GMM Estimation of Autoregressive Roots Near Unity with Panel Data

Hyungsik Roger Moon and Peter C.B. Phillips

USC Center for Law, Economics & Organization
Research Paper No. C02-27

CENTER FOR LAW, ECONOMICS AND ORGANIZATION RESEARCH PAPER SERIES

Sponsored by the John M. Olin Foundation

University of Southern California Law School
Los Angeles, CA 90089-0071

GMM Estimation of Autoregressive Roots Near Unity with Panel Data*

Hyungsik Roger Moon
Department of Economics
University of Southern California
&

Peter C.B. Phillips
Cowles Foundation, Yale University
University of Auckland & University of York

November 2002

Abstract

This paper investigates a generalized method of moments (GMM) approach to the estimation of autoregressive roots near unity with panel data and incidental deterministic trends. Such models arise in empirical econometric studies of firm size and in dynamic panel data modeling with weak instruments. The two moment conditions in the GMM approach are obtained by constructing bias corrections to the score functions under OLS and GLS detrending, respectively. It is shown that the moment condition under GLS detrending corresponds to taking the projected score on the Bhattacharya basis, linking the approach to recent work on projected score methods for models with infinite numbers of nuisance parameters (Waterman and Lindsay, 1998). Assuming that the localizing parameter takes a nonpositive value, we establish consistency of the GMM estimator and find its limiting distribution. A notable new finding is that the GMM estimator has convergence rate $n^{1/6}$, slower than $\sqrt{n}$, when the true localizing parameter is zero (i.e., when there is a panel unit root) and the deterministic trends in the panel are linear. These results, which rely on boundary point asymptotics, point to the continued difficulty of distinguishing unit roots from local alternatives, even when there is an infinity of additional data.

JEL Classification: C22 & C23

Keywords and Phrases: Bias, boundary point asymptotics, GMM estimation, local to unity, moment conditions, nuisance parameters, panel data, pooled regression, projected score.

---

*The authors are grateful to J. Horowitz and two referees for comments on earlier versions, to J. Owens for excellent research assistance, to J. Hahn and D. Bowman for helpful discussions, and to R. Waterman for sharing with us his recent joint papers with B. Lindsay. Moon thanks the Academic Research Committee of UCSB and the Faculty Development Awards of USC for research support and Phillips thanks the NSF for research support under NSFs SBR 97-30295 & SES 0092509.

1 Introduction

Recent years have seen the introduction of several important panel data sets where the cross sectional dimension (n) and the time series dimension (T) are comparable in magnitude. Some of these panel data sets, like the Penn World Tables, involve time series that are manifestly nonstationary and have persistent or slowly decaying serial correlations. These features distinguish the new data from the characteristics that are conventionally assumed in the analysis of panel data where T is very small and n is very large.

Since the early 1990's, there has been ongoing theoretical and applied research on panels whose time series components are nonstationary or persistent. For large n and fixed T panels, see Hahn *et al* (2001) and Kruiniger (2000). For large n and T panels allowing for nonstationarity in the data over time, the theoretical research includes the study of asymptotically unbiased estimation of the dynamic panel model (*e.g.*, Hahn and Kuersteiner, 2000), panel unit root tests (*e.g.*, Quah, 1994, Levin and Lin, 1993, Im *et al.*, 1996, Maddala and Wu, 1997, and Choi, 1999), panel cointegration tests (*e.g.*, Pedroni, 1999, Binder *et al.*, 1999), and the development of linear regression theories for panel estimators under nonstationarity (*e.g.*, Pesaran and Smith, 1995, and Phillips and Moon, 1999). Applied research includes tests of growth convergence theories (Bernard and Jones, 1996), purchasing power parity relations (MacDonald, 1996, Oh, 1996, Pedroni, 1996, Wu, 1996, and Wu, 1997), and studies of the international links between savings and investment (Coakley *et al.*, 1996 and Moon and Phillips, 1998).

Two recent papers by the authors (Moon and Phillips, 1999 & 2000) study panel regression models that allow for both deterministic trends and stochastic trends with roots local to unity. As we discuss in Section 2 of the present paper, such models are important empirically in studying Gibrat's law and they have received attention recently in the weak instrument literature. When the deterministic trends in nonstationary panel data are heterogeneous across individuals, Moon and Phillips (1999) show that the maximum likelihood estimator (MLE) of the local to unity parameter in the stochastic trend is inconsistent. They call this phenomenon, which arises because of the presence of an infinite number of nuisance parameters, an incidental trend problem because it is analogous to the well-known incidental parameter problem in dynamic panels when T is fixed[1]. To solve the incidental trend problem, Moon and Phillips (2000) propose various methods, including an iterative ordinary least squares (OLS) procedure and a double bias corrected estimator, and establish limit theories for these consistent estimators that can be used for statistical inference about the localizing parameter.

As a continuation of the two studies just mentioned, the present paper investigates a generalized method of moments (GMM) estimator of autoregressive roots near unity with panel data. We establish two moment conditions that form the basis for inference. The first moment condition is obtained by adjusting for the bias of the score function after conventional OLS detrending. The second moment condition is constructed by adjusting for the bias of the score function following quasi-difference (QD) detrending. Interestingly, the second moment condition is shown to correspond to the Gaussian projected score, where the projection is taken on the so-called Bhattacharya basis that has been studied recently in the conventional incidental parameter problem by Waterman and Lindsay (1996, 1998) and Hahn and Kuersteiner (2000). Unlike the conventional moment conditions used in estimating dynamic panel data models, these moment conditions do not suffer the weak instrument problem that is discussed, for example, in Kruiniger (2000) and Hahn *et al.* (2001).

Consistency of the GMM estimator is proved under the assumption that the localizing parameter takes a nonpositive value. This condition is not too restrictive because

[1] Lancaster(2000) provides a recent general survey of the incidental parameter problem in econometrics.

most econometric models consider non-explosive autoregressive regression models. Nevertheless, the restriction does matter in deriving the limiting distribution of the estimator because it is possible that the true parameter lies on the boundary of the parameter set. The most interesting case is, of course, the pure unit root case where the true localizing parameter is zero. In this case, in establishing the limiting distribution we cannot use the conventional approach that approximates the first order condition because the true parameter could be on the boundary of the parameter set. To avoid this difficulty, we use the approach that takes a quadratic approximation of the nonlinear objective function and optimize it on the parameter set (c.f. Andrews, 1999, for some recent developments of estimation and inference in boundary problems).

One of the most interesting findings in the present paper is that the GMM estimator has slower convergence rate than $\sqrt{n}$ when the time series components in the panel have unit roots (i.e., the true localizing parameter is zero), and the deterministic trends are linear. In this case the convergence rate is actually $O(n^{1/6})$ rather than $O(\sqrt{n})$. This slow convergence rate arises because of lack of information in the moment conditions when there is a unit root, i.e., at the point $c = 0$ in the space of the localizing parameter. It points to the continued difficulty of distinguishing unit roots from local alternatives in the presence of heterogeneous deterministic trends even when there is an infinity of additional data from a cross section.

The paper is organized as follows. Section 2 lays out the model and gives the basic assumptions that are maintained throughout the paper. We also discuss the empirical relevance of the model and the conventional moment conditions used in dynamic panels models of this type. In section 3 we introduce two new moment conditions and prove that the second of these moment conditions corresponds to a Gaussian projected score on the Bhattacharya basis. In Section 4 we establish consistency of the GMM estimator and obtain the limiting distributions of the GMM estimator when the true parameter is less than zero and equal to zero. The appendix contains technical derivations and proofs of the results in the main text.

2 Persistent Dynamic Panels

2.1 The Model and Assumptions

We study panel data that may show characteristics of time trends and persistent temporal shocks and whose dimension is large in both cross section (n) and time series (T) dimensions. To model such data, we extend the conventional dynamic panel model by taking the components formulation

$$z_{it} = \beta_i' g_{pt} + y_{it}, \tag{1}$$

where $g_{pt} = \left(t, t^2, ..., t^p\right)'$, the coefficients β_i are $p-$vectors that could be random, and the residuals y_{it} follow

$$y_{it} = \rho y_{it-1} + \varepsilon_{it}$$

with a common autoregressive coefficient ρ that is close to one defined in (4) below. Let $y_{i0} = z_{i0}$ be the panel observations at the initial time period. In (1), the first term $\beta_i' g_{pt}$ represents deterministic trends in the data, omitting an intercept because this is not consistently estimable from time series data when y_{it} is near integrated (*e.g.*, Phillips and Lee, 1996) and can be incorporated in the initial condition y_{i0}. Assuming that the coefficient β_i varies across i, we may treat the trends $\beta_i' g_{pt}$ as systematic individual effects or fixed effects in the panel. Since ρ is in the vicinity of unity, the components y_{it} have stochastic trends with persistent innovations ε_{it}.

