1.8	−5.54	5.8	5.82
−6.1	−11.43	−2.1	−8.34	1.9	−4.88	5.9	5.92
−6	−11.34	−2	−8.28	2	−4.14	6	6.02
−5.9	−11.26	−1.9	−8.21	2.1	−3.36	6.1	6.12
−5.8	−11.17	−1.8	−8.15	2.2	−2.56	6.2	6.21
−5.7	−11.09	−1.7	−8.09	2.3	−1.77	6.3	6.31
−5.6	−11	−1.6	−8.04	2.4	−1.03	6.4	6.41
−5.5	−10.92	−1.5	−7.98	2.5	−0.34	6.5	6.51
−5.4	−10.84	−1.4	−7.93	2.6	0.27	6.6	6.61
−5.3	−10.75	−1.3	−7.88	2.7	0.82	6.7	6.71
−5.2	−10.67	−1.2	−7.83	2.8	1.29	6.8	6.81
−5.1	−10.59	−1.1	−7.78	2.9	1.7	6.9	6.91
−5	−10.51	−1	−7.74	3	2.06	7	7.01
−4.9	−10.43	−0.9	−7.69	3.1	2.36	7.1	7.11
−4.8	−10.34	−0.8	−7.66	3.2	2.63	7.2	7.21
−4.7	−10.26	−0.7	−7.62	3.3	2.86	7.3	7.31
−4.6	−10.18	−0.6	−7.59	3.4	3.07	7.4	7.41
−4.5	−10.1	−0.5	−7.56	3.5	3.25	7.5	7.5
−4.4	−10.02	−0.4	−7.54	3.6	3.42	7.6	7.6
−4.3	−9.94	−0.3	−7.52	3.7	3.57	7.7	7.7
−4.2	−9.86	−0.2	−7.51	3.8	3.71	7.8	7.8
−4.1	−9.79	−0.1	−7.5	3.9	3.84	7.9	7.9

TABLE 2. Numerical values of asymptotic standard error $\sqrt{V_{\hat{c}^+}(c)}$ of iterative OLS estimator when $\tilde{g}_t' = (1t)'$

c	$\sqrt{V_{\hat{c}^+}(c)}$	c	$\sqrt{V_{\hat{c}^+}(c)}$	c	$\sqrt{V_{\hat{c}^+}(c)}$	c	$\sqrt{V_{\hat{c}^+}(c)}$	c	$\sqrt{V_{\hat{c}^+}(c)}$
−8	6.0079	−4.6	5.5106	−1.2	5.1166	2.2	4.9802	5.6	0.0864
−7.9	5.9932	−4.5	5.4965	−1.1	5.1097	2.3	4.6617	5.7	0.0782
−7.8	5.9785	−4.4	5.4824	−1	5.1032	2.4	4.2764	5.8	0.0709
−7.7	5.9638	−4.3	5.4683	−0.9	5.0971	2.5	3.8520	5.9	0.0642
−7.6	5.9491	−4.2	5.4544	−0.8	5.0913	2.6	3.4167	6	0.0582
−7.5	5.9344	−4.1	5.4405	−0.7	5.0858	2.7	2.9939	6.1	0.0528
−7.4	5.9197	−4	5.4267	−0.6	5.0806	2.8	2.6001	6.2	0.0479
−7.3	5.9050	−3.9	5.4131	−0.5	5.0758	2.9	2.2447	6.3	0.0435
−7.2	5.8902	−3.8	5.3995	−0.4	5.0711	3	1.9316	6.4	0.0395
−7.1	5.8755	−3.7	5.3860	−0.3	5.0667	3.1	1.6603	6.5	0.0358
−7	5.8608	−3.6	5.3727	−0.2	5.0624	3.2	1.4281	6.6	0.0325
−6.9	5.8460	−3.5	5.3595	−0.1	5.0583	3.3	1.2307	6.7	0.0295
−6.8	5.8313	−3.4	5.3465	0	5.0540	3.4	1.0637	6.8	0.0268
−6.7	5.8166	−3.3	5.3336	0.1	5.0503	3.5	0.9226	6.9	0.0244
−6.6	5.8019	−3.2	5.3208	0.2	5.0463	3.6	0.8032	7	0.0221
−6.5	5.7871	−3.1	5.3082	0.3	5.0424	3.7	0.7021	7.1	0.0201
−6.4	5.7724	−3	5.2958	0.4	5.0386	3.8	0.6160	7.2	0.0183
−6.3	5.7577	−2.9	5.2836	0.5	5.0351	3.9	0.5425	7.3	0.0166
−6.2	5.7430	−2.8	5.2716	0.6	5.0321	4	0.4794	7.4	0.0151
−6.1	5.7283	−2.7	5.2598	0.7	5.0301	4.1	0.4250	7.5	0.0137
−6	5.7136	−2.6	5.2483	0.8	5.0299	4.2	0.3779	7.6	0.0125
−5.9	5.6989	−2.5	5.2369	0.9	5.0323	4.3	0.3368	7.7	0.0114
−5.8	5.6843	−2.4	5.2258	1	5.0387	4.4	0.3009	7.8	0.0103
−5.7	5.6696	−2.3	5.2150	1.1	5.0510	4.5	0.2694	7.9	0.0094
−5.6	5.6550	−2.2	5.2045	1.2	5.0711	4.6	0.2416	8	0.0085
−5.5	5.6404	−2.1	5.1942	1.3	5.1016	4.7	0.2170		
−5.4	5.6258	−2	5.1843	1.4	5.1444	4.8	0.1951		
−5.3	5.6113	−1.9	5.1746	1.5	5.2000	4.9	0.1757		
−5.2	5.5968	−1.8	5.1653	1.6	5.2658	5	0.1584		
−5.1	5.5823	−1.7	5.1563	1.7	5.3330	5.1	0.1429		
−5	5.5679	−1.6	5.1477	1.8	5.3853	5.2	0.1290		
−4.9	5.5535	−1.5	5.1394	1.9	5.3991	5.3	0.1166		
−4.8	5.5392	−1.4	5.1314	2	5.3481	5.4	0.1054		
−4.7	5.5249	−1.3	5.1238	2.1	5.2114	5.5	0.0954		

Note: The numerical values are obtained by 10,000 iterations of the simulation with size 1,000 data.

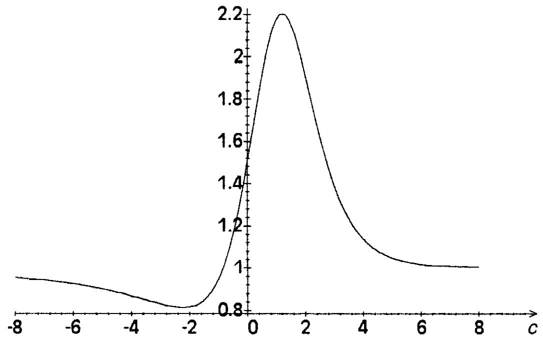

FIGURE 3. Graph of $dF(c)/dc$ when $\tilde{g}_t = 1$.

$$\frac{dc}{db} = \frac{dF^{-1}(b)}{db} = \left(1 + \frac{d}{dc}\left(\frac{\omega_2(c)}{\omega_1(c)}\right)\right)^{-1},$$

where $c = F^{-1}(b)$, and $[dF^{-1}(b)/db]$ is well defined on the region $\{b = F(c) : [dF(c)/dc] \neq 0\}$. If $b = F^{-1}(c)$ and $(dc/db) = [dF^{-1}(b)/db]$ are well defined, then by the delta method, we have

$$\sqrt{n}(\tilde{c} - c) = \sqrt{n}(F^{-1}(\hat{c}^+) - F^{-1}(F(c)))$$
$$\Rightarrow N\left(0, \left(\frac{dF^{-1}(b)}{db}\right)^2 V_{\hat{c}^+}(c)\right), \tag{19}$$

where $b = F(c)$.

(e) In Figures 3 and 4 we plot the graphs of $[dF(c)/dc]$ when $\tilde{g}_t = 1$ and $\tilde{g}_t = (1, t)'$. When $\tilde{g}_t = 1$, $[dF(c)/dc] \neq 0$. However, when $\tilde{g}_t = (1, t)'$, $[dF(c)/dc] = 0$ at two points, $c = 0$ and $c \simeq 0.895$, and at these points the derivative $[dF^{-1}(c)/dc]$ is not defined.

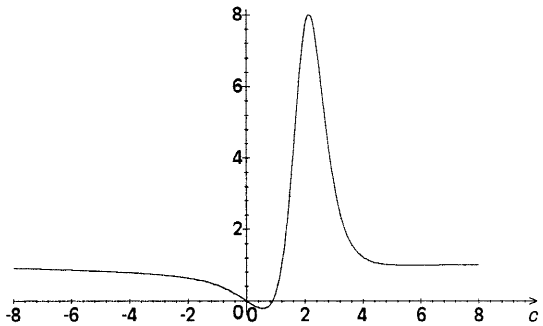

FIGURE 4. Graph of $dF(c)/dc$ when $\tilde{g}_t = (1, t)'$.

Consistent estimators $\hat{\Lambda}_i$ and $\hat{\Omega}_i$ for the individual long-run variances Λ_i and Ω_i can be obtained by employing standard kernel estimates. These estimates can then be averaged by produce consistent estimates of the quantities Λ, Φ, and Ω. More specifically, let $\tilde{a}$ be the pooled least-squares estimator of the regression model (8), that is,

$$\tilde{a} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2\right)^{-1} \sum_{i=1}^{n}\sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}\underset{\sim}{z}_{i,t},$$

and $\hat{\varepsilon}_{i,t}$ be the residual $\hat{\varepsilon}_{i,t} = \underset{\sim}{z}_{i,t} - \tilde{a}\underset{\sim}{z}_{i,t-1}$ from this regression. Define the sample covariances $\hat{\Gamma}_i(j) = (1/T)\sum \hat{\varepsilon}_{i,t}\hat{\varepsilon}_{i,t+j}$, where the summation is defined over $1 \leq t, t+j \leq T$. Then, the kernel estimators for $\hat{\Lambda}_i$ and $\hat{\Omega}_i$ are

$$\hat{\Lambda}_i = \sum_{j=1}^{T} w\left(\frac{j}{K}\right)\hat{\Gamma}_i(j), \tag{20}$$

$$\hat{\Omega}_i = \sum_{j=-T}^{T} w\left(\frac{j}{K}\right)\hat{\Gamma}_i(j), \tag{21}$$

where $w(\cdot)$ is a kernel function and K is a lag truncation parameter. Truncation occurs when $w(j/K) = 0$ for $|j| \geq K$. Averaging over cross section observations now leads to consistent estimators of Λ, Φ, and Ω. The following assumptions concern the class of admissible kernels and the choice of the bandwidth to be employed in the kernel estimates (20) and (21). These assumptions are used in our joint convergence arguments in the Appendixes, where it is shown that $(1/n)\sum_{i=1}^{n}\hat{\Lambda}_i \rightarrow_p \Lambda$ as $(T,n) \rightarrow \infty$. For sequential limits, it is possible to use weaker conditions.

Assumption 3 (Kernel Condition). The kernel function $w(\cdot): \mathbb{R} \rightarrow [-1,1]$ satisfies the following:

(a) $w(0) = 1$, $w(x) = w(-x)$, $\int_{-1}^{1} w(x)^2 dx < \infty$, and $w(\cdot)$ is continuous at zero and all but a finite number of other points.
(b) $w(x) = 0$, $|x| \geq 1$.
(c) $w_q = \lim_{x\rightarrow 0}[1 - w(x)/|x|^q]$ is finite for some $q \in (\frac{1}{2},\infty)$.

Assumption 4 (Bandwidth Condition). We assume that, as $T \rightarrow \infty$, the bandwidth parameter satisfies $K \rightarrow \infty$, $(K^2/T) \rightarrow 0$, and $(K^{2q+1}/T) \rightarrow \gamma > 0$ for some $\frac{1}{2} < q \leq b$ for which w_q is finite, where b is given in condition (c) in Assumption 1.

The Parzen exponent q in Assumption 3 is related to the smoothness of the kernel at zero. The most frequently used kernels in applications satisfy this assumption—see, for example, Andrews (1991) for details.

Remarks.

(a) The iterated OLS estimator discussed previously is a pooled least-squares estimator based on OLS detrended data. Naturally, there are many pooled least-squares estimators based on data that have been detrended in different ways. One procedure that is used widely in applications is to use first differenced data. This detrending procedure has difficulties similar to those of iterated OLS. To be specific, assume a simple linear trend in the panel model (1), so that

$$z_{i,t} = \beta_i t + y_{i,t},$$

$$y_{i,t} = a y_{i,t-1} + \varepsilon_{i,t}, \qquad a = \exp\left(\frac{c}{T}\right). \tag{22}$$

The difference detrended data are then simply

$$\underline{z}_{i,t} = z_{i,t} - \bar{\beta}_i t,$$

where $\bar{\beta}_{i,1} = (1/(T-1))\sum_{t=2}^{T} \Delta z_{i,t}$. Define

$$\bar{c}^+ = T\left(\hat{\Omega}_\varepsilon \sum_{i=1}^{n}\sum_{t=2}^{T} \underline{z}_{i,t-1}^2\right)^{-1}\left(\hat{\Omega} \sum_{i=1}^{n}\sum_{t=2}^{T} \Delta\underline{z}_{i,t}\,\underline{z}_{i,t-1}\right),$$

where $\hat{\Omega}_\varepsilon = (1/n)\sum_{i=1}^{n}(1/T)\sum_{t=2}^{T}(\Delta\underline{z}_{i,t})^2$. In this case, applying similar arguments to those used earlier in this section, we find

$$\bar{c}^+ \to_p \frac{1}{-2\omega_3(c)}, \tag{23}$$

where $\omega_3(c) = \int_0^1 \{(1/2c)(e^{2rc}-1) - 2r((1/2c)(1-e^{-2cr})e^{c(1+r)}) + r^2(1/2c)(e^{2c}-1)\}dr$. From this outcome, it is apparent that the probability limit of $\bar{c}^+$, $[1/-2\omega_3(c)]$, is different from c in general and therefore the estimator $\bar{c}^+$, like $\hat{c}^+$, is not consistent. (More details on this estimation procedure are given in the previous version of this paper, Moon and Phillips, 1999a.)

(b) The asymptotic bias in iterative OLS estimation arises because of the correlation between the detrended regressors and the regression errors. The usual econometric approach to the consistent estimation of regression coefficients when there is correlation between the regressors and the errors is instrumental variables. In the present case, an instrumental variable procedure is possible in which backward-recursive detrended data are used to produce an instrumental variable for the regressor in a forward-recursive detrended regression model. To explain this idea, take the regression model (8) and consider the following two recursive detrending procedures. First, detrend the data recursively through $t = t_0, \ldots, T$, starting at some observation $t_0 > p$, where $p = \dim(g_t)$, and calculate the backward-detrended data

$$\underline{z}_{i,t} = z_{i,t} - \tilde{g}_t'\left(\sum_{s=1}^{t} \tilde{g}_s \tilde{g}_s'\right)^{-1}\left(\sum_{s=1}^{t} \tilde{g}_s z_{i,s}\right).$$

Similarly, for $t = 1, \ldots, T - t_1$, we have the forward-detrended data as follows:

$$\bar{z}_{i,t} = z_{i,t} - \tilde{g}_t' \left(\sum_{s=t+1}^{T} \tilde{g}_s \tilde{g}_s' \right)^{-1} \left(\sum_{s=t+1}^{T} \tilde{g}_s z_{i,s} \right).$$

Then, we employ the forward-detrending procedure in the regression equation (8) and have

$$\bar{z}_{i,t} = a\bar{z}_{i,t-1} + \bar{\varepsilon}_{i,t}.$$

Now, using the backward-detrended data as an instrument, we construct the following instrumental variable (IV) estimator:

$$\hat{c}_{IV}^{+} = T(\hat{a}_{IV}^{+} - 1),$$

where

$$\hat{a}_{IV}^{+} = \left(\sum_{i=1}^{n} \sum_{t=t_0}^{T-t_1} \bar{z}_{i,t-1} \underline{z}_{i,t-1} \right)^{-1} \left(\sum_{i=1}^{n} \left(\sum_{t=t_0}^{T-t_1} \bar{z}_{i,t} \underline{z}_{i,t-1} - \hat{\Lambda}_i \right) \right).$$

The forward-recursive detrended data use future information in detrending, whereas the backward-recursive detrended data use past information in detrending. Thus, we might expect that the forward-recursive detrended error $\bar{\varepsilon}_{i,t}$ in the numerator of $\hat{a}_{IV}^{+} - 1$ might be asymptotically uncorrelated with the backward-detrended regressor $\underline{z}_{i,t-1}$. In the earlier version of the paper (Moon and Phillips, 1999a), we showed that the IV estimator $\hat{c}_{IV}^{+}$ is consistent for almost all the values for c. However, it turns out that $\hat{c}_{IV}^{+}$ also has a problem that the numerator of $\hat{c}_{IV}^{+}$ is not always nonzero. In particular, when $c = 0$, the limit of the denominator of $\hat{c}_{IV}^{+}$ degenerates to zero in probability, and so the IV estimator $\hat{c}_{IV}^{+}$ is not consistent in this case. Resolving the bias problem that arises in the numerator yields a degeneracy problem in the denominator for some values of c, and in particular at $c = 0$. In effect, there is insufficient information (in terms of persistent excitation in the regressor/instrument) about the true value $c = 0$ to deliver a consistent estimate for this value of c.

3.2. Double Bias Corrected Estimation

The iterative OLS estimator has an asymptotic bias that depends only on the unknown localizing parameter c. The idea behind the method we investigate here is to adjust for the bias that arises from the correlation of the filtered data and the regression error. In particular, we use a linear representation of the exponential term that appears in the bias producing element (17), so that the estimator of c can be adjusted directly to take the bias in OLS regression into account.

First, notice that when c is close to c_0, we can approximate the asymptotic bias, $\omega_2(c)$, of the numerator of $\hat{c}^+ - c$ by

$$\omega_2(c) \simeq -\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}\left[e^{\left(\frac{t}{T}-\frac{s}{T}\right)c_0} + \left(\frac{t}{T}-\frac{s}{T}\right)e^{\left(\frac{t}{T}-\frac{s}{T}\right)c_0}(c-c_0)\right]\tilde{h}_T(t,s).$$

If there exists a consistent estimate, say, $\tilde{c}$, for c, then, we may further approximate $\omega_2(c)$ by

$$\hat{\omega}_2(c) \simeq -\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}\left[e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}} + \left(\frac{t}{T}-\frac{s}{T}\right)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}(c-\tilde{c})\right]\tilde{h}_T(t,s).$$

Because this approximation to the bias $\omega_2(c)$ is linear in c, it is possible to adjust the estimator $\hat{c}^+$ to take the bias information into account. The adjustment is designed so that the new estimator, $\hat{c}^{++}$, satisfies the system

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2\right)(\hat{c}^{++}-c) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\Bigg[\frac{1}{T}\sum_{t=1}^{T}(\underset{\sim}{z}_{i,t-1}\varepsilon_{i,t}-\hat{\Lambda}_i) \\
&\qquad + \Omega_i\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}\left(e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}} + \left(\frac{t}{T}-\frac{s}{T}\right)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}(\hat{c}^{++}-\tilde{c})\right)\right. \\
&\qquad \left.\times\, \tilde{h}_T(t,s)\right)\Bigg] + o_p(1).
\end{aligned}$$

Then,

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2\right)(\hat{c}^{++}-c) \\
&\quad - \left(\frac{1}{n}\sum_{i=1}^{n}\Omega_i\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}\left[\left(\frac{t}{T}-\frac{s}{T}\right)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\right]\tilde{h}_T(t,s)\right)(\hat{c}^{++}-c) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\Bigg[\frac{1}{T}\sum_{t=1}^{T}(\underset{\sim}{z}_{i,t-1}\varepsilon_{i,t}-\hat{\Lambda}_i) + \Omega_i\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}e^{\left(\frac{t}{T}-\frac{s}{T}\right)c}\tilde{h}_T(t,s)\right) \\
&\qquad - \Omega_i(\tilde{c}-c)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}\left[\left(\frac{t}{T}-\frac{s}{T}\right)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\right]\tilde{h}_T(t,s)\Bigg] + o_p(1).
\end{aligned}$$

(24)

Because $\tilde{c}$ is consistent for c, the third term on the right hand side of (24) vanishes. The other two terms on the right hand side of (24) also converge in probability to zero because as $T \to \infty$ for fixed n, they converge in distribution to

$$\frac{1}{n}\sum_{i=1}^{n}\Omega_i\left(\int_0^1 \underset{\sim}{J}_{c,i}\,dW_i - \omega_2(c)\right),$$

which converges in probability to zero as $n \to \infty$ because $E(\int_0^1 \underset{\sim}{J}_{c,i}\,dW_i) = \omega_2(c)$ and $(1/n)\sum_{i=1}^{n}\Omega_i \to \Omega$.

To implement the idea in (24), we use the consistent estimator $\tilde{c} = F^{-1}(\hat{c}^{+})$ defined in Section 2, assuming that there are no problems of identification.[7] Then, the preceding heuristic analysis leads to the following panel estimator for the local to unity parameter c

$$\hat{c}^{++} = \left[\sum_{i=1}^{n}\left\{\frac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2 - \hat{\Omega}_i\frac{1}{T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\left(\frac{t}{T}-\frac{s}{T}\right)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\tilde{h}_T(t,s)\right)\right\}\right]^{-1}$$
$$\times \sum_{i=1}^{n}\left\{\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}\Delta\underset{\sim}{z}_{i,t} - \hat{\Lambda}_i + \hat{\Omega}_i\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}\left[1-\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}\right]\right.$$
$$\left.\times\, e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\tilde{h}_T(t,s)\right\}. \quad \textbf{(25)}$$

The inclusion of $\hat{\Lambda}_i$ in the formulation of $\hat{c}^{++}$ provides the usual serial correlation bias correction. The adjustment of the numerator and the denominator by

$$\hat{\Omega}_i\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}\left[1-\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}\right]e^{\left(\frac{t}{T}-\frac{s}{T}\right)\hat{c}^{+}}\tilde{h}_T(t,s)$$

and

$$\hat{\Omega}_i\frac{1}{T^3}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}(t-s)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\tilde{h}_T(t,s)\right)$$

corrects for the bias from the use of detrended data.

From the definition of $\hat{c}^{++}$, we deduce that

$$\hat{c}^{++} = c + \left[\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2 - \hat{\Omega}_i\frac{1}{T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\left(\frac{t}{T}-\frac{s}{T}\right)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\tilde{h}_T(t,s)\right)\right\}\right]^{-1}$$
$$\times\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t} - \hat{\Lambda}_i + \hat{\Omega}_i\frac{1}{T^2}\sum_{t=1}^{T}\left[1-\left(\frac{t}{T}-\frac{s}{T}\right)(\tilde{c}-c)\right]\right.$$
$$\left.\times\, e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\tilde{h}_T(t,s) + r_T\frac{1}{T^3}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2\right\},$$

where $r_T = T^2(\exp(c/T) - (1 + c/T))$ and equality holds because $\underset{\sim}{z}_{i,t} = a\underset{\sim}{z}_{i,t-1} + \underset{\sim}{\varepsilon}_{i,t}$.