The components model (1) can also be written in the more familiar format of an augmented regression form as

$$z_{it} = \rho z_{it-1} + \delta_i + \gamma_i' g_{pt} + \varepsilon_{it}, \tag{2}$$

where

$$\begin{aligned}
\delta_i &= \rho \beta_i' \iota_p, \ \iota_p = -\left(-1, (-1)^2, ..., (-1)^p\right)', \\
\gamma_i' &= \beta_i' \Upsilon(\rho), \ \Upsilon(\rho) \text{ is a } (p \times p) \text{ matrix depending on } \rho.
\end{aligned}$$

For example, when $p = 1$, the deterministic panel trends are linear and the augmented model (2) is

$$z_{it} = \rho \beta_i + (1-\rho)\beta_i t + \rho z_{it-1} + \varepsilon_{it}. \tag{3}$$

This linear trend model (3) is an extended version of the standard model for dynamic panels in which the individual effects are the incidental trends $\rho\beta_i + (1-\rho)\beta_i t$ and the autoregressive parameter is assumed to be close to one.[2]

The augmented format (2) has the drawback that linear regression leads to inefficient trend elimination, but the advantage that the detrended data is invariant to the trend parameters in (2). In the next section, we use the augmented formation (2) to define the first moment condition, and the component model (1) for the second moment condition.

To enable a rigorous development when ρ is close to one, we take the specific near unity formulation,

$$\rho = 1 + \frac{c}{T}, \tag{4}$$

or equivalently,

$$T(\rho - 1) = c,$$

in which the standardized deviation of the coefficient ρ from unity remains constant (c). In this case, the stochastic trends y_{it} are near integrated and they are characterized by the parameter c instead of the autoregressive coefficient ρ. The time series properties of the near integrated process y_{it} are well known from the nonstationary time series literature (*e.g.*, Phillips, 1987, and Stock, 1994). Recently, Hahn *el al.* (2001) and Kruiniger (2000) use the related specification $\rho = 1 + \frac{c}{n}$ to model an autoregressive coefficient near unity in a panel with large n and fixed T.

In a conventional time series autoregression (AR), the probabilistic features (and the asymptotics) are discontinuous with respect to the AR coefficient as it passes through unity. When $|\rho| < 1$, the process is stationary, reverts to its mean, converges to a steady state, and has no stochastic trend, whereas when $\rho = 1$, the process is nonstationary, not mean-reverting, and contains stochastic trends. Models with near unit roots as in (4) have probabilistic features that are continuous with respect to parameter c, while still retaining some of the implications of the three different cases: $\rho < 1$ ($c < 0$), $\rho = 1$ ($c = 0$), and $\rho > 1$ ($c > 0$). More specifically, a direct calculation shows that $Var(y_{it})$ increases at the rate t, regardless of the sign of the parameter c. Thus, y_{it} is nonstationary and has a

[2] For other examples of incidental trend models, see Section 11.2.1 of Wooldridge (2001) and the references therein.

stochastic trend regardless of the sign of the parameter c. On the other hand, when for $t = [Tr]$ with $0 < r \leq 1$, it is well known that

$$\frac{y_{it}}{\sqrt{T}} \Rightarrow J_c(r) \text{ as } T \to \infty, \tag{5}$$

where $J_c(r) = \int_0^r e^{c(r-s)}dW(s)$ is the Ornstein-Uhlenbeck process, and $W(s)$ is a Brownian motion (*e.g.*, Phillips, 1987). So, the marginal asymptotic distribution of the standardized process $\frac{y_{it}}{\sqrt{T}}$ is continuous in c. Also, when $c < 0$, the limit process $J_c(r)$ is stationary and mean reverting, while for $c = 0$, the limit process is Brownian motion. Thus, the standardized process $\frac{y_{it}}{\sqrt{T}}$ preserve some of the probabilistic implications implied by the conventional AR(1) model for $|\rho| < 1$ and $\rho = 1$. One benefit of the continuity property is that it is possible to produce confidence intervals for the AR coefficient ρ from estimates of c (Stock, 1991) even though consistent estimation of c from time series observations is not possible. In this paper we use notation c_0 to denote the true coefficient for c.

In later sections of the paper, as part of the asymptotic development, we need to verify some properties of complicated nonlinear functions of c that depend on the trend g_{pt}. These functions are so complicated that it is very difficult to establish analytic results under the general polynomial trend set up with $g_{pt} = (t, ..., t^p)'$. Instead, we rely on numerical methods for this part of the analysis. To assist the analytic development, we restrict our attention to the following two cases: (i) $g_{1t} = t$ and (ii) $g_{2t} = \left(t, t^2\right)'$. This restriction is hardly restrictive in practice because the linear and quadratic trends are the most widely used in empirical applications. The set up is formalized as follows:

Assumption 1 *(Trend Formulation)*
The polynomial trend in model (1) *is either (i)* $g_{1t} = t$ *or (ii)* $g_{2t} = \left(t, t^2\right)'$.

Assumption 2 *(Initial Condition) The initial observations* $z_{i0} = y_{i0}$ *are independent across* i *with* $\sup_i Ey_{i0}^4 < \infty$.

Assumption 3 *(Error Condition) The error terms* $\varepsilon_{it} \sim iid\left(0, \sigma^2\right)$ *across* i *and* t *with* $E\varepsilon_{it}^4 < \infty$ *and* ε_{it} *are independent of* y_{i0} *for all* i.

Assumption 4 *(Parameter Set)*
(a) The localizing parameter c *takes a value in a compact subset* $\mathbb{C} = [\ \bar{c}\ ,\ 0\ ] \subset \mathbb{R}$, *where* $\bar{c} < 0$.
(b) The true localizing parameter c_0 *is in the set* $\mathbb{C}_0 = (\ \bar{c}\ ,\ 0\ ]$.

Assumption 4(a) restricts the parameter set $\mathbb{C} = [\bar{c}\ ,\ 0]$ to be non-positive. This restriction is made because in most econometric applications, the cases $|\rho| < 1$ and $\rho = 0$ are of most interest. When the true parameter $c_0 = 0$, the model becomes nonstandard in the sense that the true parameter is on the boundary of the parameter set. Section 4.3 explores the implications of the boundary point aspect of this case.

The practical implication of the restriction of c_0 to $\mathbb{C} = [\bar{c}\ ,\ 0]$ is of some interest. One of the advantages of pooling is that consistent estimation of c_0 is achievable with panel data while it is not with a single time series. Take the case where the true value $\rho_0 = 1 + \frac{c_0}{T}$ lies in the interior of the compact interval $[1 + \frac{\bar{c}}{T}, 1]$ for some $\bar{c} < 0$. Theorem 3 below shows that the GMM estimator $\hat{c}$ is consistent for c_0 and has a limit normal distribution $\sqrt{n}(\hat{c} - c_0) \to_d Z$. The corresponding autoregressive coefficient estimate is $\hat{\rho} = 1 + \frac{\hat{c}}{T}$ and $\sqrt{n}T(\hat{\rho} - \rho_0) \to_d Z$. So, panel pooling affects the 'usual asymptotic distribution' of the autoregressive coefficient in near integrated models. The distribution is normal and lives in a shrunken neighbourhood within the compact set $[1 + \frac{\bar{c}}{T}, 1]$. In effect, $\hat{\rho}$ is distributed in an $O\left(\frac{1}{\sqrt{n}}\right)$ neighbourhood of the true value $\rho_0 \in [1+\frac{\bar{c}}{T}, 1]$. The main thrust of the present work is to put panel pooling to good effect, so that we get this increased precision about the value of ρ_0 even when it is already in the locality of unity. To the extent that many economic time series are near integrated but also may have idioscyncratic deterministic trend components, this process increases precision in estimation, confidence interval construction and local discriminatory power at unity for ρ_0 (where it increases from $O\left(T^{-1}\right)$ to $O\left(T^{-1}n^{-\frac{1}{6}}\right)$ - see section 4.3), but also near unity (where it increases from $O\left(T^{-1}\right)$ to $O\left(T^{-1}n^{-\frac{1}{2}}\right)$ - see section 4.2).

Assumption 5 *$(n, T \to \infty)$ with $\frac{\log n}{\log T} < 2 + O\left(\frac{1}{\log T}\right)$.*

Assumption 5 restricts the relative size of the cross-section and temporal dimensions in the panel in the asymptotic theory. It is assumed that n increases at a lesser speed than T^2 as $T \to \infty$. So if $n = T^{\alpha}$ then $\alpha < 2$, thereby allowing for panels in which n and T are of comparable size satisfying $\frac{n}{T} \to \kappa$, $0 < \kappa < \infty$, as in Hahn and Kuersteiner, 2000), and some panels where the size of the cross section dominates the time series count, i.e., $\frac{n}{T} \to \infty$. This assumption is required for the derivation of the limit distribution of the estimator of c in Section 4. Consistency of the estimator does not require this restriction.