To derive the probability limit of $\hat{c}^{++}$ under sequential limits, we first let $T \to \infty$ for fixed n and then let $n \to \infty$. Because $\tilde{c}$ is consistent for c, it follows that

$$\hat{c}^{++} - c = \left[\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{T^2} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 - \hat{\Omega}_i \frac{1}{T^2} \left(\sum_{t=1}^{T} \sum_{s=1}^{t} \left(\frac{t}{T} - \frac{s}{T} \right) e^{\left(\frac{t}{T} - \frac{s}{T}\right)c} \tilde{h}_T(t,s) \right) \right\} \right]^{-1}$$
$$\times \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{T} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1} \varepsilon_{i,t} - \hat{\Lambda}_i + \hat{\Omega}_i \frac{1}{T^2} \sum_{t=1}^{T} e^{\left(\frac{t}{T} - \frac{s}{T}\right)c} \tilde{h}_T(t,s) + r_T \frac{1}{T^3} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 \right\} + o_p(1).$$

When $T \to \infty$ with fixed n, by the continuous mapping theorem and cross section independence, we have

$$\left[\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{T^2} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 - \hat{\Omega}_i \frac{1}{T^2} \left(\sum_{t=1}^{T} \sum_{s=1}^{t} \left(\frac{t}{T} - \frac{s}{T} \right) e^{\left(\frac{t}{T} - \frac{s}{T}\right)c} \tilde{h}_T(t,s) \right) \right\} \right]^{-1}$$
$$\times \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{T} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1} \varepsilon_{i,t} - \hat{\Lambda}_i + \hat{\Omega}_i \frac{1}{T^2} \sum_{t=1}^{T} e^{\left(\frac{t}{T} - \frac{s}{T}\right)c} \tilde{h}_T(t,s) + r_T \frac{1}{T^3} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 \right\}$$
$$\Rightarrow \left[\frac{1}{n} \sum_{i=1}^{n} \Omega_i \left(\int_0^1 \underset{\sim}{J}_{c,i}^2(r) dr - \int_0^1 \int_0^t (r-s) \tilde{h}(r,s) ds dr \right) \right]^{-1}$$
$$\times \frac{1}{n} \sum_{i=1}^{n} \Omega_i \left(\int_0^1 \underset{\sim}{J}_{c,i}(r) dW_i(r) + \int_0^1 \int_0^r e^{(r-s)c} \tilde{h}(r,s) ds dr \right). \quad \textbf{(26)}$$

In view of (14), we have

$$E\left(\int_0^1 \underset{\sim}{J}_{c,i}^2(r) dr - \int_0^1 \int_0^r (r-s) \tilde{h}(r,s) ds dr \right)$$
$$= \omega_1(c) - \int_0^1 \int_0^r (r-s) \tilde{h}(r,s) ds dr = \omega(c), \text{ say,}$$

where

$$\omega_1(c) = \frac{-1}{2c} \left\{ 1 + \frac{1}{2c} (1 - e^{2c}) \right\} - \int_0^1 \int_0^1 e^{c(r+s)} \frac{1}{2c} (1 - e^{-2c(r \wedge s)}) \tilde{h}(r,s) dr ds.$$

We know that $\int_0^1 \underset{\sim}{J}^2(r) dr$ has finite second moments. Also, it is assumed that $\sup_i |C_i| = \bar{C} < \infty$ so $\sup_i |\Omega_i| = \bar{C}^2 < \infty$. Then, by the weak law of large numbers, as $n \to \infty$

$$\left(\frac{1}{n} \sum_{i=1}^{n} \Omega_i \left(\int_0^1 \underset{\sim}{J}_{c,i}^2(r) dr - \int_0^1 \int_0^r (r-s) \tilde{h}(r,s) ds dr \right) \right) \to_p \Omega \omega(c). \quad \textbf{(27)}$$

For the time being, assume that $\omega(c) \neq 0$ at the true value of c.

Similarly, in view of (15) we have

$$E\left(\int \underset{\sim}{J}_{c,i}\, dW_i\right) + \int_0^1 \int_0^r e^{(r-s)c}\tilde{h}(r,s)\, ds dr = 0.$$

Because $\lim_n (1/n)\sum_{i=1}^n \Omega_i = \Omega$ and $\lim_n (1/n)\sum_{i=1}^n \Lambda_i = \Lambda$, and using the weak law again, as $n \to \infty$, we have

$$\left(\frac{1}{n}\sum_{i=1}^n \left(\Omega_i \int \underset{\sim}{J}_{c,i}(r)\, dW_i + \int_0^1 \int_0^r e^{(r-s)c}\tilde{h}(r,s)\, ds dr\right)\right) \to_p 0. \quad \textbf{(28)}$$

Combining (27) and (28), and provided $\omega(c) \neq 0$, we then have under sequential limits as $(T, n \to \infty)_{seq}$

$$\hat{c}^{++} \to_p c. \quad \textbf{(29)}$$

In summary, we have the following result for the consistency of $\hat{c}^{++}$ under sequential limits. Appendix C extends this result to give consistency of $\hat{c}^{++}$ under joint limits.

THEOREM 3. *Under Assumptions* 1 *and* 2, *if* $\omega(c) \neq 0$ *and if* $\tilde{c}$ *is consistent for* c, *then as* $(T, n \to \infty)_{seq}$, $\hat{c}^{++} \to_p c$.

Remarks.

(a) The consistency of $\hat{c}^{++}$ in the preceding theorem holds only for values of c such that $\omega(c) \neq 0$. In general, $\omega(c)$ is quite a complicated function of c and is dependent on the explicit form of the deterministic trends in the model. Consequently, it is hard to find analytically the set of c such that $\omega(c) = 0$. Figures 5 and 6 plot the graphs of $\omega(c)$ for the most commonly used trends $\tilde{g}_t = 1$ and $\tilde{g}_t = (1, t)'$.

(b) These graphs show three important features of $\omega(c)$. First, we see that $\omega(c) \neq 0$ when $c < 0$; second, $\omega(c) = 0$ when $c = 0$; and, third, there is another point of c for which $\omega(c) = 0$ in the region $c > 0$ when $\tilde{g}_t = (1, t)'$.

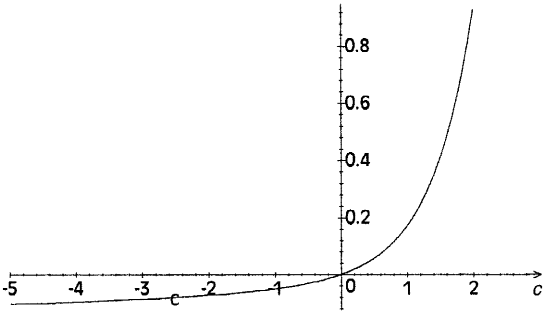

FIGURE 5. Graph of $\omega(c)$ when $\tilde{g}_t = 1$.

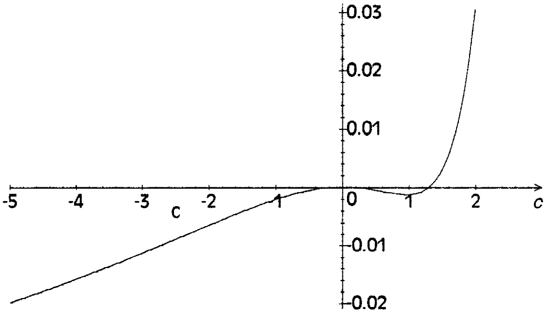

FIGURE 6. Graph of $\omega(c)$ when $\tilde{g}_t = (1, t)'$.

(c) Unfortunately, at $c = 0$, $\omega(c)$ is always zero regardless of the form of the deterministic trends assumed in the model. This can be verified by the following simple calculation for the general case. We have

$$\omega(0) = E\left(\int_0^1 \underset{\sim}{W}_i^2(r)dr - \int_0^1 \int_0^r (r-s)\tilde{h}(r,s)dsdr \right)$$

$$= \int_0^1 rdr - \int_0^1 \int_0^1 (r \wedge s)\tilde{h}(r,s)dsdr - \int_0^1 \int_0^r (r-s)\tilde{h}(r,s)dsdr$$

$$= \int_0^1 rdr - \int_0^1 \int_r^1 r\tilde{h}(r,s)dsdr - \int_0^1 \int_0^r r\tilde{h}(r,s)dsdr$$

$$= \int_0^1 rdr - \int_0^1 \int_0^1 r\tilde{h}(r,s)drds = \int_0^1 rdr - \int_0^1 sds = 0,$$

where the last line holds for the following reason. Let $L_2[0,1]$ be a space of square integrable functions on $[0,1]$ with inner product $\langle f, g \rangle = \int_0^1 f(r)g(r)dr$. Let $\mathcal{Q}$ denote a space of polynomial functions of degree p on $[0,1]$ generated by $\{1, r, \ldots, r^p\}$. Let $\tilde{g}(r) = (1, r, \ldots, r^p)'$. Then, the operator $\mathcal{P}$ from $L_2[0,1]$ to $\mathcal{Q}$ defined as $\mathcal{P}(f) = \tilde{g}(r)'(\int_0^1 \tilde{g}(r)\tilde{g}(r)'dr)(\int_0^1 \tilde{g}(s)f(s)ds)$ is a projection. Hence, when $f(r) = r$, $\mathcal{P}(f(r)) = f(r) = r$, and so we have $\int_0^1 r\tilde{h}(r,s)dr = s$.

(d) As Figures 5 and 6 show, even though $\omega(c) \neq 0$ for $c < 0$, $\omega(c)$ is very close to zero around $c = 0$. Because of this, we can expect that the estimator $\hat{c}^{++}$ may perform poorly for $c \sim 0$.

Next we derive the limit distribution of $\hat{c}^{++}$ using sequential limit arguments. Here we assume that c satisfies $\omega(c) \neq 0$, $F^{-1}(c)$ is well defined, and $[dF(c)/dc] \neq 0$. In this case, $\tilde{c} = F^{-1}(\hat{c}^+)$ is $\sqrt{n}$ consistent and $\sqrt{n}(\tilde{c} - c)$ is stochastically bounded (see (19)).

First, standardizing $\hat{c}^{++} - c$ by $\sqrt{n}$, we write

$$\begin{aligned}
\sqrt{n}(\hat{c}^{++} - c) &= \left[\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{T^2}\sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 - \hat{\Omega}_i \frac{1}{T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}(t-s)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\tilde{h}_T(t,s)\right)\right\}\right]^{-1} \\
&\quad \times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{1}{T}\sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t} - \hat{\Lambda}_t + \hat{\Omega}_i \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t}\left(1 - \left(\frac{t}{T} - \frac{s}{T}\right)(\tilde{c} - c)\right)\right. \\
&\qquad \left. \times e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\tilde{h}_T(t,s) + r_T \frac{1}{T^3}\sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2\right\}.
\end{aligned}$$

Because $\tilde{c}$ is consistent for c, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_i \frac{1}{T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}(t-s)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\tilde{h}_T(t,s)\right) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_i \frac{1}{T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}(t-s)e^{\left(\frac{t}{T}-\frac{s}{T}\right)c}\tilde{h}_T(t,s)\right) + o_p(1). \qquad \textbf{(30)}
\end{aligned}$$

Next, by the mean value theorem we write

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\Omega}_i \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t}\left(1 - \left(\frac{t}{T} - \frac{s}{T}\right)(\tilde{c} - c)\right)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\tilde{h}_T(t,s) \\
&\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\Omega}_i \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t} e^{\left(\frac{t}{T}-\frac{s}{T}\right)c}\tilde{h}_T(t,s) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\Omega}_i \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t}\left(e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}} - e^{\left(\frac{t}{T}-\frac{s}{T}\right)c}\right)\tilde{h}_T(t,s) \\
&\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\Omega}_i\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t}\left(\frac{t}{T} - \frac{s}{T}\right)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\tilde{h}_T(t,s)\right\}(\tilde{c} - c) \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_i\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t}\left(\frac{t}{T} - \frac{s}{T}\right)\left(e^{\left(\frac{t}{T}-\frac{s}{T}\right)c^*} - e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\right)\tilde{h}_T(t,s)\right) \\
&\quad \times \sqrt{n}(\tilde{c} - c), \qquad \textbf{(31)}
\end{aligned}$$

where c^* is located between c and $\tilde{c}$. Because c^* converges in probability to c and $\sqrt{n}(\tilde{c} - c)$ is stochastically bounded (see (19)), it is easy to see that

$$(31) = o_p(1). \qquad \textbf{(32)}$$

So, in view of (30)–(32), we now have

$$\sqrt{n}(\hat{c}^{++}-c)$$
$$=\left[\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2-\hat{\Omega}_i\frac{1}{T^2}\left(\sum_{t=1}^{T}\sum_{s=1}^{t}(t-s)e^{\left(\frac{t}{T}-\frac{s}{T}\right)c}\tilde{h}_T(t,s)\right)\right\}\right]^{-1}$$
$$\times\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t}-\hat{\Lambda}_i+\hat{\Omega}_i\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t}e^{\left(\frac{t}{T}-\frac{s}{T}\right)c}\tilde{h}_T(t,s)\right\}$$
$$+o_p(1).$$

For fixed n, as $T\to\infty$, the main term of $\sqrt{n}(\hat{c}^{++}-c)$ converges in distribution to

$$\left[\frac{1}{n}\sum_{i=1}^{n}\Omega_i\left(\int_0^1\underset{\sim}{J}_{c,i}^2(r)dr-\int_0^1\int_0^r(r-s)\tilde{h}(r,s)dsdr\right)\right]^{-1}$$
$$\times\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Omega_i\left(\int_0^1\underset{\sim}{J}_{c,i}(r)dW_i(r)+\int_0^1\int_0^r e^{(r-s)c}\tilde{h}(r,s)dsdr\right). \quad \textbf{(33)}$$

As shown in the previous section.

$$E\left[\Omega_i\left(\int_0^1\underset{\sim}{J}_{c,i}(r)dW_i(r)+\int_0^1\int_0^r e^{(r-s)c}\tilde{h}(r,s)dsdr\right)\right]=0.$$

Appendix A derives the variance of the numerator in (33).[8] It is

$$E\left[\Omega_i\left(\int_0^1\underset{\sim}{J}_{c,i}(r)dW_i(r)+\int_0^1\int_0^r e^{(r-s)c}\tilde{h}(r,s)dsdr\right)\right]^2$$
$$=\Omega_i^2\left\{E\left(\int_0^1\underset{\sim}{J}_{c,i}(r)dW_i(r)\right)^2-\left(\int_0^1\int_0^r e^{(r-s)c}\tilde{h}(r,s)dsdr\right)^2\right\}$$
$$=\Omega_i^2V_{\hat{c}^{++}}(c),$$

where

$$V_{\hat{c}^{++}}(c)=\int_0^1\int_0^r e^{2c(r-s)}dsdr-2\int_0^1\int_0^1\int_0^{p\wedge s}e^{c(p+s-2x)}dx\tilde{h}(p,s)dsdp$$
$$-2\int_0^1\int_0^p\int_0^s e^{c(s-r)}e^{c(s-r)}\tilde{h}(p,s)drdsdp$$
$$+\int_0^1\int_0^1\int_0^1\left(\int_0^{p\wedge q}e^{c(p+q-2x)}dx\right)\tilde{h}(q,r)\tilde{h}(p,r)drdpdq$$
$$+\int_0^1\int_0^1\left(\int_0^p e^{c(p-s)}\tilde{h}(q,s)ds\right)\left(\int_0^q e^{c(q-r)}\tilde{h}(p,r)dr\right)dpdq. \quad \textbf{(34)}$$

Because $\sup_i \Omega_i^2 < \infty$ and $\lim_{n\to\infty}(1/n)\sum_{i=1}^n \Omega_i^2 = \Phi$, it follows by the Lindeberg–Levy central limit theorem that, as $n \to \infty$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \Omega_i\left(\int_0^1 \underset{\sim}{J}_{c,i}(r)dW_i(r) + \int_0^1\int_0^r e^{(r-s)c}\tilde{h}(r,s)dsdr\right) \Rightarrow N(0, \Phi V_{\hat{c}^{++}}(c)).$$

Combining this with the probability limit for the denominator of (33), $\Omega\omega(c)$, we have established the following theorem under sequential limit arguments. The same limit theory is obtained in Appendix D under joint limit arguments.

THEOREM 4. *Suppose that Assumptions* 1 *and* 2 *hold*. *Also assume that* $\tilde{c} = F^{-1}(\hat{c}^+)$ *is consistent for* c, $dF^{-1}(c)/dc$ *is well defined*, *and* $\omega(c) \neq 0$. *Then*, *as* $(T, n \to \infty)_{seq}$,

$$\sqrt{n}(\hat{c}^{++} - c) \Rightarrow N\left(0, \frac{\Phi V_{\hat{c}^{++}}(c)}{\Omega^2\omega(c)^2}\right), \tag{35}$$

where $V_{\hat{c}^{++}}(c)$ *and* $\omega(c)$ *are defined in* (34) *and* (27), *respectively*.

Remarks.

(a) In Table 3 we calculate numerical values of $\sqrt{V_{\hat{c}^{++}}(c)/\omega(c)^2}$, $-8 \le c \le 8$, where $\tilde{g}_t = (1,t)'$. When c is close to 0 or 1.3, $\omega(c) \sim 0$ (see Figure 6), and so $\sqrt{V_{\hat{c}^{++}}(c)/\omega(c)^2}$ takes high values around $c = 0$ and $c = 1.3$.

(b) Appendixes C and D establish joint consistency as in (29) for $(n, T \to \infty)$ (see Theorem 13) and joint asymptotic normality as in (35) for $(n, T \to \infty)$ with $(n/T) \to 0$ (see Theorem 14).

(c) When a consistent preliminary estimator for c is available, one may think of an estimator that corrects for the double biases in a simpler way by subtracting the estimates of Λ_i and $\omega_2(c)$. Let $\bar{c}$ be a consistent estimator for c, for example, $\bar{c} = F^{-1}(\hat{c}^+)$ or $\bar{c} = \hat{c}^{++}$. A simple double bias corrected estimator could then be defined as

$$\hat{c}^{+++} = T(\hat{a}^{+++} - 1),$$
$$\hat{a}^{+++} = \left(\sum_{i=1}^n\sum_{t=1}^T \underset{\sim}{z}_{i,t-1}^2\right)^{-1}\sum_{i=1}^n\sum_{t=1}^T(\underset{\sim}{z}_{i,t-1}\underset{\sim}{z}_{i,t} - T\hat{\Lambda}_i - \hat{\Omega}_i\omega_2(\bar{c})).$$

Because $\bar{c}$ is consistent for c and $\omega_2(c)$ is continuous, it is straightforward that $\hat{c}^{+++} \to_p c$. However, this simple bias corrected estimator has an undesirable property—its limit distribution depends on the asymptotic distribution of the preliminary estimator that is used to estimate the bias $\omega_2(c)$. Write, by definition,

$$\begin{aligned}\sqrt{n}(\hat{c}^{+++} - c) &= \left(\frac{1}{n}\sum_{i=1}^n\frac{1}{T^2}\sum_{t=1}^T \underset{\sim}{y}_{i,t-1}^2\right)^{-1} \\ &\quad\times\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{1}{T}\sum_{t=1}^T(\underset{\sim}{y}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t} - \hat{\Lambda}_i - \hat{\Omega}_i\omega_2(c))\right. \\ &\qquad\left. - \left(\frac{1}{n}\sum_{i=1}^n\hat{\Omega}_i\right)(\sqrt{n}(\omega_2(\bar{c}) - \omega_2(c)))\right).\end{aligned} \tag{36}$$

TABLE 3. Numerical values of $\sqrt{\dfrac{V_{\hat{c}^{++}}(c)}{\omega(c)^2}}$ in Theorem 4 when $\tilde{g}_t' = (1, t)'$

c	$\sqrt{\dfrac{V_{\hat{c}^{++}}(c)}{\omega(c)^2}}$ [a]	c	$\sqrt{\dfrac{V_{\hat{c}^{++}}(c)}{\omega(c)^2}}$	c	$\sqrt{\dfrac{V_{\hat{c}^{++}}(c)}{\omega(c)^2}}$	c	$\sqrt{\dfrac{V_{\hat{c}^{++}}(c)}{\omega(c)^2}}$	c	$\sqrt{\dfrac{V_{\hat{c}^{++}}(c)}{\omega(c)^2}}$
−8	4.9136	−4.6	7.5581	−1.2	44.6729	2.2	4.8422	5.6	0.0944
−7.9	4.9572	−4.5	7.7080	−1.1	51.9287	2.3	4.0131	5.7	0.0856
−7.8	5.0019	−4.4	7.8665	−1	61.4285	2.4	3.3761	5.8	0.0777
−7.7	5.0477	−4.3	8.0343	−0.9	74.2174	2.5	2.8747	5.9	0.0705
−7.6	5.0948	−4.2	8.2123	−0.8	92.0310	2.6	2.4718	6	0.0640
−7.5	5.1430	−4.1	8.4015	−0.7	117.9273	2.7	2.1423	6.1	0.0581
−7.4	5.1926	−4	8.6030	−0.6	157.7059	2.8	1.8690	6.2	0.0527
−7.3	5.2436	−3.9	8.8181	−0.5	223.5227	2.9	1.6394	6.3	0.0478
−7.2	5.2960	−3.8	9.0481	−0.4	344.4371	3	1.4446	6.4	0.0434
−7.1	5.3499	−3.7	9.2947	−0.3	605.2925	3.1	1.2778	6.5	0.0395
−7	5.4054	−3.6	9.5597	−0.2	1,349.9285	3.2	1.1339	6.6	0.0358
−6.9	5.4625	−3.5	9.8451	−0.1	5,369.5697	3.3	1.0090	6.7	0.0325
−6.8	5.5214	−3.4	10.1535	0	—	3.4	0.9000	6.8	0.0296
−6.7	5.5822	−3.3	10.4875	0.1	5,376.2421	3.5	0.8044	6.9	0.0269
−6.6	5.6448	−3.2	10.8503	0.2	1,355.8429	3.6	0.7202	7	0.0244
−6.5	5.7095	−3.1	11.2457	0.3	612.2524	3.7	0.6459	7.1	0.0222
−6.4	5.7764	−3	11.6780	0.4	353.0930	3.8	0.5800	7.2	0.0201
−6.3	5.8455	−2.9	12.1523	0.5	234.4024	3.9	0.5215	7.3	0.0183
−6.2	5.9171	−2.8	12.6746	0.6	171.4487	4	0.4694	7.4	0.0166
−6.1	5.9912	−2.7	13.2520	0.7	135.4350	4.1	0.4228	7.5	0.0151
−6	6.0680	−2.6	13.9830	0.8	114.7475	4.2	0.3813	7.6	0.0137
−5.9	6.1477	−2.5	14.6078	0.9	104.6621	4.3	0.3440	7.7	0.0125
−5.8	6.2305	−2.4	15.4088	1	104.6484	4.4	0.3106	7.8	0.0113
−5.7	6.3165	−2.3	16.3110	1.1	120.5580	4.5	0.2806	7.9	0.0103
−5.6	6.4061	−2.2	17.3329	1.2	189.3297	4.6	0.2537	8	0.0094
−5.5	6.4993	−2.1	18.4976	1.3	—	4.7	0.2294		
−5.4	6.5965	−2	19.8338	1.4	123.3519	4.8	0.2076		
−5.3	6.6980	−1.9	21.3779	1.5	51.3654	4.9	0.1879		
−5.2	6.8040	−1.8	23.1768	1.6	29.0267	5	0.1701		
−5.1	6.9150	−1.7	25.2913	1.7	18.7699	5.1	0.1541		
−5	7.0312	−1.6	27.8017	1.8	13.1506	5.2	0.1396		
−4.9	7.1532	−1.5	30.8156	1.9	9.7317	5.3	0.1266		
−4.8	7.2813	−1.4	34.4803	2	7.4963	5.4	0.1147		
−4.7	7.4161	−1.3	39.0012	2.1	5.9536	5.5	0.1040		

Note: The numerical values are obtained by 10,000 iterations of the simulation with size 1,000 data.
[a]Because $\omega(0) = 0$ and $\omega(1.3) \simeq 0$, we do not report the values of $\sqrt{(V_{\hat{c}^{++}}(c)/\omega(c)^2)}$.