2.2 Discussion of the Model

2.2.1 Empirical Relevance in Modeling Firm Size

In the empirical industrial organization literature on firm size, dynamic panel models have been used widely to describe firm growth in terms of a simple formulation that follows the spirit of Gibrat's law of proportional effect (Gibrat, 1931). Gibrat's law states that the expected value of the increment in a firm's size each time period is proportional to the current size of the firm. Let Z_{it} denote the size of firm i at time t and e_{it} denote the (stochastic) proportionate rate of growth of firm i between time t and $t-1$. Then Gibrat's law is formalized as

$$Z_{it} - Z_{it-1} = Z_{it-1}e_{it}.$$

(e.g., Steindle, 1965, and Sutton, 1997). Let $z_{it} = \log Z_{it}$ and, using the approximation $\log(1+a) \approx a$ for small a, we may write the proportional law in autoregressive form as

$$z_{it} = z_{it-1} + e_{it}, \tag{6}$$

which McCloughan (1995) calls Gibrat's process – a law in which the growth rate of a firm is independent of its initial size. When $e_{it} \sim iid\left(\mu, \sigma_e^2\right)$, set $\varepsilon_{it} = e_{it} - \mu$, and then we may rewrite Gibrat's process (6) in the component form

$$\begin{aligned} z_{it} &= \mu t + y_{it}, \\ y_{it} &= y_{it-1} + \varepsilon_{it}, \end{aligned} \tag{7}$$

which is a special case of model (1). The panel model (1) (or (2)) can therefore be interpreted as a generalized version of Gibrat's process when applied to firm growth data. To motivate (1), we now discuss more detailed features of the model in the context of this application and indicate how we can explain in terms of this model (1) some empirical findings relating to the violation of Gibrat's law.

First we consider several well known implications of Gibrat's law. For T large, by virtue of the functional central limit theorem applied to $T^{-1/2}z_{it}$ in (6), the distribution of firm size Z_{it} is approximately lognormal and therefore skewed. Evidence supporting lognormality and skewness of firm size distributions has been reported in many past empirical papers. The more recent literature argues that firm size distributions do not follow a 'typical' distributional shape but nevertheless have skewness as their central characteristic (e.g., Sutton, 1997, and Schmalensee, 1989). Next, observing that $\Delta Var\left(z_{it}\right) := Var(z_{it}) - Var(z_{it-1}) = \sigma^2$ in Gibrat's process (7), Prais (1976) and more recently McCloughan (1995) claimed that the proportionate model implies that size inequalities between firms will increase at a constant rate, and the rate of concentration will be greater the higher the variance of the growth distribution, σ^2. For this reason, McCloughan (1995) calls $\Delta Var\left(z_{it}\right) = \sigma^2$ Gibrat's effect. For other implications of Gibrat's law, see Lucas (1978) and Sutton (1997).

Over the last two decades, many studies have found empirical violations of Gibrat's law. These are reviewed and classified in McCloughan (1995). For example, Evans (1987) found through a cross-sectional analysis that firm growth decreases with firm age and firm size, which is claimed to be consistent with the prediction made by the theory in Jovanovic (1982). More recently, Hall and Mairesse (2000) investigated the time series properties of several variables in firm-level panel data that are related to the growth of firms. One of their findings was that the growth rates of these variables vary widely across firms. Another was that the sample serial correlations typically decay very slowly.

In what follows, we compare the implications of the panel model (1) with those of Gibrat's process and consider possible explanations, in terms of (1), of the empirical violations of the law.

(i) According to (1) and (4), the stochastic shock in the logarithm of firm size is nearly integrated and involves the parameter c. As discussed earlier, for T large this formulation ensures that the distribution of firm size is asymptotically lognormal, as indicated by Gibrat's law[3]. Also, since the autoregressive coefficient ρ is close to unity for large T, the serial correlations of z_{it} will not decay. These properties coincide with what Hall and Mairesse (2000) observe in their firm panel data.

(ii) Since the random growth rate process is $\Delta z_{it} = \delta_i + \gamma_i' g_{pt} + \frac{c}{T} z_{it-1} + \varepsilon_{it}$, we have $\frac{\partial \Delta z_{it}}{\partial z_{it-1}} = \frac{c}{T} < 0$ if $c < 0$. Thus, firm growth decreases as firm size increases, giving the size effect.

(iii) If the deterministic trends are quadratic and the coefficient of the quadratic coefficient is negative, we have the age effect because $\frac{\partial E(\Delta z_{it})}{\partial t} = 2\beta_{i2} < 0$ when $z_{i0} = y_{i0} = 0$.

(iv) A simple calculation shows that

$$Var\left(z_{it}\right) = \sigma^2 \left(1 + \rho^2 + ... + \rho^{2(t-1)}\right) \sim \sigma^2 \sum_{s=0}^{t-1} \exp\left(2c\frac{s}{T}\right) \tag{8}$$

when y_{i0} is nonrandom. So,

$$\Delta Var\left(z_{it}\right) \simeq \sigma^2 \exp\left(2c\frac{t-1}{T}\right). \tag{9}$$

[3] In empirical studies, a commonly used model for representing a generalized Gibrat process is the conventional AR(1) model

$$z_{it} = \mu + \rho z_{it-1} + e_{it},$$

e.g., McCloughan, 1995, and Hall and Mairesse, 2000. In this case, when $-1 < \rho < 1$, the size distribution depends on the distribution of the error e_{it} and skewess of the size distribution is not guaranteed. By contrast, skewness is guaranteed, at least asymptotically, in the near integrated case.

Equations (8) and (9) indicate that inequalities in firm growth grow over time and the rate of change in the inequality depends on the value of the growth parameter c.

(v) Model (2) introduces individual specific effects through the trend coefficients β_i and the initial conditions z_{i0}. These effects may explain certain heterogeneous factors such as the x-inefficiency of some firms.

In view of these characteristics, the panel model (2) helps to bridge the gap between the pure form of Gibrat's law and the empirical evidence indicating certain systematic violations of the law. In addition, consistent estimation of the systematic growth parameter c makes it possible to measure the size effect and changes in growth inequality or firm concentration.

2.2.2 Conventional Moment Conditions and Weak Instruments

In recent years, one of the most widely used methods of estimating dynamic panel regression models with fixed effects, such as

$$z_{it} = (1-\rho)\beta_i + \rho z_{it-1} + \varepsilon_{it},$$

is to utilize the moment conditions implied by the assumptions imposed on the model. (For details and references, see the recent survey by Arellano and Honore, 2000). Among these moment conditions[4], the moment restrictions known as the basic moment conditions, viz.,

$$E\left(z_{is}\Delta\varepsilon_{it}\right) = 0 \text{ for } s = 0, ..., t-2, \tag{10}$$

are the most widely applied. This section considers the properties of the basic form of moment conditions implied by the model (1).

For simplicity, we consider only the linear trend case, i.e.,

$$z_{it} = \rho\beta_i + (1-\rho)\beta_i t + \rho z_{it-1} + \varepsilon_{it}.$$

Due to the presence of the incidental trends, instead of (10) the basic moment conditions in the model 1 are

$$E\left(z_{is}\Delta^2\varepsilon_{it}\right) = 0 \text{ for } s = 0, ..., t-3, \tag{11}$$

which provide the orthogonality conditions in an instrumental variable regression of

$$\Delta^2 z_{it} = \rho\Delta^2 z_{it-1} + \Delta^2\varepsilon_{it}, \tag{12}$$

with instrumental variables z_{is}, $s = 0, ..., t-3$. Notice from (12) that the basic moment conditions (11) are linear in the parameter ρ. To evaluate the orthogonality conditions in (11) in terms of their information content, we calculate the first derivative of the moments with respect to the parameter ρ. Then,

$$\frac{\partial E\left(z_{is}\left(\Delta^2 z_{it} - \rho\Delta^2 z_{it-1}\right)\right)}{\partial\rho} = E\left(z_{is}\Delta^2 z_{it-1}\right), \; s = 0, ..., t-3,$$

which is the covariance between the instrumental variable z_{is} and regressor in (12).

First, when $\rho = 1$, as is well known in the conventional dynamic panel model, the information content of the basic moment conditions is zero because $E\left(z_{is}\Delta^2 z_{it-1}\right) = 0$

[4] Other types of moment conditions used in conventional dynamic panel models are derived under extra assumptions on the initial conditions, the fixed effect parameter, and the covariance stationarity restrictions.

for $s = 0, ..., t-3$, and so the instruments are not correlated with the regressors of (12). In this case, the parameter ρ is not identifiable from the basic moment conditions.

When $\rho = 1 + \frac{c}{T}$ with $c < 0$, a direct calculation shows that

$$E\left(z_{is}\Delta^2 z_{it-1}\right) = \left(\frac{c}{T}\right)^2 E\left(\beta_i s + y_{is}\right)\left(2\beta_i + y_{t-3}\right) \sim O\left(\frac{1}{\sqrt{T}}\right).$$

This implies that if the coefficient is near unity as in (4) and the time dimension is large, the information in the basic moment conditions is small, and the instruments z_{is} become weak[5]. In consequence, GMM estimation involving only conventional moment conditions cannot estimate the model consistently. This has become an issue of some importance in empirical studies of dynamic panel regression.

We therefore proceed to consider a different approach that augments the information content of the basic moment conditions. This approach arises from a consideration of the incidental trends problem, which we now discuss.

2.2.3 Incidental Trends Problem

Two recent papers by the authors (Moon and Phillips, 1999 & 2000) find that an incidental trend problem arises in estimating the local to unity parameter c in the panel regression model (1) or (2) where infinite number of nuisance parameters are present. Moon and Phillips (1999) show that the score of the (pseudo) likelihood that concentrates out the incidental trend parameters β_i are biased (even in the limit). In consequence, the maximum likelihood estimator (MLE) of c in (1) is inconsistent. Also, according to Moon and Phillips (2000), when the incidental trend parameters β_i are eliminated by ordinary least squares (OLS) projection, the normal equation of the pooled OLS estimator is biased as well (even in the limit) due to the correlation between the OLS detrended error term and the OLS detrended regressor. It follows that the pooled OLS estimator of c obtained from OLS detrended data is also biased.

3 Moment Conditions

We now propose an estimation procedure that eliminates the effects of the incidental trend problem. The approach involves GMM and minimizes a distance criterion based on a vector of sample moment functions. The moment conditions are designed to have a limit that will identify the true localizing parameter c_0 even in the presence of the incidental trend coefficients β_i.

The principle we use for contructing moment conditions with this property is to adjust for the bias that arises in the usual score functions, the latter being explained in Moon and Phillips (1999, 2000). The adjustments are based on formulae from the explicit computation of the bias functions. The next subsections explain these constructions in detail.[6]

[5] When T is finite and n is large, Hahn *et al* (2001) and Kruiniger (2000) investigate similar weak intrument effects in the conventional dynamic panel regression model with near unit root specifications $\rho = 1 + \frac{c}{n}$.

[6] The principle employed can be applied to any biased score function for transformed data that is invariant to the incidental trend coefficients. A referee suggested such a biased score function that leads to an alternative moment condition that turns out to be closely related to the first moment condition described below. In view of this similarity and space constraints, we do not investigate this alternative moment condition here.

3.1 The First Moment Condition

The following notation is defined to assist with the analysis of the trend function asymptotics and it will be used throughout the rest of the paper. Let

$$\begin{aligned}
\tilde{\gamma}_i &= (\delta_i, \gamma_i')', \\
\tilde{g}_{pt} &= \left(1, g_{pt}'\right)', \qquad g_p(r) = (r, ..., r^p)' \text{ with } r \in [0,1], \qquad \tilde{g}_p(r) = \left(1, g_p(r)'\right)', \\
G_{p,T} &= \left(g_{p1}', ..., g_{pT}'\right)', \qquad G_{p,T,-1} = \left(g_{p0}', ..., g_{pT-1}'\right)', \qquad \tilde{G}_{p,T} = \left(\tilde{g}_{p1}', ..., \tilde{g}_{pT}'\right)', \\
\tilde{M}_{p,T} &= I_T - \tilde{G}_{p,T}\left(\tilde{G}_{p,T}'\tilde{G}_{p,T}\right)^{-1}\tilde{G}_{p,T}', \\
D_{p,T} &= diag\left(T, ..., T^p\right), \qquad \tilde{D}_{p,T} = diag\left(1, D_{p,T}\right), \\
h_{pT}(t,s) &= g_{pt}' D_{p,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} D_{p,T}^{-1} g_{pt} g_{pt}' D_{p,T}^{-1}\right)^{-1} D_{p,T}^{-1} g_{ps}, \\
\tilde{h}_{pT}(t,s) &= \tilde{g}_{pt}' \tilde{D}_{p,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} \tilde{D}_{p,T}^{-1} \tilde{g}_{pt} \tilde{g}_{pt}' \tilde{D}_{p,T}^{-1}\right)^{-1} \tilde{D}_{p,T}^{-1} \tilde{g}_{ps}, \\
h_p(r,s) &= g_p'(r)\left(\int_0^1 g_p(r) g_p(r)' dr\right)^{-1} g_p(s), \\
\tilde{h}_p(r,s) &= \tilde{g}_p'(r)\left(\int_0^1 \tilde{g}_p(r) \tilde{g}_p(r)' dr\right)^{-1} \tilde{g}_p(s).
\end{aligned}$$

Write $z_i = (z_{i1}, ..., z_{iT})'$, $z_{i,-1} = (z_{i0}, ..., z_{iT-1})'$, and $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$. Define

$$\underset{\sim}{z}_{i} = \tilde{M}_{p,T} z_i, \qquad \underset{\sim}{\varepsilon}_{i} = \tilde{M}_{p,T}\varepsilon_i, \qquad \underset{\sim}{z}_{i,-1} = \tilde{M}_{p,T} z_{i,-1}.$$

Then, it is straightforward to show that

$$\underset{\sim}{z}_{i} = \underset{\sim}{y}_{i} \text{ and } \underset{\sim}{z}_{i,-1} = \underset{\sim}{y}_{i,-1},$$

where

$$\underset{\sim}{y}_{i} = \tilde{M}_{p,T} y_i, \qquad \underset{\sim}{y}_{i,-1} = \tilde{M}_{p,T} y_{i,-1},$$

$y_i = (y_1, ..., y_T)'$, and $y_{i,-1} = (y_0, ..., y_{T-1})'$. Let

$$\left(\underset{\sim}{z}_{i,-1}\right)_t = z_{it-1} - \frac{1}{T}\sum_{s=1}^{T}\tilde{h}_{pT}(t,s)\, z_{is-1}$$

be the t^{th} element of $\underset{\sim}{z}_{i,-1}$.

One straightforward procedure of estimating c_0 (equivalently $\rho_0 = 1 + \frac{c_0}{T}$) is to eliminate the unknown trends $\delta_i + \gamma_i' g_{pt}$ by taking OLS regression residuals and then apply pooled least squares with an appropriate bias correction for the serial correlation of ε_{it}. We call this method iterative OLS. However, as noted in Moon and Phillips (2000), this iterative OLS procedure yields an inconsistent estimator of c_0 due to a nondegenerating asymptotic bias between the detrended regressor and the detrended error term. The first moment condition is obtained simply by subtraction of this asymptotic bias term in iterative OLS from the normal equation for the OLS detrended data.

More specifically, write the model (2) in vector notation as

$$z_i = \rho_0 z_{i,-1} + \tilde{G}_{p,T}\tilde{\gamma}_i + \varepsilon_i, \quad \rho_0 = 1 + \frac{c_0}{T}. \tag{13}$$

Applying $\tilde{M}_{pT}$ to both sides, we have

$$\underset{\sim}{z}_i = \rho_0 \underset{\sim}{z}_{i,-1} + \underset{\sim}{\varepsilon}_i, \tag{14}$$

where $\underset{\sim}{z}_i, \underset{\sim}{z}_{i,-1}$, and $\underset{\sim}{\varepsilon}_i$ are OLS detrended versions of z_i, $z_{i,-1}$, and ε_i, respectively. In general, the detrended regressor vector $\underset{\sim}{z}_{i,-1}$ and the detrended error vector $\underset{\sim}{\varepsilon}_i$ are correlated and the first moment condition is found by correcting for the bias due to the correlation between $\underset{\sim}{z}_{i,-1}$ and $\underset{\sim}{\varepsilon}_i$.

Before we define the first moment condition, we discuss the estimation of the error variance σ^2. Let $\hat{\rho}$ be the pooled OLS estimator of (14),

$$\hat{\rho} = \frac{\sum_{i=1}^n \underset{\sim}{z}_i' \underset{\sim}{z}_{i,-1}}{\sum_{i=1}^n \underset{\sim}{z}_{i,-1}' \underset{\sim}{z}_{i,-1}},$$

and $\hat{\underset{\sim}{\varepsilon}}_i$ be the OLS residual of (14), $\hat{\underset{\sim}{\varepsilon}}_i = \underset{\sim}{z}_i - \hat{\rho}\underset{\sim}{z}_{i,-1}$. The estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{nT}\sum_{i=1}^n \hat{\underset{\sim}{\varepsilon}}_i' \hat{\underset{\sim}{\varepsilon}}_i. \tag{15}$$

The first moment condition formula is defined as

$$\begin{aligned} m_{1,iT}(c) &= \frac{1}{T}\left(\underset{\sim}{z}_i - \left(1 + \frac{c}{T}\right)\underset{\sim}{z}_{i,-1}\right)' \underset{\sim}{z}_{i,-1} + \hat{\sigma}^2 \omega_{pT}(c) \qquad (16) \\ &= \frac{1}{T}\underset{\sim}{\varepsilon}_i' \underset{\sim}{y}_{i,-1} - (c - c_0)\frac{1}{T^2}\underset{\sim}{y}_{i,-1}' \underset{\sim}{y}_{i,-1} + \hat{\sigma}^2\omega_{pT}(c) \\ &= \frac{1}{T}\sum_{t=1}^T \varepsilon_{it} y_{i,t-1} - \frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T \varepsilon_{it} y_{i,s-1}\tilde{h}_{pT}(t,s) - (c - c_0)\frac{1}{T^2}\sum_{t=1}^T \left(\underset{\sim}{y}_{i,-1}\right)_t^2 + \hat{\sigma}^2\omega_{pT}(c), \end{aligned}$$

where

$$\omega_{pT}(c) = \frac{1}{T^2}\sum_{t=2}^T\sum_{s=1}^{t-1}\left[\left(1 + \frac{c}{T}\right)^T\right]^{\frac{t-s-1}{T}} \tilde{h}_{pT}(t,s),$$

and $\underset{\sim}{\varepsilon}_{it}$ and $(\underset{\sim}{y}_{i,-1})_t$ are the t^{th} elements of $\underset{\sim}{\varepsilon}_i$ and $\underset{\sim}{y}_{i,-1}$, respectively. The terms $\hat{\sigma}^2\omega_{pT}(c)$ corrects for the bias that arises from the correlation between $\underset{\sim}{\varepsilon}_{it}$ and $(\underset{\sim}{y}_{i,-1})_t$.