In view of (36), the asymptotic distribution of the numerator of $\sqrt{n}(\hat{c}^{+++} - c)$ depends on the joint weak limit of $(1/\sqrt{n})\sum_{i=1}^{n}(1/T)\sum_{t=1}^{T}(\underset{\sim}{y}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t} - \hat{\Lambda}_i - \hat{\Omega}_i\omega_2(c))$ and $\sqrt{n}(\omega_2(\bar{c}) - \omega_2(c))$. The asymptotic distribution of $\sqrt{n}(\omega_2(\bar{c}) - \omega_2(c))$ then depends on the weak limit of $\sqrt{n}(\bar{c} - c)$ by standard delta method arguments. Therefore, the asymptotic distribution of $\sqrt{n}(\hat{c}^{+++} - c)$ relies on the limit distribution of the consistent preliminary estimator of c.

(d) On the other hand, the double bias corrected estimator $\hat{c}^{++}$ has a limit distribution that is independent of the weak limit of the preliminary consistent estimator. Therefore, even though the double bias corrected estimator $\hat{c}^{++}$ is more complicated than the simple estimator $\hat{c}^{+++}$ and suffers from the problem of a degenerate denominator for certain specific values of c (notably, $c = 0$), we prefer to recommend $\hat{c}^{++}$.

4. APPLICATIONS

4.1. Tests on the Localizing Coefficient

The asymptotic normality of $\sqrt{n}(\hat{c}^{+} - F(c))$ and $\sqrt{n}(\hat{c}^{++} - c)$ given in Theorems 2 and 4 enables us to construct tests for many interesting hypotheses. Suppose, for instance, that we are interested in testing the null hypothesis

$$\mathbb{H}_0 : c = c_0, \tag{37}$$

where c_0 belongs to a consistently estimable parameter set.[9] Then, for example, Theorem 4 suggests the following simple t-test based on $\hat{c}^{++}$:

$$t_{stat} = \frac{\sqrt{n}(\hat{c}^{++} - c_0)}{\sqrt{\dfrac{\hat{\Phi}V_{\hat{c}}(\hat{c}^{++})}{\hat{\Omega}^2\omega(\hat{c}^{++})^2}}},$$

where $\hat{\Omega} = (1/n)\sum_{i=1}^{n}\hat{\Omega}_i$, $\hat{\Phi} = (1/n)\sum_{i=1}^{n}\hat{\Omega}_i^2$. By Theorem 4, we have

$$t_{stat} \Rightarrow N(0,1)$$

as $(T, n \to \infty)_{seq}$. The joint limit convergence of t_{stat} to $N(0,1)$ is established in Appendix D.

As mentioned earlier, the panel model specification in model (1) that allows for a common local to unity parameter across individuals can sometimes be too restrictive. In such cases it may be of interest to test the difference of local to unity parameters between specific subgroups of individuals. Suppose that I_a and I_b denote two subgroups of individuals and we are interested in testing hypotheses about the local to unity parameters of model (8) in the following form:

$$z_{i,t} = \exp\left(\frac{c_\mu}{T}\right) z_{i,t-1} + \gamma_i' g_t + \varepsilon_{i,t},$$

where $c_\mu = c_a$ if $i \in I_a$ and $c_\mu = c_b$ if $i \in i_b$. A natural hypothesis is

$$\mathbb{H}_0 : c_a = c_b.$$

Let $n_a = \#(I_a)$ and $n_b = \#(I_b)$, respectively. Also, assume that $n_a/n_b \to \kappa < \infty$ as $n_a, n_b \to \infty$. The null hypothesis can be tested by computing the Wald statistic

$$W_{a,b} = n_b \frac{(\hat{c}_a^{++} - \hat{c}_b^{++})^2}{(n_a/n_b)\hat{V}_a + \hat{V}_b},$$

where $\hat{c}_\mu^{++}$ is a consistent estimator for c in group $\mu \in \{a,b\}$ and $\hat{V}_\mu = \hat{\Phi}_\mu V_{\hat{c}^{++}}(\hat{c}_\mu^{++})(\hat{\Omega}_\mu^2 \omega(\hat{c}_\mu^{++})^2)^{-1}$, $\mu \in \{a,b\}$. By Theorem 4, as $(T, n \to \infty)_{seq}$, we know

$$W_{a,b} \Rightarrow \chi_1^2,$$

a chi-square distribution with degree of freedom one.

4.2. An Application to Efficient Trend Elimination

In this section we show how consistent estimation of the localizing coefficient c can be used for efficient estimation of the trend coefficients. Suppose a trending time series z_t is generated by the system

$$z_t = \beta_0 + \beta_1' t + y_t, \qquad t = 1, \dots, T,$$

$$y_t = a y_{t-1} + \varepsilon_t, \qquad a = e^{c/T}\left(\simeq 1 + \frac{c}{T}\right), \tag{38}$$

where c denotes a local departure from unity, ε_t has mean zero and finite variance, and $y_0 = O_p(1)$ with a finite variance as $T \to \infty$. Suppose that our primary interest is in estimating the trend coefficient β_1 and in constructing confidence intervals for β_1. We assume a linear trend in model (38) because it is the most widely used specification in empirical applications. It is straightforward to allow for general polynomial trends, but to keep the algebra simple we do not discuss the general case here.

According to recent research (Phillips and Lee, 1996; Canjels and Watson, 1997), when the residual term y_t in (38) is nearly integrated, a partial generalized least squares (GLS) procedure based on quasi-differencing the data (called quasi-differencing detrending or QD detrending) is asymptotically more efficient than OLS in estimating the trend coefficient β_1. However, to execute feasible QD detrending it is necessary to estimate the unknown local to unity parameter c. But consistent estimation of c from a single time series trajectory is not generally possible, and this complicates estimation and inference about β_1 and the construction of valid confidence intervals for β_1. However, if panel data are available,[10] then the parameter c can be consistently estimated almost everywhere, as discussed in previous sections, and this makes efficient estimation of the trend coefficients (β_{1i}, say) possible and facilitates statistical inference.

Let $\hat{c}$ be a consistent estimate of c, for instance, $\hat{c} = F^{-1}(\hat{c}^{+})$, where c is in the identifiable region of OLS estimation, or $\hat{c} = \hat{c}^{++}$ and $\omega(c) \neq 0$. Define $\tilde{g}_t = (1, t)'$ and set

$$\breve{g}_{c,t} = \tilde{g}_t - \left(1 + \frac{\hat{c}}{T}\right)\tilde{g}_{t-1} = \begin{pmatrix} -\dfrac{\hat{c}}{T} \\ 1 - \dfrac{\hat{c}}{T}(t-1) \end{pmatrix}, \qquad \breve{g}_{c,1} = g_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and

$$\breve{z}_{c,t} = z_t - \left(1 + \frac{\hat{c}}{T}\right) z_{t-1}, \qquad \breve{z}_{c,1} = z_1,$$

for $t = 2, \ldots, T$. The QD estimator of the trend coefficient in a particular equation, say, $\beta = (\beta_0, \beta_1)'$ where we omit the subscript i for convenience, is defined as

$$\breve{\beta}_c = \left(\sum_{t=1}^{T} \breve{g}_{c,t}\,\breve{g}'_{c,t}\right)^{-1} \left(\sum_{t=1}^{T} \breve{g}_{c,t}\,\breve{z}_{c,t}\right), \tag{39}$$

Let $F_T = \operatorname{diag}(1, T)$. Then, it is easy to verify that, as T (and n) $\to \infty$,

$$F_T^{-1/2}\left(\sum_{t=1}^{T} \breve{g}_{c,t}\,\breve{g}'_{c,t}\right) F_T^{-1/2} \to_p \begin{pmatrix} 1 & 0 \\ 0 & \int_0^1 (1 - cr)^2 dr \end{pmatrix}. \tag{40}$$

Because the limit in (40) is block diagonal,

$$\begin{aligned} \sqrt{T}(\breve{\beta}_{1,c} - \beta_1) = {} & \left[\frac{1}{T}\sum_{t=2}^{T}\left(1 - \frac{\hat{c}}{T}(t-1)\right)^2\right]^{-1} \\ & \times \left[\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\left(1 - \frac{\hat{c}}{T}(t-1)\right)\left(z_t - \left(1 + \frac{\hat{c}}{T}\right) z_{t-1}\right)\right] \\ & + o_p(1). \end{aligned}$$

Write $\Omega = \lim_{T\to\infty} E((t/\sqrt{T})\sum_{t=1}^{T} \varepsilon_t)^2$. As T (and n) $\to \infty$,

$$\begin{aligned} \sqrt{T}(\breve{\beta}_{1,c} - \beta_1) \Rightarrow {} & \Omega\left(\int_0^1 (1 - cr)^2 dr\right)^{-1}\left(\int_0^1 (1 - cr)\, dW(r)\right) \\ \equiv {} & N\left(0, \Omega^2\left(\int_0^1 (1 - cr)^2 dr\right)^{-1}\right) \\ \equiv {} & N\left(0, \frac{\Omega^2}{1 - c + \frac{1}{3}c^2}\right). \end{aligned}$$

This estimator has the same limit distribution as the GLS estimator of β and, hence, attains the efficiency bound for the estimation of β in this class.

Next, suppose that $\hat{\Omega}$ is a consistent estimate of Ω. Then, using the consistent estimate of $c, \hat{c}$, we can conduct statistical inference about β_1. For example, a $(1-\alpha)\%$ asymptotic confidence interval for β_1 can be constructed as

$$\left[\breve{\beta}_{1,c} - z_{1-(\alpha/2)} \frac{1}{\sqrt{T}} \sqrt{\frac{\hat{\Omega}^2}{1-\hat{c}+\frac{1}{3}\hat{c}^2}}, \breve{\beta}_{1,c} + z_{1-(\alpha/2)} \frac{1}{\sqrt{T}} \sqrt{\frac{\hat{\Omega}^2}{1-\hat{c}+\frac{1}{3}\hat{c}^2}}\right], \quad \textbf{(41)}$$

where $z_{1-(\alpha/2)}$ is the two-sided $\alpha\%$ percentage point of the $N(0,1)$ distribution. In addition, to test hypotheses such as

$$H_0 : \beta_1 = \beta_{10},$$

we can use the Wald statistic

$$W = T(\breve{\beta}_{1,c} - \beta_{10})^2 \left(\frac{\hat{\Omega}^2}{1-\hat{c}+\frac{1}{3}\hat{c}^2} \right)^{-1}.$$

Because $\hat{c} \to c$, the Wald statistic converges in distribution to χ_1^2 as T (and n) $\to \infty$.

4.3. Estimation of Distant Initialization

As a referee has mentioned, if the initial conditions are random and in the distant past, then the limit theory and confidence intervals such as (41) need to be modified to account for their effects. Thus, suppose we have, in place of the $O_p(1)$ condition on $y_{i,0}$, the alternate initialization

$$y_{i,0}^{\theta} = \sum_{j=0}^{[\theta T]} a^j \varepsilon_{i,-j} \quad \textbf{(42)}$$

(as in Phillips and Lee, 1996; Canjels and Watson, 1997), where $y_{i,0}^{\theta}$ is parameterized by the distant past parameter θ, which measures the distance into the past that the initialization extends in terms of some fraction θ of the present sample of time series data T. When $\varepsilon_{i,-j}$ satisfies Assumption 1, the distant past initialization (42) gives data at the beginning of the time series sample statistical properties similar to those of the sample itself. Then, as $T \to \infty$ we have

$$\frac{1}{\sqrt{T}} y_{i,0}^{\theta} \Rightarrow C_i K_{c,i}(\theta) =_d N\left(0, \Omega_i \int_0^{\theta} e^{2cr} dr\right), \quad \textbf{(43)}$$

where $K_{c,i}$ is a diffusion process with the same properties as $J_{c,i}$. Furthermore, in place of (10), we now have

$$\frac{1}{\sqrt{T}} y_{i,[Tr]} \Rightarrow C_i J_{c,i}(r) + e^{cr} C_i K_{c,i}(\theta), \quad \textbf{(44)}$$

where $K_{c,i}$ and $J_{c,i}$ are independent, in view of the short memory of the errors ε_t. It follows that θ, in addition to c, now plays a role in the limit theory and any confidence intervals constructed from it.

Just as c can be consistently estimated in panels by $\tilde{c} = F^{-1}(\hat{c}^+)$ under certain conditions, we might expect there to be some prospect for estimating θ in a related way. Indeed, if the initialization parameter θ is the same across i, it follows from independence across i and (43) that as $(T, n \to \infty)_{seq}$,

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{i,1}^{\theta}}{\sqrt{T}}\right)^2 \to_p \Omega \int_0^{\theta} e^{2cr}dr = \Omega\,\frac{e^{2c\theta}-1}{2c}.$$

From this formula, it is apparent that, when c is known, a simple consistent estimator of θ is given by

$$\hat{\theta}(c) = \frac{1}{2c}\log\left[1+\frac{2c\hat{\sigma}_0^2}{\hat{\Omega}}\right], \tag{45}$$

where $\hat{\Omega} = (1/n)\sum_{i=1}^{n}\hat{\Omega}_i$. Notice that in cases of panel models with unit roots (i.e., $c = 0$) the corresponding consistent estimator of θ would simply be the variance ratio $\hat{\theta}(0) = \hat{\sigma}_0^2/\hat{\Omega}$. When c is unknown, joint estimation of c and θ is possible, and the following outlines a consistent estimation procedure. With initial observations as in (42), the probability limit of $\hat{c}^+$ is dependent on the two unknown parameters c and θ. Suppose we write this dependence as $F(c,\theta)$. Then, a consistent estimator of c, say, $\check{c}$, can be found by inverting the concentrated limit function $F(c,\hat{\theta}(c))$ in the range of c where $F(c,\hat{\theta}(c))$ is monotonic, just as we did in the case of the iterated OLS estimator in Section 3. A consistent estimate of θ is then found as $\hat{\theta}(\check{c})$. Note that in all these cases $\hat{\theta}$ is also consistent when $\theta = 0$ and the initialization is $O_p(1)$.

What the preceding discussion indicates is that, under the assumption that all members of the panel originate at the same time in the distant past, there is the prospect of consistently estimating the distance parameter θ. Intuitively, estimates like $\hat{\theta}$ work because if there is distant initialization in the elements of the panel, it can be expected to show up in the extent of the observed variation in the first sample data point across the panel. The estimator $\hat{\theta}$ simply assesses this observed variation (namely, $\hat{\sigma}_0^2$) relative to a consistent estimate of the average long-run variation displayed by the panel (namely, $\hat{\Omega}$), with some adjustment to account for the presence of a root that is local to unity rather than at unity.

5. CONCLUDING REMARKS

This paper has studied the estimation of a common localizing parameter for models with near unit roots using panel data. First, it was shown that the local to unity parameter in a simple panel near-integrated regression model can be consistently estimated by straightforward pooling and ordinary least-squares regression. When deterministic trends are present in the panel regression model, the situation is much more complex. We have shown that the nice results for the model with no trends do not extend easily.

In particular, the simplest pooled estimator that is based on the use of ordinary least squares with detrended data has an asymptotic bias that depends upon the unknown localizing parameter. One solution suggested here is to use the numerical inverse of the bias function to obtain a consistent estimate of the localizing parameter. However, this suggestion works only when the true value of c is in the identifiable region.

As a second method, we developed an estimation procedure that corrects for the bias from the serial correlation and from the use of the detrended data, using a preliminary consistent estimator of c. This double bias corrected estimator is consistent except for a finite number of values in the parameter space of c. However, the set of values where this estimator is not consistent contains $c = 0$, which is an especially interesting case. Also, when the true parameter takes a value close to zero, in practice, we can expect the double bias corrected procedure to provide a poor estimate of the true localizing parameter because the denominator of the double bias corrected estimator will be close to zero in this case. Similar problems arise in the case of an IV estimator that avoids bias by prudent instrumenting. Thus, even with panel data and a common localizing coefficient, consistent estimation of the localizing parameter is a challenging task when we want to allow for deterministic trends in the model.

For those cases where consistent estimation of the localizing coefficient is possible (notably when $c < 0$), the methods are used to show how to perform efficient trend extraction for panel data. This gives us an empirically useful algorithm for efficiently estimating a deterministic trend in the presence of stochastic trends generated by near-integrated processes with a common localizing parameter. Another useful application of panel data lies in the consistent estimation of the distancing parameter that arises in the formulation of distant initial conditions. This parameter (which is expressed as a fraction of the length of the present time series sample) measures how far into the past the initialization extends in terms of the shocks that have determined it. It is shown that the observed variation in panel observations at the initial point in the time series sample provides enough information about presample observations to construct a consistent estimate of this parameter.

In the development of the asymptotic theory we make use of both sequential limits and joint limits for the indices (n, T). A limiting feature of the joint asymptotics that is sometimes needed in our development is the rate condition $(n/T) \to 0$, which means that the results are likely to be most relevant in panels where T is large and n is moderately large.

Finally, although we do not report the analysis here, the authors have been able to show that the Gaussian maximum likelihood estimator of c is also inconsistent in panel models with deterministic trends and near-integrated stochastic trends. This panel data example provides an interesting new case where maximum likelihood estimation is inconsistent in the presence of an infinite number of nuisance parameters. It is explored in Moon and Phillips (1999b).

NOTES

1. Some recent work by Phillips, Moon, and Xiao (2001) develops new block local to unity models in which the autoregressive roots are local to unity but not as close to unity as they are in conventional near-integrated models. In these block local to unity models, the authors show that it is possible to consistently estimate the block localizing coefficient from a single trajectory.

2. Recently, Moon and Phillips (1999b) study a panel model such as (1) with homogeneous trends. The present paper considers only the heterogeneous trends model, where there are special complications in estimation and inference, as we will show.

3. Suppose that $Z_i = (z_{i,1}, \ldots, z_{i,T})'$, $Z_{-1,i} = (z_{i,0}, \ldots, z_{i,T-1})'$, and $G = (g_1, \ldots, g_T)'$. Then, $\underset{\sim}{z}_{i,t-1}$ and $\Delta \underset{\sim}{z}_{i,t}$ are the tth elements of $Q_G Z_{-1,i}$ and $Q_G(Z_i - Z_{-1,i})$, respectively, where $Q_G = I_T - G(G'G)^{-1}G'$ and $t \geq 2$.

4. The correction is for serial correlation in $\varepsilon_{i,t}$, following Phillips (1987a).

5. In this case, the trend coefficients β_i are zeros, and so we can estimate c consistently as we have shown in the previous section. However, it is common in empirical practice to use demeaned data, and use of this estimator results in bias, as is apparent from the probability limit $F(c)$.

6. We consider only the linear trend case because it is the most widely used specification in empirical application.

7. For example, we may assume that the parameter set includes only negative region and zero, $c < 0$.

8. See $v_{22}(c)$ in Appendix A.

9. For example, in the case of iterative OLS estimation, $c_0 \in \{c : F(c) \text{ is monotonic}\}$.

10. The panel data are assumed to have common localizing parameter c but may have individually different unknown trends.

REFERENCES

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

Apostol, T. (1974) *Mathematical Analysis*. Reading: Addison-Wesley.

Billingsley, P. (1968) *Convergence of Probability Measures.* New York: Wiley.

Canjels, E. & M. Watson (1997) Estimating deterministic trends in presence of serially correlated errors. *Review of Economics and Statistics* 79, 184–200.

Cavanagh, C., G. Elliott, & J.H. Stock (1995) Inference in models with nearly integrated regressors. *Econometric theory* 11, 1131–1147.

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

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

Moon, H.R. & P.C.B. Phillips (1999a) Estimation of Autoregressive Roots near Unity Using Panel Data. Working paper in economics 1-99, University of California, Santa Barbara.

Moon, H.R. & P.C.B. Phillips (1999b) Maximum Likelihood Estimation in Panels with Incidental Trends. Working paper in economics 6-99, University of California, Santa Barbara.

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

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

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

Phillips, P.C.B. & H.R. Moon (2000) Nonstationary panel data analysis: An overview of some recent developments. *Econometric Reviews*, forthcoming.
Phillips, P.C.B. & H.R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.
Phillips, P.C.B., H.R. Moon, & Z. Xiao (2001) How to estimate autoregressive roots near unity. *Econometric Theory* 17, forthcoming.
Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.
Said, E. & D. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–607.
Stock, J.H. (1991) Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics* 28, 435–459.
Uhlig, H. (1994) On Jeffreys' prior when using the exact likelihood function. *Econometric Theory* 10, 633–644.
Xiao, Z. & P.C.B. Phillips (1999) Efficient detrending in cointegrating regression. *Econometric Theory* 15, 519–548.

APPENDIX A

Proof of Theorem 2. This proof derives the limit distribution of $\hat{c}^+$ under sequential limits. First, note that

$$\sqrt{n}(\hat{c}^+ - F(c)) = \sqrt{n}\left(T(\hat{a}^+ - 1) - c - \frac{\omega_2(c)}{\omega_1(c)}\right)$$
$$= \sqrt{n}\left(T(\hat{a}^+ - a) - \frac{\omega_2(c)}{\omega_1(c)}\right) + O\left(\frac{\sqrt{n}}{T}\right).$$

Let

$$A_{n,T} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2,$$
$$B_{n,T} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t} - \hat{\Lambda}_i\right).$$

Then,

$$\sqrt{n}\left(T(\hat{a}^+ - a) - \frac{\omega_2(c)}{\omega_1(c)}\right) = \sqrt{n}\left(\frac{B_{n,T}}{A_{n,T}} - \frac{\left(\frac{1}{n}\sum_{i=1}^{n}\Omega_i\right)\omega_2(c)}{\left(\frac{1}{n}\sum_{i=1}^{n}\Omega_i\right)\omega_1(c)}\right).$$

To establish asymptotic normality of $\hat{c}^+$, we first show that

$$\sqrt{n}\begin{pmatrix} A_{n,T} - \left(\frac{1}{n}\sum_{i=1}^{n}\Omega_i\right)\omega_1(c) \\ B_{n,T} - \left(\frac{1}{n}\sum_{i=1}^{n}\Omega_i\right)\omega_2(c)\end{pmatrix} \Rightarrow N(0, \Phi V(c)), \qquad \textbf{(A.1)}$$

where

$$V(c) = \begin{pmatrix} v_{11}(c) & v_{12}(c) \\ v_{12}(c) & v_{22}(c) \end{pmatrix}$$

is given at the end of this proof. Define $f(a,b) = (b/a)$. Then, application of the delta method leads directly to

$$\begin{aligned}\sqrt{n}\left(T(\hat{a}^{+} - a) - \frac{\omega_2(c)}{\omega_1(c)}\right) &= \sqrt{n}(f(A_{n,T}, B_{n,T}) - f(\Omega\omega_1(c), \Omega\omega_2(c))) \\ &\Rightarrow N\left(0, \Phi\begin{pmatrix} \dfrac{-\omega_2(c)}{\omega_1(c)^2} & \dfrac{1}{\omega_1(c)} \end{pmatrix} V(c) \begin{pmatrix} \dfrac{-\omega_2(c)}{\omega_1(c)^2} \\ \dfrac{1}{\omega_1(c)} \end{pmatrix}\right) \\ &\equiv N(0, \Phi V_{\hat{c}^{+}}(c)), \text{ say.}\end{aligned}$$

We now establish (A.1) with sequential limits.

As $T \to \infty$ for fixed n,

$$\sqrt{n}\begin{pmatrix} A_{n,T} - \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\Omega_i\right)\omega_1(c) \\ B_{n,T} - \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\Omega_i\right)\omega_2(c) \end{pmatrix} \Rightarrow \begin{pmatrix} \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\Omega_i\left(\int \underset{\sim}{J}^2_{i,c}(r)dr - \omega_1(c)\right) \\ \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\Omega_i\left(\int \underset{\sim}{J}^2_{i,c}(r)dW_i(r) - \omega_2(c)\right) \end{pmatrix},$$

where $\underset{\sim}{J}_{i,c}(r) = J_{i,c}(r) - \int J_{i,c}(s)\tilde{h}(s,r)ds$. Note that $\sup_i \Omega_i < \infty$. Then, by the multivariate Lindeberg–Feller central limit theorem with $\lim_n (1/n)\sum_{i=1}^{n}\Omega_i^2 = \Phi$, we have

$$\begin{pmatrix} \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\Omega_i\left(\int \underset{\sim}{J}^2_{i,c}(r)dr - \omega_1(c)\right) \\ \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\Omega_i\left(\int \underset{\sim}{J}_{i,c}(r)dW_i(r) - \omega_2(c)\right) \end{pmatrix} \Rightarrow N(0, \Phi V(c)),$$

where

$$v_{11}(c) = E\left(\int \underset{\sim}{J}^2_{i,c}(r)dr - \omega_1(c)\right)^2$$

$$v_{22}(c) = E\left(\int \underset{\sim}{J}_{i,c}(r)dW_i(r) - \omega_2(c)\right)^2$$

$$v_{12}(c) = E\left(\int \underset{\sim}{J}^2_{i,c}(r)dr - \omega_1(c)\right)\left(\int \underset{\sim}{J}_{i,c}(r)dW_i(r) - \omega_2(c)\right),$$

and $\Phi = \lim_n (1/n)\sum_{i=1}^{n}\Omega_i^2$. The limit covariance matrix $V(c)$ has components that are as follows. For notational brevity, we omit the individual index i in the following expressions.

(a) $v_{11}(c)$

$$
\begin{aligned}
&E\left(\int_0^1 \underset{\sim}{J}_c^2(r)dr - \omega_1(c)\right)^2 \\
&\quad = E\left(\int_0^1 \underset{\sim}{J}_c^2(r)dr\right)^2 - \omega_1(c)^2 \\
&\quad = E\left(\int_0^1 J_c^2(r)dr - \int_0^1\int_0^1 J_c(p)J_c(q)\tilde{h}(p,q)dpdq\right) \\
&\qquad \times \left(\int_0^1 J_c^2(s)ds - \int_0^1\int_0^1 J_c(x)J_c(y)\tilde{h}(x,y)dxdy\right) - \omega_1(c)^2 \\
&\quad = \int_0^1\int_0^1 E(J_c^2(r)J_c^2(s))drds \\
&\qquad - 2\int_0^1\int_0^1\int_0^1 E(J_c^2(r)J_c(s)J_c(p))\tilde{h}(s,p)drdsdp \\
&\qquad + \int_0^1\int_0^1\int_0^1\int_0^1 E(J_c(r)J_c(s)J_c(p)J_c(q))\tilde{h}(r,s)\tilde{h}(p,q)dqdpdsdr \\
&\qquad - \omega_1(c)^2 \\
&\quad = \int_0^1\int_0^1 \Psi(r,r,s,s)drds - 2\int_0^1\int_0^1\int_0^1 \Psi(r,r,s,p)\tilde{h}(s,p)drdsdp \\
&\qquad + \int_0^1\int_0^1\int_0^1\int_0^1 \Psi(r,s,p,q)\tilde{h}(r,s)\tilde{h}(p,q)dqdpdsdr - \omega_1(c)^2,
\end{aligned}
$$

where

$$
\begin{aligned}
\Psi(r,s,p,q) &= E(J_c(r)J_c(s)J_c(p)J_c(q)) \\
&= E\left(\int_0^r\int_0^s\int_0^p\int_0^q e^{c(r+s+p+q-v-x-y-z)}dW(v)dW(x)dW(y)dW(z)\right) \\
&= \int_0^{r\wedge s} e^{c(r+s-2x)}dx \int_0^{p\wedge q} e^{c(p+q-2x)}dx \\
&\quad + \int_0^{r\wedge p} e^{c(r+p-2x)}dx \int_0^{s\wedge q} e^{c(s+q-2x)}dx \\
&\quad + \int_0^{r\wedge q} e^{c(r+q-2x)}dx \int_0^{p\wedge s} e^{c(p+s-2x)}dx.
\end{aligned}
$$

(b) $v_{22}(c)$: Note that

$$E\left(\int_0^1 \underset{\sim}{J}_c(r)dW(r)\right)^2$$

$$= E\left[\left(\int_0^1 J_c(r)dW(r) - \int_0^1\int_0^1 J_c(p)\tilde{h}(p,q)dW(q)dp\right)\right.$$

$$\left.\times\left(\int_0^1 J_c(s)dW(s) - \int_0^1\int_0^1 J_c(x)\tilde{h}(x,y)dW(y)dx\right)\right]$$

$$= \int_0^1\int_0^1 E(J_c(r)J_c(s)dW(r)dW(s))$$

$$-2\int_0^1\int_0^1\int_0^1 E(J_c(r)dW(r)J_c(p)\tilde{h}(p,q)dW(q))dp$$

$$+\int_0^1\int_0^1\int_0^1\int_0^1 J_c(p)\tilde{h}(p,r)dW(r)J_c(q)\tilde{h}(q,s)dW(s)dpdq.$$

Using Lemma 5, which follows, we have

$$E\left(\int_0^1 \underset{\sim}{J}_c(r)dW(r) - \omega_2(c)\right)^2$$

$$= \int_0^1\int_0^r e^{2c(r-s)}dsdr - 2\int_0^1\int_0^1\int_0^{p\wedge q} e^{c(p+q-2x)}dx\tilde{h}(p,q)dqdp$$

$$-2\int_0^1\int_0^p\left(\int_0^r e^{c(r-s)}\tilde{h}(p,s)ds\right)e^{c(p-r)}drdp$$

$$+\int_0^1\int_0^1\int_0^1\left(\int_0^{p\wedge q} e^{c(p+q-2x)}dx\right)\tilde{h}(q,r)\tilde{h}(p,r)drdpdq$$

$$+\int_0^1\int_0^1\left(\int_0^p e^{c(p-s)}\tilde{h}(q,s)ds\right)\left(\int_0^q e^{c(q-r)}\tilde{h}(p,r)dr\right)dpdq$$

$$+\left(\int_0^1\int_0^p e^{c(p-s)}\tilde{h}(p,s)dsdp\right)^2 - \omega_2(c)^2. \tag{A.2}$$

(c) $v_{12}(c)$: Note that

$$E\left(\int_0^1 \underset{\sim}{J}_c^2(r)dr - \omega_1(c)\right)\left(\int_0^1 \underset{\sim}{J}_c(r)dW(r) - \omega_2(c)\right)$$

$$= E\left(\int_0^1 \underset{\sim}{J}_c^2(r)dr\int_0^1 \underset{\sim}{J}_c(r)dW(r)\right) - \omega_1(c)\omega_2(c)$$

$$= \int_0^1 \int_0^1 E(J_c^2(r) J_c(s) dW(s) dr)$$

$$- \int_0^1 \int_0^1 \int_0^1 E(J_c(r) J_c(s) \tilde{h}(r,s) J_c(p) dW(p)) ds dr$$

$$- \int_0^1 \int_0^1 \int_0^1 E(J_c^2(r) J_c(s) \tilde{h}(s,p) dW(p)) ds dr$$

$$+ \int_0^1 \int_0^1 \int_0^1 \int_0^1 E(J_c(r) J_c(s) \tilde{h}(r,s) J_c(p) \tilde{h}(p,q) dW(q)) dp ds dr$$

$$- \omega_1(c) \omega_2(c).$$

By simple modifications of (d) and (e) in Lemma 5, which follows, we have

$$E\left(\int_0^1 \underset{\sim}{J}_c^2(r) dr - \omega_1(c) \right) \left(\int_0^1 \underset{\sim}{J}_c(r) dW(r) - \omega_2(c) \right)$$

$$= 2 \int_0^1 \int_0^r \left(\int_0^s e^{c(r+s-2x)} dx \right) e^{c(r-s)} ds dr$$

$$- 2 \int_0^1 \int_0^1 \int_0^s \left(\int_0^{r \wedge p} e^{c(r+p-2x)} dx \right) e^{c(s-p)} dp \tilde{h}(r,s) ds dr$$

$$- 2 \int_0^1 \int_0^1 \left(\int_0^{r \wedge s} e^{c(r+s-2x)} dx \right) \left(\int_0^r e^{c(r-p)} \tilde{h}(s,p) dp \right) ds dr$$

$$- \int_0^1 \int_0^1 \left(\int_0^r e^{2c(r-x)} dx \right) \left(\int_0^s e^{c(s-p)} \tilde{h}(s,p) dp \right) ds dr$$

$$+ \int_0^1 \int_0^1 \int_0^1 \left(\int_0^{r \wedge s} e^{c(r+s-2x)} dx \right) \left(\int_0^p e^{c(p-q)} \tilde{h}(p,q) dq \right) dp \tilde{h}(r,s) ds dr$$

$$+ \int_0^1 \int_0^1 \int_0^1 \left(\int_0^{p \wedge s} e^{c(p+s-2y)} dy \right) \left(\int_0^r e^{c(r-q)} \tilde{h}(p,q) dq \right) dp \tilde{h}(r,s) ds dr$$

$$+ \int_0^1 \int_0^1 \int_0^1 \left(\int_0^{p \wedge r} e^{c(p+r-2x)} dx \right) \left(\int_0^s e^{c(s-q)} \tilde{h}(p,q) dq \right) dp \tilde{h}(r,s) ds dr$$

$$- \omega_1(c) \omega_2(c).$$ ■

LEMMA 5. *The following hold.*

(a)

$$E \int_0^1 \int_0^1 \int_0^1 J_c(s) J_c(p) dW(r) dW(s) dp$$

$$= \int_0^1 \int_0^1 \int_0^{p \wedge s} e^{c(p+s-2x)} dx ds dp + \int_0^1 \int_0^p \left(\int_0^s e^{c(s-r)} dr \right) e^{c(p-s)} ds dp.$$

(b)

$$E\int_0^1\int_0^1 J_c(r)J_c(s)dW(r)dW(s)$$

$$=\int_0^1\int_0^r e^{2c(r-s)}dsdr.$$

(c)

$$E\int_0^1\int_0^1\int_0^1\int_0^1 J_c(p)J_c(q)dW(r)dW(s)$$

$$=\int_0^1\int_0^1\int_0^1\left(\int_0^{p\wedge q} e^{c(p+q-2x)}dx\right)drdpdq$$

$$+\int_0^1\int_0^1\left(\int_0^p e^{c(p-s)}ds\right)\left(\int_0^q e^{c(q-r)}dr\right)dpdq$$

$$+\int_0^1\int_0^1\left(\int_0^p e^{c(p-r)}dr\right)\left(\int_0^q e^{c(q-s)}ds\right)dpdq.$$

(d)

$$E\int_0^1\int_0^1\int_0^1 J_c(r)J_c(s)J_c(p)dW(p)dsdr$$

$$=2\int_0^1\int_0^1\int_0^s\left(\int_0^{r\wedge p} e^{c(r+p-2x)}dx\right)e^{c(s-p)}dpdsdr.$$

(e)

$$E\int_0^1\int_0^1\int_0^1\int_0^1 J_c(r)J_c(s)J_c(p)dW(q)dpdsdr$$

$$=\int_0^1\int_0^1\int_0^1\left(\int_0^{r\wedge s} e^{c(r+s-2x)}dx\right)\left(\int_0^p e^{c(p-q)}dq\right)dpdsdr$$

$$+\int_0^1\int_0^1\int_0^1\left(\int_0^{p\wedge s} e^{c(p+s-2y)}dy\right)\left(\int_0^r e^{c(r-q)}dq\right)dpdsdr$$

$$+\int_0^1\int_0^1\int_0^1\left(\int_0^{p\wedge r} e^{c(p+r-2x)}dx\right)\left(\int_0^s e^{c(s-q)}dq\right)dpdsdr.$$

Proof.

(a) Note that

$$E\int_0^1\int_0^1\int_0^1 J_c(s)J_c(p)dW(r)dW(s)dp$$
$$=\int_0^1\int_0^1\int_0^1\int_0^s\int_0^p e^{c(s-x)}e^{c(p-y)}E(dW(x)dW(y)dW(r)dW(s))dp. \quad \textbf{(A.3)}$$

Then, part (a) holds because

$$\text{(A.3)} = \int_0^1\int_0^1\int_0^{p\wedge s} e^{c(p+s-2x)}dxdsdp \quad \text{if } x=y(<p)<s=r$$
$$=\int_0^1\int_0^p\left(\int_0^s e^{c(s-r)}dr\right)e^{c(p-s)}dsdp \quad \text{if } x=r<y=s<p$$
$$=0 \quad \text{otherwise.}$$

(b)

$$E\int_0^1\int_0^1 J_c(r)J_c(s)dW(r)dW(s)$$
$$=\int_0^1\int_0^1\int_0^r\int_0^s e^{c(r-x)}e^{c(s-y)}E(dW(x)dW(y)dW(r)dW(s))$$
$$=\int_0^1\int_0^r e^{2c(r-s)}dsdr$$

because only when $x=y<r=s$, $E(dW(x)dW(y)dW(r)dW(s))\neq 0$.

(c) Note that

$$E\int_0^1\int_0^1\int_0^1\int_0^1 J_c(p)J_c(q)dW(r)dW(s)dqdp$$
$$=\int_0^1\int_0^1\int_0^1\int_0^1\int_0^p\int_0^q e^{c(p-x)}e^{c(q-y)}E(dW(x)dW(y)dW(r)dW(s))dpdq. \quad \textbf{(A.4)}$$

Then, part (c) holds because

$$\text{(A.4)} = \int_0^1\int_0^1\int_0^1\left(\int_0^{p\wedge q} e^{c(p+q-2x)}dx\right)drdpdq \quad \text{if } x=y,\ r=s,\ x\neq r$$
$$=\int_0^1\int_0^1\left(\int_0^p e^{c(p-s)}ds\right)\left(\int_0^q e^{c(q-r)}dr\right)dpdq \quad \text{if } x=s,\ y=r,\ x\neq y$$
$$=\int_0^1\int_0^1\left(\int_0^p e^{c(p-r)}dr\right)\left(\int_0^q e^{c(q-s)}ds\right)dpdq \quad \text{if } x=r,\ y=s,\ x\neq y$$
$$=0 \quad \text{otherwise.}$$

(d)

$$E\int_0^1\int_0^1\int_0^1 J_c(r)J_c(s)J_c(p)dW(p)dsdr$$

$$= \int_0^1\int_0^1\int_0^1\int_0^r\int_0^s\int_0^p e^{c(r-x)}e^{c(s-y)}e^{c(p-z)}$$

$$\times E(dW(x)dW(y)dW(z)dW(p))dsdr \qquad \textbf{(A.5)}$$

$$= 2\int_0^1\int_0^1\int_0^s\left(\int_0^{r\wedge p} e^{c(r+p-2x)}dx\right)e^{c(s-p)}dpdsdr,$$

where the last equality holds because

$$\text{(A.5)} = \int_0^1\int_0^1\int_0^s\left(\int_0^{r\wedge p} e^{c(r+p-2x)}dx\right)e^{c(s-p)}dpdsdr \quad \text{if } z = x < p = y < s$$

$$= \int_0^1\int_0^1\int_0^r\left(\int_0^{s\wedge p} e^{c(s+p-2y)}dy\right)e^{c(r-p)}dpdsdr \quad \text{if } z = y < p = x < r$$

$$= 0 \ \text{ otherwise.}$$

(e) Note that

$$E\int_0^1\int_0^1\int_0^1\int_0^1 J_c(r)J_c(s)J_c(p)dW(q)dpdsdr$$

$$= \int_0^1\int_0^1\int_0^1\int_0^1\int_0^r\int_0^s\int_0^p e^{c(r-x)}e^{c(s-y)}e^{c(p-z)}$$

$$\times E(dW(x)dW(y)dW(z)dW(q))dpdsdr. \qquad \textbf{(A.6)}$$

Then, part (e) holds because

$$\text{(A.6)} = \int_0^1\int_0^1\int_0^1\left(\int_0^{r\wedge s} e^{c(r+s-2x)}dx\right)\left(\int_0^p e^{c(p-q)}dq\right)dpdsdr$$

if $x = y \neq z = q$

$$= \int_0^1\int_0^1\int_0^1\left(\int_0^{p\wedge s} e^{c(p+s-2y)}dy\right)\left(\int_0^r e^{c(r-q)}dq\right)dpdsdr$$

if $x = q \neq z = y$

$$= \int_0^1\int_0^1\int_0^1\left(\int_0^{p\wedge r} e^{c(p+r-2x)}dx\right)\left(\int_0^s e^{c(s-q)}dq\right)dpdsdr$$

if $x = z \neq q = y$

$$= 0 \quad \text{otherwise.}$$ ■

APPENDIX B

Background of joint convergence theory. The primary object of this paper is to develop asymptotic theories of inference for localizing coefficients in panel data models. In many applications of large n and T panel regression models, we are interested in the limit behavior of double indexed quantities such as $X_{n,T}$ that are constructed as averages of i.i.d. random variables $Y_{i,T}$, that is,

$$X_{n,T} = \frac{1}{n}\sum_{i=1}^{n} Y_{i,T}, \quad \textbf{(B.1)}$$

where the $Y_{i,T}$ are independent across i for all T. Typically, we need to find the probability limit of $X_{n,T}$ or the limit distribution of scaled quantities such as $\sqrt{n}X_{n,T}$ when $(n, T \to \infty)$. In earlier work, the authors (Phillips and Moon, 1999) provide a conceptual framework and rigorous definitions for joint convergence in probability and joint convergence in distribution for double indexed processes. This section briefly reviews some concepts and helpful results from that earlier work that will be used frequently for establishing joint limits in this paper. All of the results given in this section are proved in Phillips and Moon (1999).

As mentioned in the text of the paper, the sequential probability limit of $X_{n,T} = (1/n)\sum_{i=1}^{n} Y_{i,T}$ is established by letting the index T go to infinity first and then the second index n is passed to infinity later. Using existing time series limit theory we can often easily obtain the limit behavior of $Y_{i,T}$. For example, suppose that as $T \to \infty$

$$Y_{i,T} \Rightarrow Y_i \quad \textbf{(B.2)}$$

or

$$Y_{i,T} \to_p Y_i \quad \text{for all } i. \quad \textbf{(B.3)}$$

Then, by the independence of $Y_{i,T}$ across i for all T, it follows that $X_{n,T} \Rightarrow X_n$ or $X_{n,T} \to_p X_n$ as $T \to \infty$ for all n, where $X_n = (1/n)\sum_{i=1}^{n} Y_i$.

Also, in the case of (B.2), it is assumed that the Y_i are defined on the same probability space for all i so that the sum of the limit random variables $(1/n)\sum_{i=1}^{n} Y_i$ is meaningful. (The assumption that the Y_i are defined on the same probability space can be justified. For this, see Phillips and Moon, 1999.) By allowing $n \to \infty$ and applying an appropriate law of large numbers to

$$X_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \textbf{(B.4)}$$

with some regularity conditions we may then find the sequential limit of $X_{n,T}$. Let

$$\tilde{\mu}_X = \lim_n \frac{1}{n}\sum_{i=1}^{n} EY_i. \quad \textbf{(B.5)}$$

Then

$$X_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \to_{a.s.} \tilde{\mu}_X = \lim_n \frac{1}{n}\sum_{i=1}^{n} EY_i$$

so as $(T, n \to \infty)_{seq}$

$X_{n,T} \to_p \tilde{\mu}_X.$

In general, the sequential probability limit $\tilde{\mu}_X$ of $X_{n,T}$ is not the same as the probability limit of $X_{n,T}$ under joint divergence of the indices (n, T), and, in fact, the latter may not even exist without further conditions. The following theorem gives a set of sufficient conditions under which the joint probability limit and sequential probability limit of $X_{n,T}$ are equivalent.

THEOREM 6. *Suppose that we have* $(k \times 1)$ *random vectors* $Y_{i,T}$ *that are independent across i for all T and integrable. Assume that* $Y_{i,T} \Rightarrow Y_i$ *as* $T \to \infty$ *for all i. Define* $X_{n,T} = (1/n)\sum_{i=1}^n Y_{i,T}$ *and* $X_n = (1/n)\sum_{i=1}^n Y_i$.
Suppose the following conditions (i)–(iv) *hold*:

(i) $\limsup_{n,T}(1/n)\sum_{i=1}^n E\|Y_{i,T}\| < \infty$,
(ii) $\limsup_n(1/n)\sum_{i=1}^n E\|Y_i\| < \infty$,
(iii) $\limsup_{n,T}(1/n)\sum_{i=1}^n \|EY_{i,T} - EY_i\| = 0$,
(iv) $\limsup_{n,T}(1/n)\sum_{i=1}^n E\|Y_{i,T}\| 1\{\|Y_{i,T}\| > n\varepsilon\} = 0 \forall \varepsilon > 0$.

If $\tilde{\mu}_X = \lim_n(1/n)\sum_{i=1}^n EY_i$ *exists and* $X_n \to_p \tilde{\mu}_X$ *as* $n \to \infty$*, then* $X_{n,T} \to_p \tilde{\mu}_X$ *as* $(n, T \to \infty)$.

An interesting special case arises when the $Y_{i,T}$ are scaled versions of some i.i.d. random vectors $Q_{i,T}$. Suppose that $Y_{i,T} = C_i Q_{i,T}$, where $Q_{i,T}$ are i.i.d. across i for all T and C_i are $(k \times k)$ nonrandom matrices for all i. Suppose that $Q_{i,T} \Rightarrow Q_i$ as $T \to \infty$ for all i, so that $Y_i = C_i Q_i$. In general, $Y_{i,T}$ are heterogeneous across i unless C_i are same for all i. The source of the heterogeneity in $Y_{i,T}$ is the scale effects C_i.

COROLLARY 7. *Suppose that* $Y_{i,T} = C_i Q_{i,T}$*, where* $Q_{i,T}$ *are i.i.d. across i for all T and* C_i *are* $(k \times k)$ *nonrandom matrices for all i. Assume that* $Q_{i,T}$ *are integrable for all T and* $Q_{i,T} \Rightarrow Q_i$ *as* $T \to \infty$*. Assume that* $C = \lim_n(1/n)\sum_{i=1}^n C_i$ *exists. If* (i) $\|Q_{i,T}\|$ *are uniformly integrable in T for all i and* (ii) $\sup_i \|C_i\| < \infty$*, then* $(1/n)\sum_{i=1}^n Y_{i,T} \to_p CE(Q_i)$ as $(n, T \to \infty)$.