This first moment condition was utilized in the double bias corrected estimator of Moon and Phillips (2000). Suppose that a preliminary consistent estimate of c is available. By linearizing the bias correction term $\hat{\sigma}^2\omega_{pT}(c)$ in (16) around the consistent estimate of c, we may approximate the first moment condition as a linear function of $(c - c_0)$. The double bias corrected estimator solves this linear approximation equation. In this paper, to estimate the parameter c, we continue to use the nonlinear moment condition (16) rather than work with a linear approximation.

3.2 The Second Moment Condition

Before discussing the second moment condition, we introduce some additional notation. Let

$$\Delta_c = \left(1 - \left(1 + \frac{c}{T}\right) L\right),$$

where L is the lag operator. Define

$$\begin{aligned}
F_{p,T} &= diag\left(1, T, ..., T^{p-1}\right) = \frac{1}{T} D_{p,T}, \qquad \widehat{\Delta_c g_{pt}} = F_{p,T}^{-1} \Delta_c g_{pt} \\
\dot{g}_p(r) &= \frac{d}{dr} g_p(r) = \left(1, 2r, ..., pr^{p-1}\right)', \qquad \dot{g}_{pc}(r) = \dot{g}_p(r) - c g_p(r), \\
A_{pT}(c) &= \frac{1}{T} \sum_{t=1}^{T} \widehat{\Delta_c g_{pt}} \widehat{\Delta_c g_{pt}}', \qquad A_p(c) = \int_0^1 \dot{g}_{pc}(r) \dot{g}_{pc}(r)' dr, \\
B_{pT}(c) &= \frac{1}{T} \sum_{t=1}^{T} \widehat{\Delta_c g_{pt}} g_{pt-1}' D_{pT}^{-1}, \qquad B_p(c) = \int_0^1 \dot{g}_{pc}(r) g_p(r)' dr.
\end{aligned}$$

The second moment condition is obtained from the efficiently detrended regression equation. According to Canjels and Watson (1997) and Phillips and Lee (1996), the trend coefficient in the model (1) can be efficiently estimated in the time domain by employing a procedure that amounts to quasi-differencing the data with the operator Δ_c. That is, when the localizing parameter c is known, an estimator of β_i in (1) that is asymptotically more efficient than the OLS estimator of β_i is

$$\hat{\beta}_i(c) = \left(\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c g_{pt}'\right)^{-1} \left(\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c z_{it}\right).$$

Denoting $y_{it}(\beta_i) = z_{it} - \beta_i' g_{pt}$, we write

$$\hat{\beta}_i(c) = \beta_i + \left(\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c g_{pt}'\right)^{-1} \left(\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c y_{it}(\beta_i)\right).$$

Define $\varepsilon_{it}(c, \beta_i) = \Delta_c z_{it} - \beta_i' \Delta_c g_{pt}$.

The second moment function $m_{2,iT}(c)$ is defined as

$$m_{2,iT}(c) = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}\left(c, \hat{\beta}_i(c)\right) y_{it-1}\left(\hat{\beta}_i(c)\right) + \hat{\sigma}^2 \lambda_{pT}(c), \tag{17}$$

where

$$\lambda_{pT}(c) = \frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \left[\left(1 + \frac{c}{T}\right)^T\right]^{\frac{t-s-1}{T}} \widehat{\Delta_c g_{ps}}' A_{pT}(c)^{-1} \widehat{\Delta_c g_{pt}}.$$

Notice that $y_{it-1}\left(\hat{\beta}_i(c)\right)$ is the induced residual of the regression equation $z_{it} = \beta_i' g_{pt} + y_{it}$ and $\varepsilon_{it}\left(c, \hat{\beta}_i(c)\right)$ is the residual of the quasi-differenced equation $\Delta_c z_{it} = \beta_i' \Delta_c g_{pt} + \Delta_c y_{it}$. In the second moment function $m_{2,iT}(c)$ we correct for the asymptotic bias of $\frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}\left(c, \hat{\beta}_i(c)\right) y_{it-1}\left(\hat{\beta}_i(c)\right)$ by substracting off the estimate $\hat{\sigma}^2 \lambda_{pT}(c)$.

As mentioned above, Moon and Phillips (1999) showed that the Gaussian MLE of the panel regression model (3) with linear incidental trends is inconsistent. The main reason for inconsistency of the MLE is that the concentrated score of the (standardized) Gaussian likelihood function, $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}(c,\hat{\beta}_i(c))y_{it-1}(\hat{\beta}_i(c))$, has non-zero mean in the limit. In the second moment formulation of $m_{2,iT}(c)$, by subtracting off the estimate $\hat{\sigma}^2\lambda_{pT}(c)$, we eliminate the asymptotic bias of the concentrated Gaussian score function.

3.3 The Relationship between the Second Moment Condition and the Projected Score

This section shows that the second moment function $m_{2,iT}(c)$ is a projected score of the panel regression model (1) under Gaussian errors. Suppose that the error process ε_{it} in the model (1) is an iid standard normal process across i and over t. For convenience we assume that $z_{i0} = y_{i0} = 0$ for all i.

Under general regularity conditions, it is well known that the asymptotic properties of the MLE, and most notably its consistency, are closely related to the unbiasedness of the score function at the true parameter. However, it is also well known that in dynamic panel regression models with incidental parameters the MLE is not consistent (*e.g.*, see Neyman and Scott, 1948, and Nickel, 1981) as $n \to \infty$ with T fixed. Recently, Moon and Phillips (1999) found that this incidental parameter problem also arises in nonstationary panel regression models with incidental trends and roots local to unity when both $n \to \infty$ and $T \to \infty$, covering models such as (1).

The main reason for the inconsistency of the MLE is that the score function in an incidental trend model has a bias at the true parameter. Therefore, in order to obtain a consistent estimate, one needs to correct for the bias in the score function. One recently investigated method to correct for this bias is to use a projected score function, where the projection is taken onto the so-called Bhattacharyya basis. The resulting approach is called 'a projected score method'.

To define a projected score in the present case, assuming that ε_{it} are iid $N(0,1)$, we introduce the following notation. Denote the joint density of z_i

$$f_i(z_i;c,\beta_i) = \left(\frac{1}{\sqrt{2\pi}}\right)^T \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\left(\Delta_c z_{it} - \beta_i'\Delta_c g_{pt}\right)^2\right) \tag{18}$$

and set

$$\begin{aligned} U_{1i}(c,\beta_i) &= \frac{\partial f_i/\partial c}{f_i}, \qquad V_{1i}(c,\beta_i) = \frac{\partial f_i/\partial \beta_i}{f_i}, \\ V_{2i}(c,\beta_i) &= \frac{\frac{\partial^2 f_i}{\partial\beta_i\partial\beta_i'}}{f_i}, \qquad V_i(c,\beta_i) = \begin{pmatrix} V_{1i}(c,\beta_i) \\ D_p^+ vecV_{2i}(c,\beta_i)\end{pmatrix}, \end{aligned}$$

where $D_p^+ = \left(D_p'D_p\right)^{-1}D_p'$ and D_p is the duplication matrix.

For convenience, we mix notation U_{1i}, V_i, and V_{ki} for $U_{1i}(c,\beta_i)$, $V_i(c,\beta_i)$, and $V_{ki}(c,\beta_i)$, $k = 1,2$, respectively. In the statistics literature, V_{1i} and V_{2i} are known as the Bhattacharyya basis of order 1 and 2, respectively (see Bhattacharyya, 1946 and Waterman and Lindsay, 1996). The projected score U_{2i} is defined as the residual in the L_2- projection of U_{1i} on the closed linear space spanned by V_{1i} and V_{2i}, *i.e.*,

$$U_{2i} = U_{1i} - \xi_1'V_{1i} - \xi_2'D_p^+(vecV_{2i}). \tag{19}$$

Recently, using the projected score method, Waterman and Lindsay (1998) and Hahn (1998) were able to solve similar nuisance parameter problems in the classical Neyman and Scott panel regression model and in a simple dynamic panel regression model with fixed effects, respectively.