Remarks. Here we present four useful ways to verify condition (i) of Corollary 7, the uniform integrability of $\|Q_{i,T}\|$ in T. For notational simplicity, we omit the individual index, i.

(a) Suppose that $Q_T \Rightarrow Q$ as $T \to \infty$. Then, uniform integrability of $\|Q_T\|$ is equivalent to $E\|Q_T\| \to E\|Q\|$ as $T \to \infty$ (see Billingsley, 1968, Theorem 5.4).
(b) Suppose that $E\|Q_T\|^r < \infty$ for some $0 < r < \infty$ and $Q_T \to_p Q$ as $T \to \infty$. Then the following are equivalent. (i) $\|Q_T\|^r$ are uniformly integrable in T; (ii) $E\|Q_T\|^r \to E\|Q\|^r$; and (iii) $E\|Q_T - Q\|^r \to 0$. This is the Vitali theorem.
(c) Suppose that there exists a sequence of random variables U_T such that $U_T \geq \|Q_T\|$ for all T. Then, uniform integrability of U_T implies the uniform integrability of $\|Q_T\|$.
(d) Suppose that $Q_T = W_T Z_T$. If $\|W_T\|^2$ and $\|Z_T\|^2$ are uniformly integrable in T, then $\|Q_T\|$ are uniformly integrable.

Next we consider the joint convergence in distribution of the $\sqrt{n}$-standardized double indexed sequence $\sqrt{n}X_{n,T}$ given by

$$\sqrt{n}X_{n,T} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_{i,T}.$$

In many nonstationary panel applications, we find that a standardized sum of the time series for individual $i, Y_{i,T}$, can be approximated by a scaled version of i.i.d. random variables (or vectors), that is,

$$Y_{i,T} \approx C_i Q_{i,T},$$

where C_i is a constant and $Q_{i,T}$ is i.i.d. over the cross section with mean zero and finite variance.

The following lemma is helpful in deriving the joint limit distribution of a double indexed process such as $\sqrt{n}X_{n,T} = (1/\sqrt{n})\sum_{i=1}^{n} Y_{i,T}$, when $Y_{i,T} = C_i Q_{i,T}$.

THEOREM 8. *Suppose that* $Y_{i,T} = C_i Q_{i,T}$, *where the* $(k \times 1)$ *random vectors* $Q_{i,T}$ *are i.i.d.* $(0, \Sigma_T)$ *across i for all T and* C_i *are nonrandom matrices. Assume the following hold*:

(i) *Let* $\sigma_T^2 = \lambda_{\min}(\Sigma_T)$ *and* $\liminf_T \sigma_T^2 > 0$,
(ii) $[\max_{i\le n}\|C_i\|^2/\lambda_{\min}(\sum_{i=1}^{n} C_i C_i')] = O(1/n)$ *as* $n \to \infty$,
(iii) $\|Q_{i,T}\|^2$ *are uniformly integrable*,
(iv) $\lim_{n,T}(1/n)\sum_{i=1}^{n} C_i \Sigma_T C_i' = \Omega > 0$.

Then,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_{i,T} \Rightarrow N(0,\Omega) \quad \textit{as } (n, T \to \infty).$$

Some preliminary results. This section gives some useful results that will be used repeatedly in the following subsections.

(a) A particularly useful tool in treating the linear process $\varepsilon_{i,t}$ is the BN decomposition, which decomposes the linear filter into long-run and transitory elements. Phillips and Solo (1992) give details of how this method can be used to derive a large number of limit results. Under Assumption 1, the linear process $\varepsilon_{i,t}$ is decomposed as

$$\varepsilon_{i,t} = C_i u_{i,t} + \tilde{\varepsilon}_{i,t-1} - \tilde{\varepsilon}_{i,t}, \tag{B.6}$$

where $C_i = C_i(1)$, $\tilde{\varepsilon}_{i,t} = \sum_{j=0}^{\infty} \tilde{C}_{i,j} u_{i,t-j}$, and $\tilde{C}_{i,j} = \sum_{k=j+1}^{\infty} C_{i,k}$. Under the summability condition (c) in Assumption 1,

$$|C_i| \le \sum_{j=0}^{\infty} \bar{C}_j < \infty \tag{B.7}$$

and

$$E\tilde{\varepsilon}_{i,t}^2 \leq \left(\sum_{j=0}^{\infty} j\bar{C}_j \right)^2 \leq \left(\sum_{j=0}^{\infty} j^b \bar{C}_j \right)^2 < \infty, \tag{B.8}$$

where $b \geq 1$ and $\bar{C}_j = \sup_i C_{i,j}$ (see Phillips and Solo, 1992).

(b) Let $\bar{C} = \sum_{j=0}^{\infty} \bar{C}_j$. Define

$$E_{i,t} = \sum_{j=0}^{\infty} \bar{C}_j |u_{i,t-j}| \tag{B.9}$$

and

$$\tilde{E}_{i,t} = \sum_{j=0}^{\infty} \breve{C}_j |u_{i,t-j}|, \tag{B.10}$$

where $\breve{C}_j = \sum_{k=j+1}^{\infty} \bar{C}_k$. The two random variables defined in (B.9) and (B.10) are dominating random variables for $\varepsilon_{i,t}$ and $\tilde{\varepsilon}_{i,t}$, respectively, in the sense that $E_{i,t} \geq |\varepsilon_{i,t}|$ and $\tilde{E}_{i,t} \geq |\tilde{\varepsilon}_{i,t}|$ for all i and t. By definition, $E_{i,t}$ and $\tilde{E}_{i,t}$ are i.i.d. across i for all t and satisfy

$$E(E_{i,t}) = E|u_{i,1}| \sum_{j=0}^{\infty} \bar{C}_j < M$$

and

$$E(\tilde{E}_{i,t}^2) \leq Eu_{i,1}^2 \left(\sum_{j=0}^{\infty} \breve{C}_j \right)^2 < M \tag{B.11}$$

for some $M < \infty$. Throughout this Appendix and elsewhere in the paper we use M to denote a generic constant.

(c) Next, recall that

$$\tilde{h}_T(t,s) = \tilde{D}_T \tilde{g}_t' \left(\frac{1}{T} \sum_{t=1}^{T} \tilde{D}_T \tilde{g}_t \tilde{g}_t' \tilde{D}_T \right)^{-1} \tilde{g}_s \tilde{D}_T.$$

It is easy to see that when $t = [Tr]$ and $s = [Tp]$, as $T \to \infty$

$$\tilde{h}_T(t,s) \to \tilde{g}'(r) \left(\int \tilde{g}\tilde{g}' \right)^{-1} \tilde{g}(p) = \tilde{h}(r,p)$$

uniformly in $(r,p) \in [0,1] \times [0,1]$. The following limit also holds:

$$\sup_{1 \leq t,s \leq T} \tilde{h}_T(t,s) \to \sup_{0 \leq r,p \leq 1} \tilde{h}(r,p). \tag{B.12}$$

(d) Using the BN decomposition of $\varepsilon_{i,t}$, we can decompose $y_{i,t}$ into two terms—a long-run component of $y_{i,t}$ and a transitory component. By virtue of the definition of $y_{i,t}$,

$$y_{i,t} = \sum_{s=1}^{t} e^{\frac{(t-s)}{T} c} \varepsilon_{i,s} + e^{\frac{t}{T} 2c} y_{i,0}.$$

Using the BN decomposition (B.6) of $\varepsilon_{i,t}$, we can decompose $y_{i,t}$ as

$$y_{i,t} = C_i x_{i,t} + R_{i,t}, \tag{B.13}$$

where

$$x_{i,t} = \sum_{s=1}^{t} e^{\frac{(t-s)}{T}c} u_{i,s}$$

and

$$R_{i,t} = e^{\frac{(t-1)}{T}c}\tilde{\varepsilon}_{i,0} - \tilde{\varepsilon}_{i,t} + \sum_{s=1}^{t} e^{\frac{(t-s-1)}{T}c}\tilde{\varepsilon}_{i,s}(1-e^{\frac{c}{T}}) + e^{\frac{t}{T}c} y_{i,0}.$$

Next we introduce bounds for the moments of some random variables that will be frequently used in the following proofs. In particular:

$$E\left(\frac{x_{i,t}^2}{T}\right) = \frac{1}{T}\sum_{s=1}^{t} e^{\frac{t-s}{T}2c} \to \int_0^r e^{(r-s)2c} ds < M \quad \text{if } t = [Tr], \tag{B.14}$$

$$\frac{1}{T}\sum_{t=1}^{T}\sqrt{E\left(\frac{x_{i,t}^2}{T}\right)} = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{T}\sum_{s=1}^{t} e^{\frac{t-s}{T}2c}}$$
$$\to \int_0^1 \left(\int_0^r e^{(r-s)2c} ds\right)^{1/2} dr < M, \tag{B.15}$$

and

$$\begin{aligned}
&\sup_i \sup_{1\le t\le T} ER_{i,t}^2 \\
&\le \sup_{1\le t\le T} 4\Big\{ e^{\frac{t-1}{T}2c} \sup_i E\tilde{\varepsilon}_{i,0}^2 + \sup_i E\tilde{\varepsilon}_{i,t}^2 + (1-e^{c/T})^2 \\
&\qquad \times \sum_{p=1}^{t}\sum_{s=1}^{t} e^{\frac{t-1-p}{T}c} e^{\frac{t-1-s}{T}c} \sup_i E(\tilde{\varepsilon}_{i,s}\tilde{\varepsilon}_{i,p}) + e^{\frac{t}{T}2c}\sup_i Ey_{i,0}^2 \Big\} \\
&\le 4 \sup_{1\le t\le T}\Big\{ (e^{\frac{t-1}{T}2c}+1)\left(\sum_{j=0}^{\infty} j\bar{C}_j\right)^2 + (1-e^{c/T})^2 \left(\sup_{1\le p,s\le t} e^{\frac{2t-2-p-sc}{T}}\right) \\
&\qquad \times \sum_{p=1}^{t}\sum_{s=1}^{t} \sup_i |E(\tilde{\varepsilon}_{i,s}\tilde{\varepsilon}_{i,p})| \Big\} + 4\sup_{1\le t\le T} e^{(t/T)2c} \sup_i Ey_{i,0}^2 \\
&\le 4\left(\sum_{j=0}^{\infty} j\bar{C}_j\right)^2 \Big\{ \sup_{1\le t\le T}(e^{\frac{t-1}{T}2c}+1) + \frac{1}{T^2}(1-e^{c/T})^2 \\
&\qquad \times \left(\sup_{1\le p,s,t\le T} e^{\frac{2t-2-p-s}{T}c}\right) \sup_{1\le t\le T}\frac{t^2}{T^2}\Big\} \\
&\quad + 4\sup_{1\le t\le T} e^{\frac{t}{T}2c}\sup_i \sigma_{i,0}^2 \\
&\to 4\left(\sum_{j=0}^{\infty} j\bar{C}_j\right)^2 \Big\{ \sup_{0\le r\le 1}(e^{2rc}+1) + c^2\left(\sup_{0\le p,s,r\le 1} e^{(2r-p-s)c}\right)\Big\} \\
&\quad + 4\sup_{0\le r\le 1} e^{2rc}\sup_i \sigma_{i,0}^2 \\
&< M \quad \text{as } T\to\infty \quad \text{if } t = [Tr],
\end{aligned} \tag{B.16}$$

where the first inequality uses $(a+b+c+d)^2 \leq 4(a^2+b^2+c^2+d^2)$, the second inequality uses (B.8), and the third inequality holds by applying the Cauchy–Schwarz inequality and (B.8) to $|E(\tilde{\varepsilon}_{i,s}\tilde{\varepsilon}_{i,p})|$.

The next two lemmas will be useful in proving joint limits.

LEMMA 9. *Under Assumptions* 1 *and* 2, *as* $(T, n \to \infty)$ *the following hold*:

(a) $(1/n)\sum_{i=1}^n (1/T^2)\sum_{t=1}^T \underset{\sim}{z}_{i,t-1}^2 \to_p \Omega\omega_1(c)$

(b) $(1/n)\sum_{i=1}^n (1/T)\sum_{t=1}^T \underset{\sim}{z}_{i,t-1}\varepsilon_{i,t} \to_p \Lambda + \Omega\omega_2(c)$, *where* $\omega_1(c) = (-1/2c)\times\{1+(1/2c)(1-e^{2c})\} - \int_0^1\int_0^1 e^{c(r+s)}(1/2c)(1-e^{-2c(r\wedge s)})\tilde{h}(r,s)dsdr$, $\omega_2(c) = -\int_0^1\int_0^r e^{(r-s)c}\tilde{h}(r,s)dsdr$, *and* $\tilde{h}(r,s) = \tilde{g}(r)'(\int_0^1 \tilde{g}\tilde{g}')^{-1}\tilde{g}(s)$.

Proof. Part (a). Because $(1/n)\sum_{i=1}^n (1/T^2)\sum_{t=1}^T \underset{\sim}{z}_{i,t-1}^2 = (1/n)\sum_{i=1}^n (1/T^2)\times \sum_{t=1}^T \underset{\sim}{y}_{i,t-1}^2$ by (9), we will establish $(1/n)\sum_{i=1}^n(1/T^2)\sum_{t=1}^T \underset{\sim}{y}_{i,t-1}^2 \to_p \Omega\omega_1(c)$.

Define $\underset{\sim}{x}_{i,t-1} = x_{i,t-1} - (1/T)\sum_{s=1}^T x_{i,s-1}\tilde{h}_T(s,t)$ and $\underset{\sim}{R}_{i,t-1} = R_{i,t-1} - (1/T)\times\sum_{s=1}^T R_{i,s-1}\tilde{h}_T(s,t)$. From the decomposition (B.13), we have

$$\frac{1}{n}\sum_{i=1}^n \frac{1}{T^2}\sum_{t=1}^T \underset{\sim}{y}_{i,t-1}^2$$
$$= \frac{1}{n}\sum_{i=1}^n C_i^2 \frac{1}{T^2}\sum_{t=1}^T \underset{\sim}{x}_{i,t-1}^2 + 2\frac{1}{n}\sum_{i=1}^n C_i \frac{1}{T^2}\sum_{t=1}^T \underset{\sim}{x}_{i,t-1}\underset{\sim}{R}_{i,t-1} + \frac{1}{n}\sum_{i=1}^n \frac{1}{T^2}\sum_{t=1}^T \underset{\sim}{R}_{i,t-1}^2$$
$$= I_a + 2II_a + III_a, \quad \text{say.}$$

In what follows we show that $I_a \to_p \Omega\omega_1(c)$ and $II_a, III_a \to_p 0$ as $(n, T\to\infty)$.

For I_a, recall that

$$I_a = \frac{1}{n}\sum_{i=1}^n C_i^2 \frac{1}{T^2}\sum_{t=1}^T \underset{\sim}{x}_{i,t-1}^2.$$

Define $Q_{i,T} = (1/T^2)\sum_{t=1}^T \underset{\sim}{x}_{i,t-1}^2$. Note that $\{Q_{i,T}\}_i$ are i.i.d. across i and $Q_{i,T} \Rightarrow Q_i = \int_0^1 \underset{\sim}{J}_{c,i}^2(r)dr$. Because $EQ_i = \omega_1(c)$ and $\lim_n (1/n)\sum_{i=1}^n C_i^2 = \Omega$, $I_a \to_p \Omega\omega_1(c)$ as $(T, n\to\infty)_{seq}$. Thus, to conclude that $I_a \to_p \Omega\omega_1(c)$ as $(T, n\to\infty)$ it suffices to verify conditions (i)–(ii) in Corollary 7. Condition (ii) of Corollary 7 clearly holds by Assumption 1(c). For condition (i) of Corollary 7 observe that

$$EQ_{i,T} = \frac{1}{T}\sum_{t=2}^T \frac{1}{T}\sum_{s=1}^t e^{\frac{t-s}{T}2c}$$
$$- \frac{1}{T}\sum_{t=1}^T \frac{1}{T}\sum_{s=1}^T \left\{ e^{\frac{t+2}{T}c}\left(\frac{1}{T}\sum_{k=2}^{t\wedge s} e^{\frac{-k}{T}2c}\right)\tilde{h}_T(t,s)\right\}$$
$$\to \int_0^1\int_0^r e^{(r-s)2c}ds - \int_0^1\int_0^1 e^{c(r+s)}\left(\int_0^{r\wedge s} e^{-2kc}dk\right)\tilde{h}(r,s)drds$$
$$= \omega_1(c) = EQ_i \quad \text{as } T\to\infty.$$

Because $Q_{i,T} \geq 0$, $Q_{i,T} \Rightarrow Q_i$, and $EQ_{i,T} \to EQ_i$ as $T \to \infty$ are enough to assert that $\{Q_{i,T}\}_T$ are uniformly integrable by Theorem 5.4 in Billingsley (1968), it follows that (i) in Corollary 7 is satisfied.

Next, we prove that

$$II_a = \frac{1}{n}\sum_{i=1}^{n} C_i \frac{1}{T^2}\sum_{t=1}^{T} \underset{\sim}{x}_{i,t-1}\underset{\sim}{R}_{i,t-1} \to_p 0$$

and

$$III_a = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{T^2}\sum_{t=1}^{T} \underset{\sim}{R}_{i,t-1}^2 \to_p 0 \quad \text{as } n, T \to \infty$$

by showing that $E|II_a|, E|III_a| \to 0$ as $n, T \to \infty$.

First, we have

$$\begin{aligned}
E|II_a| &= E\left|\frac{1}{n}\sum_{i=1}^{n} C_i \frac{1}{T^2}\sum_{t=1}^{T} \underset{\sim}{x}_{i,t-1}\underset{\sim}{R}_{i,t-1}\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left\{|C_i| E\left|\frac{1}{T^2}\sum_{t=1}^{T} \underset{\sim}{x}_{i,t-1}\underset{\sim}{R}_{i,t-1}\right|\right\} \\
&\leq \bar{C}\frac{1}{n}\sum_{i=1}^{n} E\left|\frac{1}{T^2}\sum_{t=1}^{T} x_{i,t-1}R_{i,t-1} + \frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T} x_{i,t-1}R_{i,s-1}\tilde{h}_T(t,s)\right| \\
&\leq \bar{C}\frac{1}{n}\sum_{i=1}^{n} E\left|\frac{1}{T^2}\sum_{t=1}^{T} x_{i,t-1}R_{i,t-1}\right| \\
&\quad + \bar{C}\sup_{1\leq t,s\leq T}|\tilde{h}_T(t,s)|\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}|x_{i,t-1}||R_{i,s-1}|\right).
\end{aligned}$$

Observe that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} E\left|\frac{1}{T^2}\sum_{t=1}^{T} x_{i,t-1}R_{i,t-1}\right| \\
&\quad\leq \frac{1}{\sqrt{T}}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T} E\left|\frac{x_{i,t-1}}{\sqrt{T}} R_{i,t-1}\right| \\
&\quad\leq \frac{1}{\sqrt{T}}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\sqrt{E\left|\frac{x_{i,t-1}}{\sqrt{T}}\right|^2 E|R_{i,t-1}|^2} = O\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}$$

where the equality holds because $\sup_i \sup_{1\leq t\leq T} E|R_{i,t-1}|^2 = O(1)$ by (B.16) and the $x_{i,t-1}$ are i.i.d. across i with $(1/T)\sum_{t=1}^{T}\sqrt{E|(x_{i,t-1}/\sqrt{T})|^2} = O(1)$ by (B.15).

Next

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}|x_{i,t-1}||R_{i,s-1}|\right)\\
&\quad\le \frac{1}{\sqrt{T}}\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}|x_{i,t-1}|\right)\left(\frac{1}{T}\sum_{s=1}^{T}|R_{i,s-1}|\right)\\
&\quad\le \frac{1}{\sqrt{T}}\sqrt{E\left(\frac{1}{T}\sum_{t=1}^{T}\frac{|x_{i,t-1}|}{\sqrt{T}}\right)^2}\sup_i\sqrt{E\left(\frac{1}{T}\sum_{s=1}^{T}|R_{i,s-1}|\right)^2}\\
&\quad\le \frac{1}{\sqrt{T}}\left(\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{Ex_{i,t-1}^2}{T}}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sqrt{\sup_i ER_{i,t-1}^2}\right) = O\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}
$$

Because $\sup_{1\le s,t\le T}|\tilde{h}_T(s,t)| = O(1)$, we conclude that $E|II_a| = O(1/\sqrt{T}) \to 0$ as $T\to\infty$.

Similarly, $III_a \to_p 0$ as $(n,T\to\infty)$ because $E|III_a|\to 0$ as $(n,T\to\infty)$ in II_a, and we have all the required results to complete the proof of part (a).

Part (b). From (9) we know that $(1/n)\sum_{i=1}^{n}(1/T)\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}\varepsilon_{i,t} = (1/n)\sum_{i=1}^{n}\times(1/T)\sum_{t=1}^{T}\underset{\sim}{y}_{i,t-1}\varepsilon_{i,t}$. By definition,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{y}_{i,t-1}\varepsilon_{i,t}\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}\varepsilon_{i,t} - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}y_{i,t-1}\varepsilon_{i,s}\tilde{h}_T(t,s).
\end{aligned}
$$

To complete the proof we show $(1/n)\sum_{i=1}^{n}(1/T)\sum_{t=1}^{T}y_{i,t-1}\varepsilon_{i,t}\to_p \Lambda$ and $(1/n)\times\sum_{i=1}^{n}(1/T^2)\sum_{t=1}^{T}\sum_{s=1}^{T}y_{i,t-1}\varepsilon_{i,s}\tilde{h}_{1T}(t,s)\to_p -\Omega\omega_2(c)$ as $(n,T\to\infty)$.

Recall that $y_{i,t} = ay_{i,t-1}+\varepsilon_{i,t}$, where $a = e^{c/T}$. Note that by squaring $y_{i,t} = ay_{i,t-1}+\varepsilon_{i,t}$ and averaging over t and i we have

$$
\begin{aligned}
&a\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}\varepsilon_{i,t}\\
&\quad= a^2\frac{1}{2n}\sum_{i=1}^{n}\frac{1}{T}(y_{i,T-1}^2-y_{i,0}^2) - (a^2-1)\frac{1}{2n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}^2 - \frac{1}{2n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t}^2\\
&\quad= a^2 I_b - II_b - \tfrac{1}{2}III_b,\quad\text{say.}
\end{aligned}
$$

Modifying the arguments in the proof of part (a) by substituting $\tilde{h}_T(t,s) = 0$, we have $(1/2n)\sum_{i=1}^{n}(1/T^2)\sum_{t=1}^{T}y_{i,t-1}^2\to_p \frac{1}{2}\Omega\int_0^1\int_0^r e^{(r-s)2c}ds$ as $(n,T\to\infty)$. Also, it follows that $T(a^2-1)\to 2c$ as $T\to\infty$. Combining these two results, we have

$$
II_b = T(a^2-1)\frac{1}{2n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}y_{i,t-1}^2 \to_p c\Omega\int_0^1\int_0^r e^{(r-s)2c}ds \qquad \textbf{(B.17)}
$$

as $(n,T\to\infty)$.

Next, we show, using Theorem 6, that as $(n, T \to \infty)$

$$III_b = \frac{1}{2n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t}^2 \to_p \frac{1}{2}\lim_n \frac{1}{n}\sum_{i=1}^{n}\Omega_{\varepsilon_i} = \frac{1}{2}\Omega - \Lambda. \tag{B.18}$$

Define $Y_{i,T} = (1/T)\sum_{t=1}^{T}\varepsilon_{i,t}^2$ and $Y_i = E\varepsilon_{i,t}^2 = \Omega_{\varepsilon_i}$. Then, by the ergodic theorem, as $T \to \infty$

$$Y_{i,T} = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t}^2 \to_p \Omega_{\varepsilon_i} = Y_i,$$

so, for fixed n, $(1/n)\sum_{i=1}^{n} Y_{i,T} \to_p (1/n)\sum_{i=1}^{n} Y_i$ as $T \to \infty$. Also $(1/n)\sum_{i=1}^{n} Y_i \to \lim_n (1/n)\sum_{i=1}^{n}\Omega_{\varepsilon_i} = \Omega - 2\Lambda$ as $n \to \infty$. Thus, according to Theorem 6, verifying conditions (i)–(iv) is enough to ensure that (B.18) holds under joint limits as $(n, T \to \infty)$. Conditions (ii) and (iii) clearly hold by the preceding arguments. Condition (i) holds because

$$\limsup_{n,T} \frac{1}{n}\sum_{i=1}^{n} E|Y_{i,T}| = \limsup_{n,T}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}E\varepsilon_{i,t}^2 = \lim_n \frac{1}{n}\sum_{i=1}^{n}\Omega_{\varepsilon_i} = \Omega - 2\Lambda.$$

For condition (iv), note by the definition of $E_{i,t}$ in (B.9) that

$$0 \le Y_{i,T} = |Y_{i,T}| \le \frac{1}{T}\sum_{t=1}^{T}E_{i,t}^2.$$

Because the sequence $E_{i,t}^2$ is strictly stationary and ergodic in t for all i, $(1/T)\sum_{t=1}^{T} \times E_{i,t}^2 \to_p E(E_{i,t}^2)$ by the ergodic theorem. Then, by Vitali's theorem (see Remark (b) following Corollary 7), $(1/T)\sum_{t=2}^{T}E_{i,t}^2$ are uniformly integrable in T. Hence, for given $\varepsilon > 0$,

$$\limsup_{n,T}\frac{1}{n}\sum_{i=1}^{n}E|Y_{i,T}|\{|Y_{i,T}| > n\varepsilon\}$$
$$\le \limsup_{n,T} E\left(\frac{1}{T}\sum_{t=1}^{T}E_{i,t}^2\right)\left\{\left|\frac{1}{T}\sum_{t=1}^{T}E_{i,t}^2\right| > n\varepsilon\right\} = 0,$$

and we have verified condition (iv).

Finally, consider I_b. Because it holds that $(1/2n)\sum_{i=1}^{n}(1/T)E(y_{i,0}^2) \le (1/2T) \times \sup_i \sigma_{i,0}^2 \to 0$ as $(n, T \to \infty)$ by Assumption 1(d), we consider only $(1/2n) \times \sum_{i=1}^{n}(1/T)y_{i,T-1}^2$.

$$\begin{aligned}
&\frac{1}{2n}\sum_{i=1}^{n}\frac{1}{T}y_{i,T-1}^2 \\
&\quad = \frac{1}{2n}\sum_{i=1}^{n}C_i^2\frac{1}{T}x_{i,T-1}^2 + \frac{1}{n}\sum_{i=1}^{n}C_i\frac{1}{T}x_{i,T-1}R_{i,T-1} + \frac{1}{2n}\sum_{i=1}^{n}\frac{1}{T}R_{i,T-1}^2 \\
&\quad = I_{b1} + I_{b2} + I_{b3}, \quad \text{say},
\end{aligned} \tag{B.19}$$

where the second equality holds by the decomposition (B.13) of $y_{i,T-1}$.

We now employ the principle used in the proof of part (a). Write $Q_{i,T} = (1/T)x_{i,T-1}^2$ and $Q_i = J_{c,i}^2(1)$. Then,

$$I_{b1} = \frac{1}{2n}\sum_{i=1}^{n} C_i^2 Q_{i,T} \Rightarrow \frac{1}{2n}\sum_{i=1}^{n} C_i^2 Q_i$$

as $T \to \infty$ for fixed n. Because $\sup_i C_i = \bar{C} < \infty$ and Q_i have finite second moments (note that $J_{c,i}(r)$ is a Gaussian process), by the strong law of large numbers for independent, nonidentically distributed random variables, we have

$$\frac{1}{2n}\sum_{i=1}^{n} C_i^2 Q_i \to_p \frac{1}{2}\,\Omega E(Q_i) = \frac{1}{2}\,\Omega\left(\frac{1}{2c}\,(e^{2c}-1)\right).$$

Thus, as $(T, n \to \infty)_{seq}$

$$I_{b1} \to_p \frac{1}{2}\,\Omega\left(\frac{1}{2c}\,(e^{2c}-1)\right).$$

We now verify conditions (i) and (ii) of Corollary 7 to obtain the joint probability limit of I_{b1}. First, condition (ii) clearly holds by Assumption 1(c). For condition (i), note by (B.14) that as $T \to \infty$

$$EQ_{i,T} = E\,\frac{1}{T}\,x_{i,T-1}^2 \to \int_0^1 e^{2(1-s)c}ds = EJ_{c,i}(1)^2 = EQ_i.$$

Clearly, the $Q_{i,T}$ are positive. It follows therefore from the preceding discussion and Theorem 5.4 of Billingsley (1968) that the $Q_{i,T}$ are uniformly integrable. Combining this with the fact that $a^2 \to 1$, as $n, T \to \infty$, finally we have

$$a^2 I_{b1} \to_p \frac{1}{2}\,\Omega\int_0^1 e^{2(1-s)c}ds = \frac{1}{2}\,\Omega\left(\frac{1}{2c}\,(e^{2c}-1)\right). \tag{B.20}$$

By similar arguments to those used in the discussion of term II_a in part (a), we can show that as $n, T \to \infty$

$$I_{b2}, I_{b3} \to_p 0.$$

Now, in view of (B.17), (B.18), and (B.20), we have as $n, T \to \infty$

$$a\,\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{i,t}$$

$$\to_p \Omega\,\frac{1}{4c}\,(e^{2c}-1) - \Omega\left(\frac{1}{4c}\,e^{2c} - \frac{1}{4c} - \frac{1}{2}\right) - \frac{1}{2}\,\Omega + \Lambda = \Lambda.$$

Because as $(n, T \to \infty)$ we have

$$a\,\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{i,t} - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{i,t} \to_p 0,$$

it follows that as $(n,T \to \infty)$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{i,t} \to_p \Lambda.$$

Next, we prove that $(1/n)\sum_{i=1}^{n}(1/T^2)\sum_{t=1}^{T}\sum_{s=1}^{T} y_{i,t-1}\varepsilon_{i,s}\tilde{h}_T(t,s) \to_p \Omega\omega_2(c)$ as $(T,n \to \infty)$. Using the decomposition of $y_{i,t-1}$ in (B.13) and the BN decomposition of $\varepsilon_{i,t}$ and noticing that $E|(1/n)\sum_{i=1}^{n} y_{i,0}(1/T^2)\sum_{s=1}^{T}\varepsilon_{i,s}\tilde{h}_T(0,s)| \le (1/T)\big(\sup_i\sqrt{\sigma_{i,0}^2\Omega_\varepsilon}\big) \times (\sup_{1\le s\le T}\tilde{h}_T(0,s)) = O(1/T)$,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} y_{i,t-1}\varepsilon_{i,s}\tilde{h}_T(t,s)\\
&\quad= \frac{1}{n}\sum_{i=1}^{n} C_i^2\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{T} x_{i,t-1}u_{i,s}\tilde{h}_T(t,s)\\
&\qquad+ \frac{1}{n}\sum_{i=1}^{n} C_i\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{T} x_{i,t-1}(\tilde{\varepsilon}_{i,s-1}-\tilde{\varepsilon}_{i,s})\tilde{h}_T(t,s)\\
&\qquad+ \frac{1}{n}\sum_{i=1}^{n} C_i\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{T} R_{i,t-1}u_{i,s}\tilde{h}_T(t,s)\\
&\qquad+ \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{T} R_{i,t-1}(\tilde{\varepsilon}_{i,s-1}-\tilde{\varepsilon}_{i,s})\tilde{h}_T(t,s) + o_p(1)\\
&\quad= I_{bb} + II_{bb} + III_{bb} + IV_{bb}, \quad \text{say.}
\end{aligned}$$

We show that as $(T,n \to \infty)$, $I_{bb} \to_p \Omega\omega_2(c)$ and II_{bb}, III_{bb}, $IV_{bb} \to_p 0$.

Note that

$$\begin{aligned}
EI_{bb} &= \frac{1}{n}\sum_{i=1}^{n} C_i^2\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{T} E(x_{i,t-1}u_{i,s})\tilde{h}_T(t,s)\\
&= \left(\frac{1}{n}\sum_{i=1}^{n} C_i^2\right)\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} e^{\left(\frac{t-s}{T}\right)c}\tilde{h}_T(t,s)\\
&\to \Omega\int_0^1\int_0^r e^{(r-s)c}\tilde{h}(r,s)\,ds\,dr = \Omega\omega_2(c) \quad \text{as } (T,n\to\infty).
\end{aligned}$$

Thus, for $I_{bb} \to_p \Omega\omega_2(c)$ as $(T,n \to \infty)$ it remains to show that

$$\begin{aligned}
I_{bb} - EI_{bb} &= \frac{1}{n}\sum_{i=1}^{n} C_i^2\left(\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{T} x_{i,t-1}u_{i,s}\tilde{h}_T(t,s) - \frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} e^{\left(\frac{t-s}{T}\right)c}\tilde{h}_T(t,s)\right)\\
&\to_p 0 \quad \text{as } (T,n\to\infty).
\end{aligned}$$

Define

$$Q_{i,T} = \frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{T} x_{i,t-1}u_{i,s}\tilde{h}_T(t,s) - \frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} e^{\left(\frac{t-s}{T}\right)c}\,\tilde{h}_T(t,s).$$

Then, as $T \to \infty$

$$Q_{i,T} = \left(\frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{T} x_{i,t-1} u_{i,s} \tilde{h}_T(t,s) - \frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{t-1} e^{\left(\frac{t-s}{T}\right)c} \tilde{h}_T(t,s) \right)$$
$$\Rightarrow \int_0^1 \left\{ \int_0^1 J_{c,i}(t) \tilde{h}(t,s) dW_i(s) - \int_0^t e^{\left(\frac{t-s}{T}\right)c} \tilde{h}(t,s) ds \right\} dt = Q_i, \quad \text{say.}$$

For fixed n, as $T \to \infty$

$$\frac{1}{n} \sum_{i=1}^{n} C_i^2 Q_{i,T} \Rightarrow \frac{1}{n} \sum_{i=1}^{n} C_i^2 Q_i.$$

Note that

$$EQ_i = E \int_0^1 \left\{ \int_0^t J_{c,i}(t) \tilde{h}(t,s) dW_i(s) - \int_0^t e^{\left(\frac{t-s}{T}\right)c} \tilde{h}(t,s) ds \right\} dt$$
$$+ \int_0^1 \left\{ \int_t^1 E(J_{c,i}(t) \tilde{h}(t,s) dW_i(s)) \right\} dt = 0.$$

Because $\sup_i C_i^2 = \bar{C}^2 < \infty$, and the second moments of Q_i are finite (Q_i is a stochastic integral of a Gaussian process with respect to a standard Brownian motion), by the weak law of large numbers, as $n \to \infty$ we have

$$\frac{1}{n} \sum_{i=1}^{n} C_i^2 Q_i \to_p \lim_n \frac{1}{n} \sum_{i=1}^{n} C_i^2 EQ_i = 0,$$

and so, as $(T, n \to \infty)_{seq}$,

$$\frac{1}{n} \sum_{i=1}^{n} C_i^2 Q_{i,T} \to_p 0.$$

Now, verifying conditions (i) and (ii) in Corollary 7 is enough to conclude that $I - EI = (1/n) \sum_{i=1}^{n} C_i^2 Q_{i,T} \to_p 0$ as $(n, T \to \infty)$. Condition (ii) holds by the assumption that $\sup_i C_i^2 = \bar{C}^2 < \infty$. To verify condition (i), note that

$$|Q_{i,T}| \leq \left| \frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{T} x_{i,t-1} u_{i,s} \tilde{h}_T(t,s) \right| + \left| \frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{t-1} e^{\left(\frac{t-s}{T}\right)c} \tilde{h}_T(t,s) \right|.$$

Because the nonstochastic term $|(1/T^2) \sum_{t=2}^{T} \sum_{s=1}^{t-1} e^{\left(\frac{t-s}{T}\right)c} \tilde{h}_T(t,s)| = O(1)$, for the uniform integrability of the $|Q_{i,T}|$ it is enough to prove that $|(1/T^2) \sum_{t=2}^{T} \times \sum_{s=1}^{T} x_{i,t-1} u_{i,s} \tilde{h}_T(t,s)|$ are uniformly integrable, which holds by Remark (d) following Corollary 7 if $\|(1/T\sqrt{T}) \sum_{t=2}^{T} x_{i,t-1} \widetilde{D}_T \tilde{g}_t\|^2$ and $\|(1/\sqrt{T}) \sum_{s=1}^{T} u_{i,s} \widetilde{D}_T \tilde{g}_s\|^2$ are uniformly integrable. Because

$$\left\| \frac{1}{T\sqrt{T}} \sum_{t=2}^{T} x_{i,t-1} \widetilde{D}_T \tilde{g}_t \right\|^2 \Rightarrow \left\| \int J_{c,i}(r) \tilde{g}(r) dr \right\|^2 \quad \text{as } T \to \infty$$

and

$$
\begin{aligned}
E\left\| \frac{1}{T\sqrt{T}} \sum_{t=2}^{T} x_{i,t-1} \tilde{D}_T \tilde{g}_t \right\|^2
&= \operatorname{tr}\left(\frac{1}{T^3} \sum_{t=2}^{T} \sum_{s=2}^{T} E x_{i,t-1} x_{i,s-1} \tilde{D}_T \tilde{g}_t \tilde{g}_s' \tilde{D}_T \right) \\
&= \operatorname{tr}\left(\frac{1}{T^3} \sum_{t=2}^{T} \sum_{s=2}^{T} \sum_{q=1}^{(t\wedge s)-1} e^{\left(\frac{t+s-2-q}{T}\right)c} \tilde{D}_T \tilde{g}_t \tilde{g}_s' \tilde{D}_T \right) \\
&\to \operatorname{tr}\left(\int_0^1 \int_0^1 \int_0^{(r\wedge s)} e^{(r+s-q)c} \tilde{g}(r)\tilde{g}(s)' dq ds dr \right) \\
&= \operatorname{tr}\left(\iint E J_{c,i}(r) J_{c,i}(s) \tilde{g}(r) \tilde{g}(s)' ds dr \right),
\end{aligned}
$$

it follows that $\|1/(T\sqrt{T})\sum_{t=2}^{T} x_{i,t-1} \tilde{D}_T \tilde{g}_t\|^2$ are uniformly integrable in T. Similarly we can verify that $\|(1/\sqrt{T})\sum_{s=1}^{T} u_{i,s} \tilde{D}_T \tilde{g}_s\|^2$ are uniformly integrable. So condition (i) is satisfied.

Next, we show

$$
II_{bb} = \frac{1}{n}\sum_{i=1}^{n} C_i \frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{T} x_{i,t-1}(\tilde{\varepsilon}_{i,s-1} - \tilde{\varepsilon}_{i,s}) \tilde{h}_T(t,s) \to_p 0.
$$

Write

$$
\begin{aligned}
II_{bb} &= \frac{1}{n}\sum_{i=1}^{n} C_i \frac{1}{T^2} \sum_{t=2}^{T} x_{i,t-1} \tilde{\varepsilon}_{i,1} \tilde{h}_T(t,1) - \frac{1}{n}\sum_{i=1}^{n} C_i \frac{1}{T^2} \sum_{t=2}^{T} x_{i,t-1} \tilde{\varepsilon}_{i,T} \tilde{h}_T(t,T) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=2}^{T} C_i x_{i,t-1} \tilde{g}_t' \tilde{D}_T \left(\frac{1}{T} \sum_{t\le T} \tilde{D}_T \tilde{g}_t \tilde{g}_t' \tilde{D}_T \right)^{-1} \sum_{s=1}^{T-1} \tilde{D}_T (\tilde{g}_{s+1} - \tilde{g}_s) \tilde{\varepsilon}_{i,s} \\
&= II_{bb1} + II_{bb2} + II_{bb3}, \quad \text{say.}
\end{aligned}
$$

For $II_{bb} \to_p 0$ as $(T,n \to \infty)$, we show that $E|II_{bbi}| \to 0$ as $(T,n\to\infty)$ for all $i = 1,2,3$.

First, $E|II_{bb2}| \to 0$ as $(T,n \to \infty)$ because

$$
\begin{aligned}
E|II_{bb2}| &\le \sup_{1\le t\le T} \tilde{h}_T(t,T) E\left(\frac{1}{n}\sum_{i=1}^{n} C_i \frac{1}{T^2} \sum_{t=2}^{T} |x_{i,t-1} \tilde{\varepsilon}_{i,T}| \right) \\
&\le \sup_{1\le t\le T} \tilde{h}_T(t,T) \frac{1}{n}\sum_{i=1}^{n} \left\{ C_i \frac{1}{\sqrt{T}} E\left[\left(\frac{1}{T} \sum_{t=2}^{T} \left| \frac{x_{i,t-1}}{\sqrt{T}} \right| \right) |\tilde{\varepsilon}_{i,T}| \right] \right\} \\
&\le \sup_{1\le t\le T} \tilde{h}_T(t,T) \frac{1}{n}\sum_{i=1}^{n} \left(C_i \frac{1}{\sqrt{T}} \sqrt{ E\left(\frac{1}{T} \sum_{t=2}^{T} \left| \frac{x_{i,t-1}}{\sqrt{T}} \right| \right)^2 E|\tilde{\varepsilon}_{i,T}|^2 } \right) \\
&\le \sup_{1\le t\le T} \tilde{h}_T(t,T) \bar{C} \frac{1}{\sqrt{T}} \sqrt{ E\left(\frac{1}{T} \sum_{t=2}^{T} \left| \frac{x_{i,t-1}}{\sqrt{T}} \right| \right)^2 E|\tilde{E}_{i,T}|^2 } = O\left(\frac{1}{\sqrt{T}} \right),
\end{aligned}
$$

where the third inequality uses Cauchy–Schwarz, the fourth inequality holds by the definitions of $\bar{C}$ and $\tilde{E}_{i,T}$, and the equality holds by (B.12) with $s = T$ and $q = 1$, (B.15) and (B.11).

By similar arguments to those given earlier, we can also show that

$$E|II_{bb1}| \to 0 \quad \text{as } n, T \to \infty.$$

For II_{bb3}, observe that

$$\begin{aligned}|II_{bb3}| &= \left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=2}^{T} C_i x_{i,t-1}\tilde{g}_t'\left(\sum_{t=1}^{T}\tilde{g}_t\tilde{g}_t'\right)^{-1}\sum_{s=1}^{T-1}(\tilde{g}_{s+1}-\tilde{g}_s)\tilde{\varepsilon}_{i,s}\right| \\ &\le \sup_{1\le t,s\le T}\tilde{f}_T(t,s)\,\frac{1}{n}\sum_{i=1}^{n} C_i\,\frac{1}{T^3}\sum_{t=2}^{T}\sum_{s=1}^{T-1}|x_{i,t-1}\tilde{\varepsilon}_{i,t}|,\end{aligned}$$

where $\tilde{f}_T(t,s) = \tilde{g}_t'\tilde{D}_T'((1/T)\sum_{t=1}^{T}\tilde{D}_T\tilde{g}_t\tilde{g}_t'\tilde{D}_T')^{-1}T\tilde{D}_T(\tilde{g}_{s+1} - \tilde{g}_s)$. Then $II_{bb3} \to_p 0$ as $(T, n \to \infty)$ because

$$\begin{aligned}E|II_{bb3}| &\le \sup_{1\le t,s\le T}\tilde{f}_T(t,s)\bar{C}\,\frac{1}{\sqrt{T}}\sqrt{E\left(\frac{1}{T}\sum_{t=2}^{T}\left|\frac{x_{i,t-1}}{\sqrt{T}}\right|\right)^2 E\left(\frac{1}{T}\sum_{t=1}^{T-1}|\tilde{E}_{i,t}|\right)^2} \\ &\le \sup_{1\le t,s\le T}\tilde{f}_T(t,s)\bar{C}\,\frac{1}{\sqrt{T}}\sqrt{E\left(\frac{1}{T}\sum_{t=2}^{T}\left|\frac{x_{i,t-1}}{\sqrt{T}}\right|\right)^2 E\left(\frac{1}{T}\sum_{t=1}^{T-1}|\tilde{E}_{i,t}|^2\right)} = O\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

To show that $III \to_p 0$ as $(T, n \to \infty)$, it is enough to show that $E|III| \to 0$ as $(T, n \to \infty)$. Write $\tilde{G}_T = ((1/T)\sum_{t=1}^{T}\tilde{D}_T\tilde{g}_t\tilde{g}_t'\tilde{D}_T)^{-1}$. By the definitions of $\tilde{h}_T(t,s)$, $\bar{C} = \sup_i|C_i|$, and by the triangle inequality, we have

$$\begin{aligned}E|III_{bb}| &= E\left|\frac{1}{n}\sum_{i=1}^{n} C_i\,\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{T} R_{i,t-1}u_{i,s}\tilde{h}_T(t,s)\right| \\ &= E\left|\frac{1}{n}\sum_{i=1}^{n} C_i\,\frac{1}{T^2}\left(\sum_{t=2}^{T} R_{i,t-1}\tilde{D}_T\tilde{g}_t'\right)\tilde{G}_T\left(\sum_{s=1}^{T}\tilde{D}_T\tilde{g}_s u_{i,s}\right)\right| \\ &\le \bar{C}\,\frac{1}{n}\sum_{i=1}^{n} E\left|\frac{1}{T^2}\left(\sum_{t=2}^{T} R_{i,t-1}\tilde{D}_T\tilde{g}_t'\right)\tilde{G}_T\left(\sum_{s=1}^{T}\tilde{D}_T\tilde{g}_s u_{i,s}\right)\right| \\ &\le \bar{C}\|\tilde{G}_T\|\,\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}E\left\|\frac{1}{T}\sum_{t=2}^{T} R_{i,t-1}\tilde{D}_T\tilde{g}_t\right\|\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\tilde{D}_T\tilde{g}_s u_{i,s}\right\|,\end{aligned}$$

where the last inequality uses Cauchy–Schwarz and the inequality $\|AB\| \le \|A\|\|B\|$. Again, by Cauchy–Schwarz, the last term in the preceding expression is less than

$$\begin{aligned}&\frac{1}{\sqrt{T}}\bar{C}\|\tilde{G}_T\|\,\frac{1}{n}\sum_{i=1}^{n}\sqrt{E\left\|\frac{1}{T}\sum_{t=2}^{T} R_{i,t-1}\tilde{D}_T\tilde{g}_t\right\|^2 E\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\tilde{D}_T\tilde{g}_s u_{i,s}\right\|^2} \\ &\quad\le \frac{1}{\sqrt{T}}\bar{C}\|\tilde{G}_T\|\sqrt{\sup_i E\left\|\frac{1}{T}\sum_{t=2}^{T} R_{i,t-1}\tilde{D}_T\tilde{g}_t\right\|^2 E\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\tilde{D}_T\tilde{g}_s u_{i,s}\right\|^2}. \qquad \textbf{(B.21)}\end{aligned}$$

Note that

$$\sup_i E\left\|\frac{1}{T}\sum_{t=2}^{T} R_{i,t-1}\widetilde{D}_T\tilde{g}_t\right\|^2 \le \sup_{1\le t\le T}\|\widetilde{D}_T\tilde{g}_t\|^2 \sup_i E\left(\frac{1}{T}\sum_{t=2}^{T}|R_{i,t-1}|\right)^2$$
$$\le \sup_{1\le t\le T}\|\widetilde{D}_T\tilde{g}_t\|^2 \sup_i \sup_{t\le T} ER_{i,t}^2$$
$$= O(1) \quad \text{as } T\to\infty, \tag{B.22}$$

where the last equality holds by (B.16) and $\sup_{1\le t\le T}\|\widetilde{D}_T\tilde{g}_t\|^2 = O(1)$. Also it is easily seen that

$$E\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\widetilde{D}_T\tilde{g}_s u_{i,s}\right\|^2 = \operatorname{tr}\left(\frac{1}{T}\sum_{s=1}^{T}\widetilde{D}_T\tilde{g}_s\tilde{g}_s'\widetilde{D}_T\right) = O(1) \quad \text{as } T\to\infty. \tag{B.23}$$

In view of (B.21)–(B.23) it follows that $E|III_{bb}| = O(1/\sqrt{T})$, and so we have the required result. The proof of $IV_{bb}\to_p 0$ as $(T,n\to\infty)$ is analogous to that of II_{bb3}. ■

LEMMA 10. *Let* $\hat{\Lambda} = (1/n)\sum_{i=1}^n \hat{\Lambda}_i$ *and* $\hat{\Omega} = (1/n)\sum_{i=1}^n \hat{\Omega}_i$, *where* $\hat{\Lambda}_i$ *and* $\hat{\Omega}_i$ *are defined in* (20) *and* (21), *respectively. Suppose that Assumptions* 1–4 *hold. Then, as* $n,T\to\infty$ $\hat{\Lambda}\to_p \Lambda$ *and* $\hat{\Omega}\to_p \Omega$.