When the joint density of z_i is given in (18), U_{1i}, V_{1i}, and V_{2i} are found to be

$$\begin{aligned} U_{1i}(c, \beta_i) &= \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}(c, \beta_i)\, y_{it-1}(\beta_i), \\ V_{1i}(c, \beta_i) &= \sum_{t=1}^{T} \varepsilon_{it}(c, \beta_i)\, \Delta_c g_{pt}, \\ V_{2i}(c, \beta_i) &= -\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c g_{pt}' + \left(\sum_{t=1}^{T} \varepsilon_{it}(c, \beta_i)\, \Delta_c g_{pt}\right)\left(\sum_{t=1}^{T} \varepsilon_{it}(c, \beta_i)\, \Delta_c g_{pt}\right)'. \end{aligned}$$

Some algebra verifies that $EV_{1i}U_{1i} = 0$ and $EV_{1i}\left[D_p^{+}(vecV_{2i})\right]' = 0$[7]. Hence, the two L_2- projection coefficients ξ_1 and ξ_2 in (19) are given by

$$\xi_1 = \left[EV_{1i}V_{1i}'\right]^{-1} EV_{1i}U_{1i} = 0,$$

and

$$\xi_2 = \left[D_p^{+} E(vecV_{2i})(vecV_{2i})' D_p^{+\prime}\right]^{-1} D_p^{+} E(vecV_{2i})\, U_{1i}.$$

Next

$$\begin{aligned} & E(vecV_{2i})(vecV_{2i})' \\ = & \sum_{t=1}^{T}\sum_{s=1}^{T}\left(\Delta_c g_{pt}\Delta_c g_{pt}' \otimes \Delta_c g_{ps}\Delta_c g_{ps}'\right) + \sum_{t=1}^{T}\sum_{s=1}^{T}\left(\Delta_c g_{pt}\Delta_c g_{ps}' \otimes \Delta_c g_{ps}\Delta_c g_{pt}'\right), \end{aligned}$$

and

$$\begin{aligned} & E(vecV_{2i})\, U_{1i} \\ = & \frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left[\Delta_c g_{pt} \otimes \Delta_c g_{ps} + \Delta_c g_{ps} \otimes \Delta_c g_{pt}\right]\left[\left(1+\frac{c}{T}\right)^{T}\right]^{\frac{t-s-1}{T}}. \end{aligned}$$

Therefore, the projected score $U_{2i}(c, \beta_i)$ is

$$\begin{aligned} & U_{2i}(c, \beta_i) \\ = & \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}(c, \beta_i)\, y_{it-1}(\beta_i) + \xi_2' D_p^{+} \sum_{t=1}^{T}\left(\Delta_c g_{pt} \otimes \Delta_c g_{pt}\right) \\ & - \xi_2' D_p^{+}\left(\sum_{t=1}^{T} \varepsilon_{i,t}(c, \beta_i)\, \Delta_c g_{pt}\right) \otimes \left(\sum_{s=1}^{T} \varepsilon_{is}(c, \beta_i)\, \Delta_c g_{ps}\right), \end{aligned}$$

where

$$\begin{aligned} & \xi_2 \\ = & \left[\sum_{t=1}^{T}\sum_{s=1}^{T} D_p^{+}\left\{\left(\Delta_c g_{pt}\Delta_c g_{pt}' \otimes \Delta_c g_{ps}\Delta_c g_{ps}'\right) + \left(\Delta_c g_{pt}\Delta_c g_{ps}' \otimes \Delta_c g_{ps}\Delta_c g_{pt}'\right)\right\}\left(D_p^{+}\right)'\right]^{-1} \\ & \times \frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1} D_p^{+}\left[\Delta_c g_{pt} \otimes \Delta_c g_{ps} + \Delta_c g_{ps} \otimes \Delta_c g_{pt}\right]\left[\left(1+\frac{c}{T}\right)^{T}\right]^{\frac{t-s-1}{T}}. \end{aligned}$$

[7] The expectation is evaluated at the parameters c and β_i.

Since β_i in U_{2i} is unknown, we replace it with the estimate

$$\hat{\beta}_i(c) = \left(\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c g'_{pt}\right)^{-1} \left(\sum_{t=1}^{T} \Delta_c g_{pt} \Delta_c z_{it}\right).$$

Then, we have the following concentrated projected score

$$U_{2i}\left(c, \hat{\beta}_i(c)\right) = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}\left(\hat{\beta}_i(c), c\right) y_{it-1}\left(\hat{\beta}_i(c)\right) + \xi_2' D_p^+ \sum_{t=1}^{T} \left(\Delta_c g_{pt} \otimes \Delta_c g_{pt}\right), \quad (20)$$

because $\sum_{t=1}^{T} \varepsilon_{it}\left(c, \hat{\beta}_i(c)\right) \Delta_c g_{pt} = 0.$

When the error process ε_{it} is iid$(0,1)$ across i and over t, the second moment function $m_{2,iT}(c)$ is

$$m_{2,iT}(c) = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}\left(c, \hat{\beta}_i(c)\right) y_{it-1}\left(\hat{\beta}_i(c)\right) + \lambda_{pT}(c).$$

The following lemma states that the bias correction term $\lambda_{pT}(c)$ in $m_{2,iT}(c)$ is equivalent to $\xi_2' D_p^+ \sum_{t=1}^{T} \left(\Delta_c g_{pt} \otimes \Delta_c g_{pt}\right)$. So, the second moment function actually corresponds to the concentrated projected score function of the Gaussian model. The proof of the lemma is in the appendix.

Lemma 1 *(Equivalence) Suppose that the errors in model 1 are iid normal with mean zero and variance 1 across i and over t and* $y_{i0} = z_{i0} = 0$ *for all i. Then, the second moment condition* $m_{2,iT}(c)$ *is equivalent to the concentrated projected score function* $U_{2i}\left(c, \hat{\beta}_i(c)\right)$.

4 GMM Estimation and Asymptotics

This section investigates the asymptotic properties of a GMM estimator of c that is based on the two moment conditions introduced in the previous section. Let

$$M_{nT}(c) = \frac{1}{n}\sum_{i=1}^{n} m_{iT}(c),$$

where $m_{iT}(c)' = (m_{1,iT}(c), m_{2,iT}(c))$, and $m_{1,iT}(c)$ and $m_{2,iT}(c)$ are defined in (16) and (17), respectively. Let $\hat{W}$ be a (2×2) random weight matrix and B_{nT} be a sequence of real numbers that converges to infinity as $(n, T \to \infty)$. The GMM estimator $\hat{c}$ for the unknown parameter c_0 in (13) is defined as the extremum estimator for which

$$Z_{nT}(\hat{c}) \leq \min_{c\in\mathbb{C}} Z_{nT}(c) + o_p\left(B_{nT}^{-2}\right), \quad (21)$$

where

$$Z_{nT}(c) = M_{nT}(c)' \hat{W} M_{nT}(c).$$

Since the objective function $Z_{nT}(c)$ is continuous in c and the parameter set $\mathbb{C} = [\bar{c}, 0]$ for some $\bar{c} < 0$ that contains c_0 is assumed to be compact, it is possible to find a global minimum of $Z_{nT}(c)$ over $\mathbb{C}$. The main purpose in allowing for an $o_p\left(B_{nT}^{-1}\right)$ deviation bound from the global minimum $\min_{c\in\mathbb{C}} Z_{nT}(c)$ is to reduce the computational burden and allow for potential numerical computational errors within a range of $o_p\left(B_{nT}^{-1}\right)$. Later in the paper, depending on the convergence rate of $\hat{c}$ to c_0, we will determine the sequence B_{nT}.

4.1 Consistency of the GMM Estimator

Define

$$M(c) = \begin{pmatrix} m_1(c) \\ m_2(c) \end{pmatrix},$$

where

$$m_1(c) = \omega_p(c) - \omega_p(c_0) - (c - c_0)\,\psi_p(c_0),$$

$$\begin{aligned}
\omega_p(c) &= \int_0^1 \int_0^r e^{c(r-s)} \tilde{h}_p(r,s)\, ds dr, \\
\psi_p(c_0) &= -\frac{1}{2c_0}\left(1 + \frac{1}{2c_0}\left(1 - e^{2c_0}\right)\right) \\
&\quad - \int_0^1 \int_0^1 e^{c_0(r+s)} \frac{1}{2c_0}\left(1 - e^{-2c_0(r \wedge s)}\right) \tilde{h}_p(r,s)\, ds dr,
\end{aligned}$$

and

$$\begin{aligned}
& m_2(c) \\
= \; & -(c - c_0)\left(\int_0^1 \int_0^r e^{2c_0(r-s)} ds dr\right) \\
& + (c - c_0) \int_0^1 \int_0^1 \int_0^{r \wedge s} e^{c_0(r+s-2v)} \dot{g}_{pc}(s)' A_p(c)^{-1} \dot{g}_{pc}(r)\, dv ds dr \\
& + (c - c_0) \int_0^1 \int_0^r e^{c_0(r-s)} \dot{g}_{pc}(r)' A_p(c)^{-1} g_p(s)\, ds dr \\
& + (c - c_0) \int_0^1 \int_0^r e^{c_0(r-s)} \dot{g}_{pc}(s)' A_p(c)^{-1} g_p(r)\, ds dr \\
& - (c - c_0)^2 \int_0^1 \int_0^1 \int_0^{r \wedge s} e^{c_0(r+s-2v)} \dot{g}_{pc}(s)' A_p(c)^{-1} g_p(r)\, dv ds dr \\
& - (c - c_0) \int_0^1 \int_0^r e^{c_0(r-s)} \dot{g}_{pc}(r)' A_p(c)^{-1} B_p(c) A_p(c)^{-1} \dot{g}_{pc}(s)\, ds dr \\
& - (c - c_0) \int_0^1 \int_0^r e^{c_0(r-s)} \dot{g}_{pc}(s)' A_p(c)^{-1} B_p(c)' A_p(c)^{-1} \dot{g}_{pc}(r)\, ds dr \\
& + (c - c_0)^2 \int_0^1 \int_0^1 \int_0^{r \wedge s} e^{c_0(r+s-2v)} \dot{g}_{pc}(s)' A_p(c)^{-1} B_p(c) A_p(c)^{-1} \dot{g}_{pc}(r)\, dv ds dr \\
& - \int_0^1 \int_0^r e^{c_0(r-s)} \dot{g}_{pc}(s)' A_p(c)^{-1} \dot{g}_{pc}(r)\, ds dr + \lambda_p(c),
\end{aligned}$$

where

$$\lambda_p(c) = \int_0^1 \int_0^r e^{c(r-s)} \dot{g}_{pc}(s)' A_p(c)^{-1} \dot{g}_{pc}(r)\, ds dr.$$

The following lemma derives the convergence rate of $\hat{\sigma}^2$ defined in (15).[8]

[8] In this paper, we assume that ε_{it} are iid (Assumption 3). If the error process ε_{it} has both serial correlation and heteroskedasticity, we may need to estimate the one-side long-run variance (Λ_i) and the

Lemma 2 *Suppose that Assumptions 1 – 3 hold. Define*

$$\tilde{\sigma}^2 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \varepsilon_{it}^2$$

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

$$\hat{\sigma}^2 = \tilde{\sigma}^2 + O_p\left(\frac{1}{T}\right),$$

and

$$\tilde{\sigma}^2 = \sigma^2 + O_p\left(\frac{1}{\sqrt{nT}}\right).$$

The next lemma shows that the sample moment condition $M_{nT}(c)$ has a uniform limit in c.