Proof of Lemma 10. In this proof we show only that $\hat{\Lambda}\to_p\Lambda$ as $(T,n\to\infty)$. Then, by the same principle as that used in the proof of $\hat{\Lambda}\to_p\Lambda$, we find that $\hat{\Omega}\to_p\Omega$ as $(T,n\to\infty)$ holds by a simple change of the summation of the lag window from $\sum_{j=1}^T$ to $\sum_{j=-T}^T$.

Define

$$\hat{\Lambda}_{i,\varepsilon,\varepsilon} = \sum_{j=1}^{T} w\left(\frac{j}{K}\right)\frac{1}{T}\sum_{t=1}^{T-j}\varepsilon_{i,t}\varepsilon_{i,t+j}, \qquad \hat{\Lambda}_{\varepsilon,\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_{i,\varepsilon,\varepsilon}. \tag{B.24}$$

Before we start the proof of Lemma 10 we introduce the following useful lemma.

LEMMA 11. *Suppose the assumptions in Lemma* 10 *hold. Then, as* $(n,T\to\infty)$, $\hat{\Lambda}_{\varepsilon,\varepsilon}\to_p\Lambda$.

Proof of Lemma 11. We show that as $(n,T\to\infty)$

$$E\left(\hat{\Lambda}_{\varepsilon,\varepsilon} - \frac{1}{n}\sum_{i=1}^{n}\Lambda_i\right)^2 = 0. \tag{B.25}$$

Then, because $(1/n)\sum_{i=1}^n\Lambda_i\to\Lambda$ it follows $\hat{\Lambda}_{\varepsilon,\varepsilon}\to_p\Lambda$. Observe that

$$E\left(\hat{\Lambda}_{\varepsilon,\varepsilon} - \frac{1}{n}\sum_{i=1}^{n}\Lambda_i\right)^2 \le \frac{1}{n}\sum_{i=1}^{n}E(\hat{\Lambda}_{i,\varepsilon,\varepsilon}-\Lambda_i)^2$$
$$\le \sup_i \operatorname{var}(\hat{\Lambda}_{i,\varepsilon,\varepsilon}) + \sup_i\{\operatorname{bias}(\hat{\Lambda}_{i,\varepsilon,\varepsilon})\}^2.$$

Let $\Gamma_i(j) = E(\varepsilon_{i,t}\varepsilon_{i,t+j})$ and $\mathrm{cum}_i(0,k,l,m)$ denote the fourth order cumulant of $(\varepsilon_{i,t}, \varepsilon_{i,t+k}, \varepsilon_{i,t+l}, \varepsilon_{i,t+m})$. We know by Assumption 1 that

$$\sup_i \Gamma_i(j) \leq \Gamma(j) = \sum_{k=0}^{\infty} \bar{C}_k \bar{C}_{k+j}$$

and

$$\sup_i |\mathrm{cum}_i(0,k,l,m)| \leq |\sigma_4 - 3| \sum_{j=0}^{\infty} \bar{C}_j \bar{C}_{j+k} \bar{C}_{j+l} \bar{C}_{j+m}.$$

Also, from the summability condition (c) in Assumption 1, it follows that if $q \leq b$ where b is given in condition (c) in Assumption 1, then

$$\sum_{j=0}^{\infty} j^q \Gamma(j) = \sum_{j=0}^{\infty} j^q \sum_{k=0}^{\infty} \bar{C}_k \bar{C}_{k+j} \leq \sum_{k=0}^{\infty} \bar{C}_k \sum_{j=0}^{\infty} (j+k)^q \bar{C}_{k+j} < \infty \tag{B.26}$$

and

$$\begin{aligned}\sup_i \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty} \mathrm{cum}_i(0,k,l,m) &\leq |\sigma_4 - 3| \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\infty} \bar{C}_j \bar{C}_{j+k} \bar{C}_{j+l} \bar{C}_{j+m} \\ &\leq |\sigma_4 - 3| \left(\sum_{j=0}^{\infty} \bar{C}_j\right)^4 < \infty.\end{aligned}$$

Choosing q as in the condition of the lemma and following the same lines of proof as that in Theorems 9 and 10 of Hannan (1970), we have as $T \to \infty$

$$\begin{aligned}&\frac{T}{K} \sup_i \mathrm{var}(\hat{\Lambda}_{i,\varepsilon,\varepsilon}) = O(1), \\ &K^q \sup_i \mathrm{bias}(\hat{\Lambda}_{i,\varepsilon,\varepsilon}) = O(1),\end{aligned} \tag{B.27}$$

which leads to

$$E\left(\hat{\Lambda}_{\varepsilon,\varepsilon} - \frac{1}{n}\sum_{i=1}^{n} \Lambda_i\right)^2 = O\left(\frac{K}{T}\right) \to 0 \quad \text{as } n, T \to \infty. \tag{B.28}$$

Now, we start the proof of Lemma 10. Let

$$\tilde{a} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2\right)^{-1} \sum_{i=1}^{n}\sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}\underset{\sim}{z}_{i,t}$$

and

$$\hat{\varepsilon}_{i,t} = \underset{\sim}{z}_{i,t} - \tilde{a}\underset{\sim}{z}_{i,t-1} = \varepsilon_{i,t} - (\tilde{a} - a)\underset{\sim}{z}_{i,t-1} - \frac{1}{T}\sum_{s=1}^{T} \varepsilon_{i,s}\tilde{h}_T(t,s).$$

Also, let

$$\hat{\Gamma}_i(j) = \frac{1}{T}\sum_{t=1}^{T-j}\hat{\varepsilon}_{i,t}\hat{\varepsilon}_{i,t+j}.$$

If we prove that as $(T, n \to \infty)$

$$\hat{\Lambda} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{T-1} w\left(\frac{j}{K}\right)\hat{\Gamma}_i(j) \to_p \hat{\Lambda}_{\varepsilon,\varepsilon},$$

then, in view of Lemma 11, we have the required result, namely,

$$\hat{\Lambda} \to_p \Lambda = \lim_n \frac{1}{n}\sum_{i=1}^{n}\Lambda_i \quad \text{as } (T, n \to \infty).$$

Notice by the triangle inequality that

$$\begin{aligned}
|\hat{\Lambda} - \hat{\Lambda}_{\varepsilon,\varepsilon}| &= \left|\sum_{j=1}^{T-1} w\left(\frac{j}{K}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T-j}(\hat{\varepsilon}_{i,t}\hat{\varepsilon}_{i,t+j} - \varepsilon_{i,t}\varepsilon_{i,t+j})\right| \\
&\le \left|\sum_{j=1}^{T-1} w\left(\frac{j}{K}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T-j}(\hat{\varepsilon}_{i,t} - \varepsilon_{i,t})\hat{\varepsilon}_{i,t+j}\right| \\
&\quad + \left|\sum_{j=1}^{T-1} w\left(\frac{j}{K}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T-j}\varepsilon_{i,t}(\hat{\varepsilon}_{i,t+j} - \varepsilon_{i,t+j})\right|. \qquad \textbf{(B.29)}
\end{aligned}$$

By the Cauchy–Schwarz inequality the first term in (B.29) is less than

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1} w\left(\frac{j}{K}\right)\sqrt{\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}(\hat{\varepsilon}_{i,t} - \varepsilon_{i,t})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{i,t}^2} \\
&\quad = \sqrt{\frac{K^2}{T}}\left(\frac{1}{K}\sum_{j=1}^{K-1} w\left(\frac{j}{K}\right)\right)\sqrt{\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}(\hat{\varepsilon}_{i,t} - \varepsilon_{i,t})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{i,t}^2},
\end{aligned}$$

where the equality holds because $w(j/K)$ vanishes, by assumption, if $j > K$. First, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}(\hat{\varepsilon}_{i,t} - \varepsilon_{i,t})^2 \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[(\tilde{a} - a)\underset{\sim}{z}_{i,t-1} + \frac{1}{T}\sum_{s=1}^{T}\varepsilon_{i,s}\tilde{h}_T(t,s)\right]^2 \\
&\quad \le 2T^2(\tilde{a} - a)^2\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2 + \frac{2}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{1}{T}\sum_{s=1}^{T}\varepsilon_{i,s}\tilde{h}_T(t,s)\right)^2 \\
&\quad = 2I + 2II, \quad \text{say.} \qquad \textbf{(B.30)}
\end{aligned}$$

It follows by Lemma 9(a) that $(1/n)\sum_{i=1}^{n}(1/T^2)\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2 = O_p(1)$ as $n, T \to \infty$. Also, from Lemma 9, it is not difficult to see that $T^2(\tilde{a} - a)^2 = O_p(1)$ as $n, T \to \infty$. In consequence, $I = O_p(1)$ as $n, T \to \infty$. Note that

$$E|II| = E\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{1}{T}\sum_{s=1}^{T}\varepsilon_{i,s}\tilde{h}_T(t,s)\right)^2\right]$$
$$= E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{p=1}^{T}\varepsilon_{i,s}\varepsilon_{i,p}\tilde{h}_T(t,s)\tilde{h}_T(t,p)\right]$$
$$\leq \sup_{1\leq t,s\leq T}\tilde{h}_T(t,s)\sup_{1\leq t,p\leq T}\tilde{h}_T(t,p)\frac{1}{n}\sum_{i=1}^{n}\sum_{h=-T+1}^{T-1}\left(1-\frac{|h|}{T}\right)\Gamma_i(h) < \infty,$$

where the last inequality holds by (B.26), which implies that $II = O_p(1)$ as $(T, n \to \infty)$. Hence, $(1/n)\sum_{i=1}^{n}\sum_{t=1}^{T}(\hat{\varepsilon}_{i,t} - \varepsilon_{i,t})^2 = O_p(1)$ as $(T, n \to \infty)$.

Also, in a similar way, $(1/n)\sum_{i=1}^{n}(1/T)\sum_{t=1}^{T}\hat{\varepsilon}_{i,t}^2 = O_p(1)$ as $(T, n \to \infty)$ because

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{i,t}^2$$
$$\leq \frac{3}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t}^2 + eT^2(\tilde{a}-a)^2\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^3}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2$$
$$+ \frac{3}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{T}\sum_{s=1}^{T}\varepsilon_{i,s}\tilde{h}_T(t,s)\right]^2 = O_p(1).$$

Finally,

$$\sqrt{\frac{K^2}{T}}\frac{1}{K}\sum_{j=1}^{K}w\left(\frac{j}{K}\right) \to 0 \quad \text{as } T \to \infty, \tag{B.31}$$

where the convergence holds because $(K^2/T) \to 0$ and $(1/K)\sum_{j=1}^{K}w(j/K) \to \int_0^1 w(x)$. Hence, the first term in (B.29) is $o_p(1)$, and the second term in (B.29) is also $o_p(1)$. We now have all the desired results. ■

APPENDIX C: JOINT CONSISTENCY

THEOREM 12. *Under Assumptions* 1–4,

$$\hat{c}^+ \to_p F(c)$$

as $(T, n \to \infty)$.

Proof. The theorem holds by Lemmas 9 and 10. ■

THEOREM 13. *Suppose that the assumptions in Theorem* 14 *hold. Also assume that* $\omega(c) \neq 0$. *Then, as* $(T, n \to \infty)$,

$$\hat{c}^{++} \to_p c.$$

Proof. Recall that

$$\hat{c}^{++} = \left[\sum_{i=1}^{n} \left\{ \frac{1}{T^2} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 - \hat{\Omega}_i \frac{1}{T^2} \left(\sum_{t=1}^{T} \sum_{s=1}^{t} \left(\frac{t}{T} - \frac{s}{T} \right) e^{\left(\frac{t}{T} - \frac{s}{T}\right)\tilde{c}} \tilde{h}_T(t,s) \right) \right\} \right]^{-1}$$

$$\times \sum_{i=1}^{n} \left\{ \frac{1}{T} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1} \Delta \underset{\sim}{z}_{i,t} - \hat{\Lambda}_i \right.$$

$$\left. + \hat{\Omega}_i \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{t} \left[1 - \left(\frac{t}{T} - \frac{s}{T} \right) \tilde{c} \right] e^{\left(\frac{t}{T} - \frac{s}{T}\right)\tilde{c}} \tilde{h}_T(t,s) \right\}.$$

Then, the consistency $\hat{c}^{++} \to_p c$ as $(n,T) \to \infty$ is straightforward in view of the fact that $\tilde{c} \to_p c$ as $(n, T \to \infty)$, $r_T = T^2(\exp(c/T) - (1 + (c/T))) = O(1)$, and Lemmas 9 and 10. ∎

APPENDIX D: JOINT WEAK CONVERGENCE

To establish asymptotic normality of $\hat{c}^{+}$ and $\hat{c}^{++}$ under joint convergence, we need a stronger assumption on the bandwidth parameter used for the estimation of the long-run variance.

Assumption 5 (Bandwidth Condition′). As $(n, T \to \infty)$, the bandwidth parameter satisfies $K \to \infty$, $(nK^2/T) \to 0$, and $(nK^{2q+1}/T) \to \gamma > 0$ for some $\frac{1}{2} < q \le b$ for which w_q is finite, where b is given in condition (c) in Assumption 1.

THEOREM 14. *Suppose that Assumptions* 1–3 *and* 5 *hold. Also assume that* $\tilde{c} = F^{-1}(\hat{c}^{+})$ *is consistent for* c, $dF^{-1}(c)/dc$ *is well defined, and* $\omega(c) \neq 0$. *Then, as* $(n, T \to \infty)$ *with* $(n/T) \to 0$,

$$\sqrt{n}(\hat{c}^{++} - c) \Rightarrow N\left(0, \frac{\Phi V_{\hat{c}^{++}}(c)}{\Omega^2 \omega(c)^2}\right),$$

where $V_{\hat{c}^{++}}(c)$ *and* $\omega(c)$ *are defined in* (34) *and* (27), *respectively*.

Proof. To establish the joint limit distribution of $\hat{c}^{++}$ it is enough to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{T} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1} \underset{\sim}{\varepsilon}_{i,t} - \hat{\Lambda}_i + \hat{\Omega}_i \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{t} \left[1 - \left(\frac{t}{T} - \frac{s}{T} \right) (\tilde{c} - c) \right] \right.$$

$$\left. \times\, e^{\left(\frac{t}{T} - \frac{s}{T}\right)\tilde{c}} \tilde{h}_T(t,s) + r_T \frac{1}{T^3} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 \right\}$$

$$\Rightarrow N(0, \Phi V_{\hat{c}^{++}}(c)) \quad \text{as } (T, n \to \infty).$$

Notice that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t}-\hat{\Lambda}_i+\hat{\Omega}_i\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}\left[1-\left(\frac{t}{T}-\frac{s}{T}\right)(\tilde{c}-c)\right]\right.$$
$$\left.\times e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}\tilde{h}_T(t,s)+r_T\frac{1}{T^3}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2\right\}$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t}-\hat{\Lambda}_i+\hat{\Omega}_i\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}e^{\left(\frac{t}{T}-\frac{s}{T}\right)c}\tilde{h}_T(t,s)\right\}$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}r_T\frac{1}{T^3}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\Omega}_i\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}\left\{e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}-e^{\left(\frac{t}{T}-\frac{s}{T}\right)c}-\left(\frac{t}{T}-\frac{s}{T}\right)e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}}(\tilde{c}-c)\right\}\right.$$
$$\left.\times\tilde{h}_T(t,s)\right\}. \qquad \textbf{(D.1)}$$

As discussed in Section 3.2, by the mean value theorem, the third term of (D.1) equals to

$$\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_i\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t}\left(\frac{t}{T}-\frac{s}{T}\right)(e^{\left(\frac{t}{T}-\frac{s}{T}\right)c^*}-e^{\left(\frac{t}{T}-\frac{s}{T}\right)\tilde{c}})\tilde{h}_T(t,s)\right)\sqrt{n}(\hat{c}-c),$$

where c^* is located between c and $\tilde{c}$. Notice that $c^*\to_p c$ as $(n,T\to\infty)$ because $\tilde{c}\to_p c$ as $(n,T\to\infty)$. Also, it is possible to show that $\sqrt{n}(\tilde{c}-c)$ is stochastically bounded by applying Theorem 15, which follows, and the delta method in joint limit as $(n,T\to\infty)$. Therefore, the third term in (D.1) is $o_p(1)$ in joint limit as $(n,T\to\infty)$.

The remaining proof consists of the following two steps.

Step 1.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{y}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t}-\Lambda_i+\Omega_i\int_0^1\int_0^r e^{(r-s)c}\tilde{h}(r,s)dsdr\right\}\Rightarrow N(0,\Phi V_{\hat{c}^{++}}(c))$$

as $(T,n\to\infty)$.

Step 2.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t}-\hat{\Lambda}_i+\hat{\Omega}_i\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t}e^{\left(\frac{t}{T}-\frac{s}{T}\right)c}\tilde{h}_T(t,s)+r_T\frac{1}{T^3}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2\right\}$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{y}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t}-\Lambda_i+\Omega_i\int_0^1\int_0^r e^{(r-s)c}\tilde{h}(r,s)dsdr\right\}$$
$$=o_p(1)\quad\text{as }(T,n\to\infty).$$

To establish step 2, it is enough to show that as $(T, n \to \infty)$

$$II_1 : \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\Lambda}_i - \Lambda_i) = o_p(1)$$

$$II_2 : \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \hat{\Omega}_i \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{t} e^{(r-s)c} \tilde{h}_T(t,s) - \Omega_i \int_0^1 \int_0^r e^{(r-s)c} \tilde{h}(r,s) ds dr \right\} = o_p(1)$$

$$II_3 : \frac{1}{\sqrt{n}} \sum_{i=1}^{n} r_T \frac{1}{T^3} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 = o_p(1),$$

where $r_T = T^2(\exp(c/T) - (1 + (c/T)))$. First, II_1 holds because

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\Lambda}_i - \Lambda_i) = \sqrt{\frac{nK^2}{T}} O_p(1) \to_p 0 \tag{D.2}$$

by (B.28), (B.31) and the condition in Assumption 5 that $(nK^2/T) \to 0$. In the same fashion, it holds also that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\Omega}_i - \Omega_i) = \sqrt{\frac{nK^2}{T}} O_p(1) \to_p 0. \tag{D.3}$$

Write

$$\begin{aligned} II_2 = {} & \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\Omega}_i - \Omega_i) \right) \left(\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{t} e^{(r-s)c} \tilde{h}_T(t,s) \right) \\ & + \left(\frac{1}{n} \sum_{i=1}^{n} \Omega_i \right) \sqrt{n} \left(\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{t} e^{(r-s)c} \tilde{h}_T(t,s) - \int_0^1 \int_0^r e^{(r-s)c} \tilde{h}(r,s) ds dr \right). \end{aligned} \tag{D.4}$$

From

$$\sup_{1 \le t \le T} \sup_{(t-1)/T \le r \le t/T} \left| \left(\frac{t}{T} \right)^k - r^k \right| = \frac{1}{T} O(1) \quad \text{for } k = 1, \ldots, p,$$

we can show that

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{t} e^{(r-s)c} h_T(t,s) = \int_0^1 \int_0^r e^{(r-s)c} h(r,s) ds dr + O\left(\frac{1}{T} \right).$$

Thus, we have $II_2 = o_p(1)O(1) + O(1)O(\sqrt{n}/T) = o_p(1)$. Finally, II_3 holds because

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} r_T \frac{1}{T^3} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 = \frac{\sqrt{n}}{T} r_T \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T} \sum_{t=1}^{T} \underset{\sim}{z}_{i,t-1}^2 \right) = O_p\left(\frac{\sqrt{n}}{T} \right) = o_p(1).$$

We now show step 1. Recall $\omega_2(c) = -\int_0^1 \int_0^r e^{(r-s)c} h(r,s)dsdr$. Using the decompositions (B.6) and (B.13), we write

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\frac{1}{T}\sum_{t=1}^T \underset{\sim}{y}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t} - \Lambda_i - \Omega_i\omega_2(c)\right)\\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\frac{1}{T}\sum_{t=2}^T \underset{\sim}{y}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t} - \Lambda_i - \Omega_i\omega_2(c)\right) + \frac{\sqrt{n}}{T}O_p(1)\\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^n\Omega_i\left(\frac{1}{T}\sum_{t=2}^T \underset{\sim}{x}_{i,t-1}\underset{\sim}{u}_{i,t} - \omega_2(c)\right)\\
&\qquad + \frac{1}{2\sqrt{n}}\sum_{i=1}^n\left(\frac{1}{T}\sum_{t=2}^T(\varepsilon_{i,t}^2 - \Omega_{\varepsilon_i})\right) + \frac{1}{2\sqrt{n}}\sum_{i=1}^n\Omega_i\left(\frac{1}{T}\sum_{t=2}^T(u_{i,t}^2-1)\right)\\
&\qquad + \frac{1}{\sqrt{n}}\sum_{i=1}^n(r_{1,i,T} - r_{2,i,T}) + \frac{\sqrt{n}}{T}O_p(1),
\end{aligned}$$

where $r_{1,i,T} = a^2(1/T)(2C_iR_{i,T}x_{i,T} + R_{i,T}^2) - T(a^2-1)(1/T^2)\sum_{t=2}^T(2C_ix_{i,t-1}\times R_{i,t-1} + R_{i,t-1}^2)$,

$$\begin{aligned}
r_{2,i,T} = \frac{1}{T^2}\sum_{t=2}^T\sum_{s=2}^T\{&R_{i,t-1}h_T(t,s)u_{i,s} + R_{i,t-1}h_T(t,s)(\tilde{\varepsilon}_{i,s-1} - \tilde{\varepsilon}_{i,s})\\
&+ C_ix_{i,t-1}h_T(t,s)(\tilde{\varepsilon}_{i,s-1} - \tilde{\varepsilon}_{i,s})\}.
\end{aligned}$$

The first line holds because $E|(1/\sqrt{n})\sum_{i=1}^n(1/T)\underset{\sim}{y}_{i,0}\underset{\sim}{\varepsilon}_{i,1}| = (\sqrt{n}/T)O_p(1)$, and the second line holds because $(a-1)(1/\sqrt{n})\sum_{i=1}^n(1/T)\sum_{t=1}^T y_{i,t-1}\varepsilon_{i,t}$, $(1/n)\times\sum_{i=1}^n(y_{i,0}^2/T) = (\sqrt{n}/T)O_p(1)$.

In view of I_{b2}, I_{b3}, II_{bb}, III_{bb}, and IV_{bb} in the proof of Lemma 9(b), it follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n(r_{1,i,T} + r_{2,i,T}) = \sqrt{\frac{n}{T}}O_p(1) = o_p(1).$$

Next, $(1/\sqrt{n})\sum_{i=1}^n\Omega_i((1/T)\sum_{t=2}^T(u_{i,t}^2-1))$ because

$$\begin{aligned}
&E\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\Omega_i\left(\frac{1}{T}\sum_{t=2}^T(u_{i,t}^2-1)\right)\right)^2\\
&\quad = \left(\frac{1}{n}\sum_{i=1}^n\Omega_i^2\right)E\left(\frac{1}{T}\sum_{t=2}^T(u_{i,t}^2-1)\right)^2\\
&\quad = \frac{T-1}{T^2}\left(\frac{1}{n}\sum_{i=1}^n\Omega_i^2\right)E(u_{i,t}^2-1)^2 \to 0.
\end{aligned}$$

Also, $(1/\sqrt{n})\sum_{i=1}^{n}((1/T)\sum_{t=2}^{T}(\varepsilon_{i,t}^2 - \Omega_{\varepsilon_i})) \to_p 0$ because

$$
\begin{aligned}
E&\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=2}^{T}(\varepsilon_{i,t}^2 - \Omega_{\varepsilon_i}\right)\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}E\left(\frac{1}{T}\sum_{t=2}^{T}(\varepsilon_{i,t}^2 - \Omega_{\varepsilon_i})\right)^2 \\
&= \frac{1}{T}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{h=-T+2}^{T-2}\left(\frac{T-|h|}{T}\right)E(\varepsilon_{i,t}^2\varepsilon_{i,t+h}^2 - \Omega_{\varepsilon_i}^2)\right] \\
&\le \frac{1}{T}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{h=-T+2}^{T-2}\left(\frac{T-|h|}{T}\right)\right. \\
&\qquad \left.\times\left\{(2+\sigma_4)\left(\sum_{j=0}^{\infty}C_{i,j}C_{i,j+|h|}\right)^2 + \sigma_4\left(\sum_{j=0}^{\infty}C_{i,j}^2C_{i,j+|h|}^2\right)\right\}\right] \\
&\le \frac{1}{T}\left[2(2+\sigma_4)\sum_{h=0}^{\infty}\left(\sum_{j=0}^{\infty}\bar{C}_j\bar{C}_{j+h}\right)^2 + 2\sigma_4\sum_{h=0}^{\infty}\sum_{j=0}^{\infty}\bar{C}_j^2\bar{C}_{j+h}^2\right] \\
&\le \frac{1}{T}\left[2(2+\sigma_4)\left(\sum_{j=0}^{\infty}j^{1/2}\bar{C}_j^2\right)^2 + 2\sigma_4\left(\sum_{j=0}^{\infty}\bar{C}_j^2\right)\right] \to 0.
\end{aligned}
$$

Next, we write

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Omega_i\left(\frac{1}{T}\sum_{t=2}^{T}\underset{\sim}{x}_{i,t-1}\underset{\sim}{u}_{i,t} - \omega_2(c)\right) \\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Omega_i\left(\frac{1}{T}\sum_{t=2}^{T}\underset{\sim}{x}_{i,t-1}\underset{\sim}{u}_{i,t} - \omega_{2T}(c)\right) - \left(\frac{1}{n}\sum_{i=1}^{n}\Omega_i\right)\sqrt{n}(\omega_{2T}(c) - \omega_2(c)),
\end{aligned}
$$

where

$$
\begin{aligned}
\omega_{2T}(c) &= E\left(\frac{1}{T}\sum_{t=2}^{T}\underset{\sim}{x}_{i,t-1}\underset{\sim}{u}_{i,t}\right) \\
&= \frac{1}{T}\sum_{t=2}^{T}E(x_{i,t-1}u_{i,t}) - \frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{T}E(x_{i,t-1}u_{i,s})\tilde{h}_T(t,s) \\
&= -\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{T}\sum_{p=1}^{t-1}e^{\left(\frac{t-1-p}{T}\right)c}E(u_{i,p}u_{i,s})\tilde{h}_T(t,s) \\
&= -\frac{1}{T^2}\sum_{t=2}^{T}\sum_{p=1}^{t-1}e^{\left(\frac{t-1-p}{T}\right)c}\tilde{h}_T(t,p).
\end{aligned}
$$

From $\sup_{1\le t\le T}\ \sup_{(t-1)/T\le r\le t/T}|(t/T)^k - r^k| = (1/T)O(1)$ for $k = 1,\ldots,p$ and

$$
\sup_{1\le t\le T}\ \sup_{(t-1)/T\le r\le t/T}|e^{t/T} - e^r| = \frac{1}{T}O(1),
$$

we can show $\omega_{2T}(c) = \omega_2(c) + O(1/T)$. Thus, because it is assumed that $(1/n)\sum_{i=1}^n \Omega_i \to \Omega$ and $(n/T) \to 0$, we have $((1/n)\sum_{i=1}^n \Omega_i)\sqrt{n}(\omega_{2T}(c) - \omega_2(c)) = O(1)O(\sqrt{n}/T) = o(1)$.

To finish the proof of step 1, it remains to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Omega_i \left(\frac{1}{T} \sum_{t=2}^{T} \underset{\sim}{x}_{i,t-1} \underset{\sim}{u}_{i,t} - \omega_{2T}(c) \right) \Rightarrow N(0, \Phi V_{\hat{c}^{++}}(c)).$$

Let

$$V_{T,\hat{c}^{++}}(c) = E\left(\frac{1}{T} \sum_{t=2}^{T} \underset{\sim}{x}_{i,t-1} \underset{\sim}{u}_{i,t} \right)^2 - \omega_{2T}(c)^2$$
$$= E\left(\frac{1}{T} \sum_{t=2}^{T} x_{i,t-1} u_{i,t} \right)^2 - 2E\left(\frac{1}{T} \sum_{t=2}^{T} x_{i,t-1} u_{i,t} \frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{T} x_{i,t-1} u_{i,s} \tilde{h}_T(t,s) \right)$$
$$+ E\left(\frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{T} E(x_{i,t-1} u_{i,s}) \tilde{h}_T(t,s) \right)^2 - \omega_{2T}(c)^2.$$

Because

$$E\left(\frac{1}{T} \sum_{t=2}^{T} x_{i,t-1} u_{i,t} \right)^2 = \frac{1}{T^2} \sum_{t=2}^{T} \sum_{p=1}^{t-1} \sum_{s=2}^{T} \sum_{q=1}^{s-1} e^{\left(\frac{t-1-p}{T}\right)c} e^{\left(\frac{s-1-q}{T}\right)c} E(u_{i,p} u_{i,t} u_{i,q} u_{i,s})$$
$$= \frac{1}{T^2} \sum_{t=2}^{T} \sum_{p=1}^{t-1} e^{\left(\frac{t-1-p}{T}\right)2c} + O\left(\frac{1}{T}\right),$$

$$E\left(\frac{1}{T} \sum_{t=2}^{T} x_{i,t-1} u_{i,t} \frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{T} x_{i,t-1} u_{i,s} \tilde{h}_T(t,s) \right)$$
$$= \frac{1}{T^3} \sum_{t=2}^{T} \sum_{p=1}^{t-1} \sum_{s=2}^{T} \sum_{q=1}^{s-1} \sum_{w=1}^{T} e^{\left(\frac{t-1-p}{T}\right)c} e^{\left(\frac{s-1-q}{T}\right)c} E(u_{i,p} u_{i,t} u_{i,q} u_{i,w}) \tilde{h}_T(s,w)$$
$$= \frac{1}{T^3} \sum_{t=2}^{T} \sum_{s=2}^{T} \sum_{p=1}^{(t\wedge s)-1} e^{\left(\frac{t+s-2-2p}{T}\right)c} \tilde{h}_T(t,s)$$
$$+ \frac{1}{T^3} \sum_{s=3}^{T} \sum_{t=2}^{s-1} \sum_{p=1}^{t-1} e^{\left(\frac{t-1-p}{T}\right)c} e^{\left(\frac{s-1-t}{T}\right)c} \tilde{h}_T(s,p) + O\left(\frac{1}{T}\right),$$

and

$$E\left(\frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{T} E(x_{i,t-1} u_{i,s}) \tilde{h}_T(t,s) \right)^2$$
$$= \frac{1}{T^4} \sum_{t=2}^{T} \sum_{p=1}^{t-1} \sum_{r=1}^{T} \sum_{s=2}^{T} \sum_{q=1}^{s-1} \sum_{w=1}^{T} e^{\left(\frac{t-1-p}{T}\right)c} e^{\left(\frac{s-1-q}{T}\right)c}$$
$$\times E(u_{i,p} u_{i,r} u_{i,q} u_{i,w}) \tilde{h}_T(t,r) \tilde{h}_T(s,w)$$

$$= \frac{1}{T^4}\sum_{t=2}^{T}\sum_{s=2}^{T}\sum_{p=1}^{(t\wedge s)-1}\sum_{r=1}^{T} e^{\left(\frac{t+s-2-2p}{T}\right)c}\tilde{h}_T(t,r)\tilde{h}_T(s,r)$$

$$+ \frac{1}{T^4}\sum_{t=2}^{T}\sum_{p=1}^{t-1}\sum_{s=2}^{T}\sum_{q=1}^{s-1} e^{\left(\frac{t-1-p}{T}\right)c}e^{\left(\frac{s-1-q}{T}\right)c}\tilde{h}_T(t,p)\tilde{h}_T(s,q)$$

$$+ \frac{1}{T^4}\sum_{t=2}^{T}\sum_{p=1}^{t-1}\sum_{s=2}^{T}\sum_{q=1}^{s-1} e^{\left(\frac{t-1-p}{T}\right)c}e^{\left(\frac{s-1-q}{T}\right)c}\tilde{h}_T(t,q)\tilde{h}_T(s,p) + O\left(\frac{1}{T}\right),$$

we have

$$V_{T,\hat{c}^{++}}(c)$$

$$= \frac{1}{T^2}\sum_{t=2}^{T}\sum_{p=1}^{t-1} e^{\left(\frac{t-1-p}{T}\right)2c} - 2\frac{1}{T^3}\sum_{t=2}^{T}\sum_{s=2}^{T}\sum_{p=1}^{(t\wedge s)-1} e^{\left(\frac{t+s-2-2p}{T}\right)c}\tilde{h}_T(t,s)$$

$$- 2\frac{1}{T^3}\sum_{s=3}^{T}\sum_{t=2}^{s-1}\sum_{p=1}^{t-1} e^{\left(\frac{t-1-p}{T}\right)c}e^{\left(\frac{s-1-t}{T}\right)c}\tilde{h}_T(p,s)$$

$$+ \frac{1}{T^4}\sum_{t=2}^{T}\sum_{s=2}^{T}\sum_{p=1}^{(t\wedge s)-1}\sum_{r=1}^{T} e^{\left(\frac{t+s-2-2p}{T}\right)c}\tilde{h}_T(t,r)\tilde{h}_T(s,r)$$

$$+ \frac{1}{T^4}\sum_{t=2}^{T}\sum_{p=1}^{t-1}\sum_{s=2}^{T}\sum_{q=1}^{s-1} e^{\left(\frac{t-1-p}{T}\right)c}e^{\left(\frac{s-1-q}{T}\right)c}h_T(t,q)\tilde{h}_T(s,p) + O\left(\frac{1}{T}\right). \quad \textbf{(D.5)}$$

Now employ Theorem 8. Write $Q_{i,T} = (1/T)\sum_{t=1}^{T} \underset{\sim}{x}_{i,t-1}\underset{\sim}{u}_{i,t} - \omega_{2T}(c)$. The $Q_{i,T}$ are i.i.d. with mean zero and variance v_T. Also, we know that

$$Q_{i,T} \Rightarrow Q_i = \int \underset{\sim}{J}_{c,i}(r)dW(r) - \omega_2(c),$$

and it is not difficult to verify that

$$E(Q_{i,T})^2 = V_{T,\hat{c}^{++}}(c) \to V_{\hat{c}^{++}}(c) = E(Q_i)^2.$$

From $\sup_i \Omega_i^2 < \infty$, the convergence of $V_{T,\hat{c}^{++}}(c) \to V_{\hat{c}^{++}}(c)$, and $(1/n)\sum_{i=1}^{n}\Omega_i^2 \to \Phi$, we verify conditions (i), (ii), and (iv) of Theorem 8. Also, condition (iii) is satisfied by applying Remark (a) following Corollary 7 (see also Billingsley, 1968, Theorem 5.4) with $Q_{i,T}^2 \Rightarrow Q_i^2$ (by the continuous mapping theorem) and $E(Q_{i,T})^2 \to E(Q_i)^2$. Thus, by Theorem 8, $(1/\sqrt{n})\sum_{i=1}^{n}\Omega_i((1/T)\sum_{t=1}^{T}\underset{\sim}{x}_{i,t-1}\underset{\sim}{u}_{i,t} - \omega_{2T}(c)) \Rightarrow N(0,\Phi V_{\hat{c}^{++}}(c))$, and we have all the required results. ∎

THEOREM 15. *Suppose Assumptions* 1–3 *and* 5 *hold. Then, as* $(n,T \to \infty)$ *with* $(n/T) \to 0$,

$$\sqrt{n}(\hat{c}^{+} - F(c)) \Rightarrow N(0,\Phi V_{\hat{c}^{+}}(c)),$$

where

$$V_{\hat{c}^+}(c) = \begin{pmatrix} \dfrac{-\omega_2(c)}{\omega_1(c)^2} & \dfrac{1}{\omega_1(c)} \end{pmatrix} V(c) \begin{pmatrix} \dfrac{-\omega_2(c)}{\omega_1(c)^2} \\ \dfrac{1}{\omega_1(c)} \end{pmatrix}$$

and $V(c)$ is defined in the previous section.

Proof. The proof is entirely analogous to that of Theorem 14. We simply sketch the proof here. To establish the joint limit distribution of $\hat{c}^+$ it is enough to show that (A.1) holds under joint limits as $(n, T \to \infty)$. The idea of the proof is similar to that of Theorem 14.

First, by definition,

$$\sqrt{n}\begin{pmatrix} A_{n,T} - \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\Omega_i\right)\omega_1(c) \\ B_{n,T} - \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\Omega_i\right)\omega_2(c) \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\left(\dfrac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}^2 - \Omega_i\,\omega_1(c)\right) \\ \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\left(\dfrac{1}{T}\sum_{t=1}^{T}\underset{\sim}{z}_{i,t-1}\underset{\sim}{\varepsilon}_{i,t} - \hat{\Lambda}_i - \Omega_i\,\omega_2(c)\right) \end{pmatrix}. \qquad \textbf{(D.6)}$$

By applying similar arguments to those in the proof of Theorem 14 we can show that the main component for the joint asymptotic normality of (D.6) is the following:

$$\begin{pmatrix} \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\Omega_i\left(\dfrac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{x}_{i,t-1}^2 - \omega_1(c)\right) \\ \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\Omega_i\left(\dfrac{1}{T}\sum_{t=1}^{T}\underset{\sim}{x}_{i,t-1}\underset{\sim}{u}_{i,t} - \omega_2(c)\right) \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\Omega_i\left(\dfrac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{x}_{i,t-1}^2 - E\left(\dfrac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{x}_{i,t-1}^2\right)\right) \\ \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\Omega_i\left(\dfrac{1}{T}\sum_{t=1}^{T}\underset{\sim}{x}_{i,t-1}\underset{\sim}{u}_{i,t} - E\left(\dfrac{1}{T}\sum_{t=1}^{T}\underset{\sim}{x}_{i,t-1}\underset{\sim}{u}_{i,t}\right)\right) \end{pmatrix}$$

$$+ \begin{pmatrix} \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\Omega_i\left(E\left(\dfrac{1}{T^2}\sum_{t=1}^{T}\underset{\sim}{x}_{i,t-1}^2\right) - \omega_1(c)\right) \\ \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^{n}\Omega_i\left(E\left(\dfrac{1}{T}\sum_{t=1}^{T}\underset{\sim}{x}_{i,t-1}\underset{\sim}{u}_{i,t}\right) - \omega_2(c)\right) \end{pmatrix}$$

$$= I_d + II_d.$$

In much the same fashion as for step 2 in the proof of Theorem 13, we can show that $II_d \to 0$ as $(T, n \to \infty)$. Thus, to complete the proof it remains to show that $I_d \Rightarrow N(0, \Phi V_{\hat{c}^+}(c))$, for which we use Theorem 8. Conditions (a) and (b) are obvious. Also, after some tedious algebra similar to the derivation of (D.5) in the proof of Theorem 14, we can show that

$$E(I_d I_d') \to \Phi V_{\hat{c}^+}(c),$$

which is enough by Remark (a) following Corollary 7 to assert that conditions (iii) and (iv) are satisfied. ■

Proof of the Joint Weak Convergence of t_{stat}. For the joint limit of t_{stat} to $N(0,1)$, in view of the joint asymptotic normality in Theorem 14, it is enough to show that as $(T, n \to \infty)$

$$\hat{\Phi} \to_p \Phi,$$

that is,

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\Omega}_i^2 - \Omega_i^2) = o_p(1).$$

Note that

$$\left|\frac{1}{n}\sum_{i=1}^{n}(\hat{\Omega}_i^2 - \Omega_i^2)\right|$$

$$= \left|\frac{1}{n}\sum_{i=1}^{n}(\hat{\Omega}_i + \Omega_i)(\hat{\Omega}_i - \Omega_i)\right| \le \frac{1}{n}\sum_{i=1}^{n}(\hat{\Omega}_i - \Omega_i)^2 + 2\sup_i|\Omega_i|\left(\frac{1}{n}\sum_{i=1}^{n}|\hat{\Omega}_i - \Omega_i|\right)$$

$$\le \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|\hat{\Omega}_i - \Omega_i|\right)^2 + \frac{2\sup_i|\Omega_i|}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|\hat{\Omega}_i - \Omega_i|\right).$$

We know $\sup_i|\Omega_i|$ is finite, and so, to show $(1/n)\sum_{i=1}^{n}(\hat{\Omega}_i^2 - \Omega_i^2) = o_p(1)$, it is enough to show that $(1/\sqrt{n})\sum_{i=1}^{n}|\hat{\Omega}_i - \Omega_i| = o_p(1)$.

Let $\hat{\Omega}_{i,\varepsilon,\varepsilon}$ be a kernel estimator for Ω_i using the unknown errors $\varepsilon_{i,t}$, defined in an analogous way to (B.24). By the triangle inequality,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|\hat{\Omega}_i - \Omega_i| \le \frac{1}{\sqrt{n}}\sum_{i=1}^{n}|\hat{\Omega}_i - \hat{\Omega}_{i,\varepsilon\varepsilon}| + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}|\hat{\Omega}_{i,\varepsilon\varepsilon} - \Omega_i|.$$

By Cauchy–Schwarz, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}E|\hat{\Omega}_{i,\varepsilon\varepsilon} - \Omega_i| \le \sqrt{\sum_{i=1}^{n}(E|\hat{\Omega}_{i,\varepsilon\varepsilon} - \Omega_i|)^2} \le \sqrt{\sum_{i=1}^{n}E(\hat{\Omega}_{i,\varepsilon\varepsilon} - \Omega_i)^2}.$$

The square of the last term is less than

$$n\sup_i E(\hat{\Omega}_{i,\varepsilon\varepsilon} - \Omega_i)^2 \le n\sup_i \operatorname{var}(\hat{\Omega}_{i,\varepsilon\varepsilon}) + n\sup_i[\operatorname{bias}(\hat{\Omega}_{i,\varepsilon\varepsilon})]^2 = \frac{nK}{T}O(1), \tag{D.7}$$

where the last equality holds by (B.27), and so $(1/\sqrt{n})\sum_{i=1}^{n}|\hat{\Omega}_{i,\varepsilon\varepsilon} - \Omega_i| = o_p(1)$.

Next, by the triangle inequality again,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|\hat{\Omega}_i - \hat{\Omega}_{i,\varepsilon\varepsilon}| \\
&\quad \le \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=-T+1}^{T-1} w\left(\frac{j}{K}\right)\left|\frac{1}{T}\sum_t (\hat{\varepsilon}_{i,t}\hat{\varepsilon}_{i,t+j} - \varepsilon_{i,t}\varepsilon_{i,t+j})\right| \\
&\quad \le \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=-T+1}^{T-1} w\left(\frac{j}{K}\right)\left|\frac{1}{T}\sum_t (\hat{\varepsilon}_{i,t} - \varepsilon_{i,t})\hat{\varepsilon}_{i,t+j}\right| \\
&\qquad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=-T+1}^{T-1} w\left(\frac{j}{K}\right)\left|\frac{1}{T}\sum_t \varepsilon_{i,t}(\hat{\varepsilon}_{i,t+j} - \varepsilon_{i,t+j})\right|, \qquad \textbf{(D.8)}
\end{aligned}$$

where the summations over t satisfy $1 \le t, t+j \le T$. Then, following the same lines as in the argument following (B.29) with a change of summation in the lag kernel to $\sum_{j=-K}^{K}$, we are led to

$$(\text{D.8}) = \sqrt{\frac{nK^2}{T}}\frac{1}{K}\sum_{j=-K}^{K} w\left(\frac{j}{K}\right)O_p(1) = o_p(1).$$

This, together with $(1/\sqrt{n})\sum_{i=1}^{n}|\hat{\Omega}_i - \hat{\Omega}_{i,\varepsilon\varepsilon}|, (1/\sqrt{n})\sum_{i=1}^{n}|\hat{\Omega}_{i,\varepsilon\varepsilon} - \Omega_i| = o_p(1)$, gives us $(1/\sqrt{n})\sum_{i=1}^{n}|\hat{\Omega}_i - \Omega_i| = o_p(1)$. ■