Lemma 3 (Uniform Convergence) *Under Assumptions 1-3,*

$$M_{nT}(c) \to_p \sigma^2 M(c) \text{ uniformly in } c,$$

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

Assumption 6 *As* $(n, T \to \infty)$, $\hat{W} \to_p W$, *where* W *is positive definite.*

Observe that the limit function $M(c)$ is continuous on the compact parameter set $\mathbb{C}$. Also, note that $M(c) = 0$ at the true parameter $c = c_0$. In Appendix D, we confirm numerically that $M(c) = 0$ only when $c = c_0$. Then, by a standard result (e.g., theorem 2.1 of Newey and McFadden, 1994), the GMM estimator $\hat{c}$ is consistent for the true parameter c_0. Summarizing, we have the following theorem.

Theorem 1 (Consistency) *Suppose Assumptions 1-3 and Assumption 6 hold. Then, as* $(n, T \to \infty)$, $\hat{c} \to_p c_0$.

4.2 Limiting Distribution of the GMM Estimator when $c_0 < 0$

By inspection the objective function $Z_{nT}(c)$ is differentiable in c on the region $c \in (\bar{c}, 0)$, and it has right and left derivatives at $c = \bar{c}$ and 0, respectively. To derive the limit distribution of the GMM estimator, we employ an approach that approximates the objective function $Z_{nT}(c)$ uniformly in terms of a quadratic function in a shrinking neighborhood of the true parameter.

For this purpose, we define

$$dM_{nT}(c) = \frac{1}{n} \sum_{i=1}^{n} dm_{iT}(c),$$

long-run variance (Ω_i) of ε_{it}. Letting $\hat{\Lambda}_i$ and $\hat{\Omega}_i$ denote the estimates of Λ_i and Ω_i, respectively, we need $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\hat{\Lambda}_i - \Lambda_i\right) = o_p(1)$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\hat{\Omega}_i - \Omega_i\right) = o_p(1)$. Typically, we use nonparametric kernel estimation for $\hat{\Lambda}_i$ and $\hat{\Omega}_i$. In this case, in order to have the desired property, we may need to choose a proper bandwidth parameter and a kernel as well as strenthen the restriction on the relative convergence rates between n and T to be $\frac{n}{T} \to 0$. For details, the reader is referred to Moon and Perron (2002).

where $dm_{iT}(c)$ denotes the derivative of $m_{iT}(c)$ with respect to c when $c \in (\bar{c}\ , 0)$ and the right and left derivatives when $c = \bar{c}$ and 0, respectively. By the mean value theorem, for $c \neq c_0$,

$$M_{nT}(c) = M_{nT}(c_0) + dM_{nT}(c_0)(c - c_0) + r_{nT}(c, c_0)(c - c_0),$$

where

$$\begin{aligned} r_{nT}(c, c_0) &= (r_{1nT}(c, c_0), r_{2nT}(c, c_0))', \\ r_{knT}(c, c_0) &= \frac{1}{n}\sum_{i=1}^{n}\left(dm_{kiT}\left(c_k^+\right) - dm_{kiT}(c_0)\right), \end{aligned}$$

and c_k^+ lies between c and c_0 for $k = 1, 2$.

Define

$$\mathcal{S}_{nT} = dM_{nT}(c_0)' \hat{W} M_{nT}(c_0),$$

and

$$\mathcal{H}_{nT} = dM_{nT}(c_0)' \hat{W} dM_{nT}(c_0).$$

Then, we can write

$$\begin{aligned} Z_{nT}(c) &= M_{nT}(c_0)' \hat{W} M_{nT}(c_0) + 2(c - c_0)\mathcal{S}_{nT} + (c - c_0)^2 \mathcal{H}_{nT} \\ &\quad + (c - c_0)\mathcal{R}_{1nT}(c, c_0) + (c - c_0)^2 \mathcal{R}_{2nT}(c, c_0), \end{aligned}$$

where

$$\mathcal{R}_{1nT}(c, c_0) = 2M_{nT}(c_0)' \hat{W} r_{nT}(c, c_0),$$

and

$$\mathcal{R}_{2nT}(c, c_0) = 2dM_{nT}(c_0)' \hat{W} r_{nT}(c, c_0) + r_{nT}(c, c_0)' \hat{W} r_{nT}(c, c_0).$$

Next, we give some asymptotic results that are useful in establishing the limit distribution of $\hat{c}$.

Lemma 4 *Suppose that Assumptions 1-3 hold. When the true parameter is* c_0,

$$dM_{nT}(c) \to_p \sigma^2 dM(c) = \sigma^2 \begin{pmatrix} dM_1(c) \\ dM_2(c) \end{pmatrix} \text{ uniformly in } c \text{ as } (n, T \to \infty)$$

for some continuous function $dM(c)$ *with*

$$dM_1(c_0) = \int_0^1 \int_0^r e^{c_0(r-s)}(r - s)\tilde{h}_p(r, s)\, ds dr - \psi_p(c_0),$$

and

$$
\begin{aligned}
& dM_2\left(c_0\right) \\
= \; & -\int_0^1 \int_0^r e^{2c_0(r-s)} dsdr \\
& + \int_0^1 \int_0^1 \int_0^{r \wedge s} e^{c_0(r+s-2v)} \dot{g}_{pc_0}\left(s\right)' A_p\left(c_0\right)^{-1} \dot{g}_{pc_0}\left(r\right) dvdsdr \\
& + \int_0^1 \int_0^r e^{c_0(r-s)} \dot{g}_{pc_0}\left(r\right)' A_p\left(c_0\right)^{-1} g_p\left(s\right) dsdr \\
& + \int_0^1 \int_0^r e^{c_0(r-s)} \dot{g}_{pc_0}\left(s\right)' A_p\left(c_0\right)^{-1} g_p\left(r\right) dsdr \\
& - \int_0^1 \int_0^r e^{c_0(r-s)} \dot{g}_{pc_0}\left(r\right)' A_p\left(c_0\right)^{-1} B_p\left(c_0\right) A_p\left(c_0\right)^{-1} \dot{g}_{pc_0}\left(s\right) dsdr \\
& - \int_0^1 \int_0^r e^{c_0(r-s)} \dot{g}_{pc_0}\left(s\right)' A_p\left(c_0\right)^{-1} B_p\left(c_0\right)' A_p\left(c_0\right)^{-1} \dot{g}_{pc_0}\left(r\right) dsdr \\
& + \int_0^1 \int_0^r \left(r-s\right) e^{c_0(r-s)} \dot{g}_{pc_0}\left(r\right)' A_p\left(c_0\right)^{-1} \dot{g}_{pc_0}\left(s\right) dsdr.
\end{aligned}
$$

Lemma 5 *Suppose that Assumptions 1-3 hold and set* $B_{nT} = \sqrt{n}$. *Then, as* $(n, T \to \infty)$ *following Assumption 5,*

$$B_{nT} M_{nT}\left(c_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{iT}\left(c_0\right) \Rightarrow N\left(0, \sigma^4 J' \Phi\left(c_0\right) J\right),$$

where $J = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 1 \end{pmatrix}'$ *and* $\Phi\left(c_0\right)$ *is defined in* (46).

Remarks

(a) The proof of Lemma 5 is similar to that of Lemma 3 and is omitted.

(b) Figs. 3-4 plot the graphs of $dM_1\left(c_0, c_0\right)$ in the cases $\tilde{g}_{1t} = \left(1, t\right)'$ and $\tilde{g}_{2t} = \left(1, t, t^2\right)'$, respectively. The graphs reveal that $dM_1\left(c_0, c_0\right) < 0$ for $c_0 < 0$, and, therefore, $\mathcal{H}_{nT} > 0$ for $c_0 < 0$.

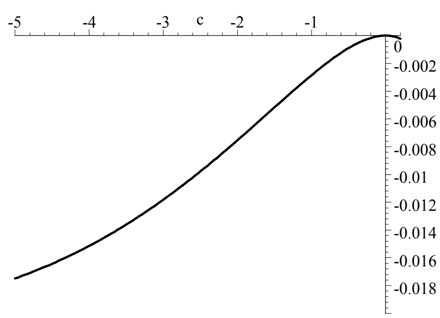

Fig. 3. Graph of $dM_1\left(c_0\right)$ when $\tilde{g}_{1t} = \left(1, t\right)'$.

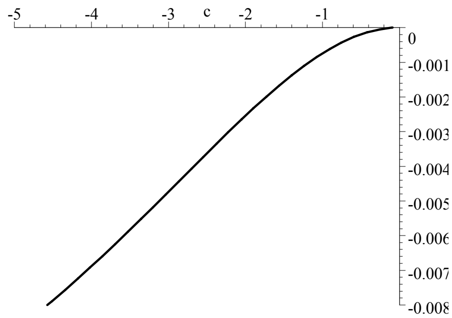

Fig. 4. Graph of $dM_1(c_0)$ when $\tilde{g}_{2t} = (1, t, t^2)'$.

(c) According to Moon and Phillips (2000), when $c_0 = 0$, $dM_1(c_0, c_0) = 0$ holds for all polynomial trends $\tilde{g}_{pt} = (1, ..., t^p)'$. Also, for $c_0 = 0$, direct calculations show that $dM_2(c_0, c_0) = 0$ for $g_{1t} = t$ and $dM_2(c_0, c_0) = 0$ for $g_{2t} = (t, t^2)'$. Therefore, $\mathcal{H}_{nT} \to_p 0$ when $c_0 = 0$, $g_{1t} = t$, and $g_{2t} = (t, t^2)'$.

From Lemma 4 and the following remarks and by Assumption 6, it follows that $\mathcal{H}_{nT}$ has a positive limit as $(n, T \to \infty)$ when $c_0 < 0$. Thus, $\mathcal{H}_{nT}^{-1} = O_p(1)$. Then, we can write

$$\begin{aligned} & B_{nT}^2 Z_{nT}(c) \\ = \; & M_{nT}(c_0)' \hat{W} M_{nT}(c_0) - \frac{(B_{nT}\mathcal{S}_{nT})^2}{\mathcal{H}_{nT}} \\ & + \mathcal{H}_{nT}\left(B_{nT}(c - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2 \\ & + B_{nT}(c - c_0) B_{nT}\mathcal{R}_{1nT}(c, c_0) + (B_{nT}(c - c_0))^2 \mathcal{R}_{2nT}(c, c_0). \end{aligned} \tag{22}$$

Lemma 6 *Under Assumptions 1-3 and Assumption 6, for every sequence $\gamma_{nT} \to 0$, we have as $(n, T \to \infty)$ following Assumption 5,*

(a)

$$\sup_{c \in \mathbb{C}: |c - c_0| \le \gamma_{nT}} |B_{nT}\mathcal{R}_{1nT}(c, c_0)| = o_p(1),$$

and
(b)

$$\sup_{c \in \mathbb{C}: |c - c_0| \le \gamma_{nT}} |\mathcal{R}_{2nT}(c, c_0)| = o_p(1).$$

Theorem 2 *Suppose that Assumptions 1-3 and Assumption 6 hold. Then, as $(n, T \to \infty)$ following Assumption 5,*

$$B_{nT}(\hat{c} - c_0) = O_p(1).$$

Lemma 6 establishes that the two remainder terms $B_{nT}\mathcal{R}_{1nT}(c, c_0)$ and $\mathcal{R}_{2nT}(c, c_0)$ converge in probability to zero uniformly in the shrinking neighborhood of the true parameter. Also, Theorem 2 shows that the GMM estimator is $B_{nT}\,(= \sqrt{n}) -$ consistent. This implies that in the shrinking neighborhood of the true parameter, the scaled objective function $B_{nT}^2 Z_{nT}(c)$ is uniformly approximated by the following quadratic function

$$\begin{aligned} & B_{nT}^2 Z_{q,nT}(c) \\ = \; & M_{nT}(c_0)' \hat{W} M_{nT}(c_0) - \frac{(B_{nT}\mathcal{S}_{nT})^2}{\mathcal{H}_{nT}} + \mathcal{H}_{nT}\left(B_{nT}(c - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2. \end{aligned}$$

The heuristic ideas of the limit theory are as follows. Let $B_{nT}(\hat{c}_q - c_0) = \arg\min_{c\in\mathbb{C}} B_{nT}^2 Z_{q,nT}(c)$. Then, we may expect that a minimizer of $B_{nT}^2 Z_{nT}(c)$ will be close to the minimizer of $B_{nT}^2 Z_{q,nT}(c)$, suggesting that the GMM estimator $B_{nT}(\hat{c} - c_0)$ will be close to

$$\begin{aligned} B_{nT}(\hat{c}_q - c_0) &= \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} \text{ if } \left\{ B_{nT}(\bar{\mathrm{c}} - c_0) \leq \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} \leq -B_{nT}c_0 \right\} \\ &= B_{nT}(\bar{\mathrm{c}} - c_0) \text{ if } \left\{ B_{nT}(\bar{\mathrm{c}} - c_0) > \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} \right\} \\ &= -B_{nT}c_0 \text{ if } \left\{ \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} > -B_{nT}c_0 \right\}. \end{aligned}$$

Notice that $\frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} = O_p(1)$ and recall that it is assumed that the true parameter satisfies $\bar{c} < c_0 < 0$. In this case, the probabilities of the events $\left\{ B_{nT}(\bar{c} - c_0) > \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} \right\}$ and $\left\{ \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} > -B_{nT}c_0 \right\}$ will be very small and the scaled and centred estimator $B_{nT}(\hat{c}_q - c_0)$ will therefore be close with high probability to the random variable

$$\hat{\phi}_{nT} = \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}.$$

In view of Lemmas 4 and 5 and Assumption 6,

$$B_{nT}\mathcal{S}_{nT} \Rightarrow \mathcal{S} \overset{d}{=} N\left(0, \sigma^8 \left[dM(c_0)' WJ'\Phi(c_0) JWdM(c_0)\right]\right)$$

and

$$\mathcal{H}_{nT} \to_p \mathcal{H} = \sigma^4 dM(c_0)' WdM(c_0) > 0$$

as $(n, T \to \infty)$ with $\frac{n}{T} \to 0$. Thus, when $c_0 \in \mathbb{C}_0/\{0\}$,

$$\hat{\phi}_{nT} \Rightarrow \phi \overset{d}{=} \mathcal{H}^{-1}\mathcal{S} \overset{let}{=} \mathcal{Z}.$$

The proof of the following theorem confirms the heuristic argument above.

Theorem 3 *Suppose that Assumptions 1-3 and Assumption 6 hold. Suppose that* $c_0 \in \mathbb{C}_0/\{0\}$ *and* $\hat{c}$ *be the GMM estimator defined in* (21). *Then, as* $(n, T \to \infty)$ *following Assumption 5,*

$$\sqrt{n}(\hat{c} - c_0) \Rightarrow \mathcal{Z},$$

where

$$\mathcal{Z} \overset{d}{=} N\left(0, \frac{dM(c_0)' WJ'\Phi(c_0) JWdM(c_0)}{\left[dM(c_0)' WdM(c_0)\right]^2}\right).$$

Remarks

(a) When $c_0 \in \mathbb{C}_0/\{0\}$ and $J'\Phi(c_0) J$ is invertible, the optimal weight matrix is found as

$$\hat{W}_{opt} = \left(J'\Phi(\hat{c}) J\right)^{-1}.$$

The limiting distribution of $\sqrt{n}(\hat{c} - c_0)$ is then

$$\sqrt{n}(\hat{c} - c_0) \Rightarrow \mathcal{Z}_{opt} \overset{d}{=} N\left(0, \frac{\sigma^4}{\left[dM(c_0)' WdM(c_0)\right]^2}\right). \quad (23)$$

(b) Fig. 5 plots the graph of the minimum eigenvalue of $J'\Phi(c_0)J$ as a function of c_0. As we see through the graph, $J'\Phi(c_0)J$ is positive definite except for the case of $c_0 = 0$.

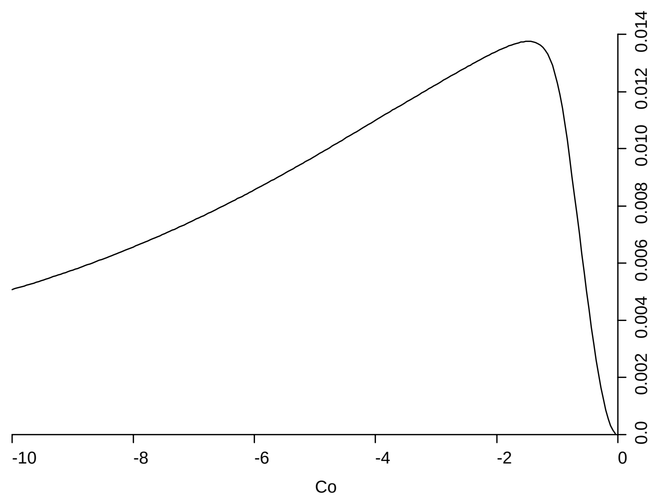

Fig. 5. Graph of the Minimum Eigenvalue of $J'\Phi(c_0)J$ when $g_{1t} = t$.

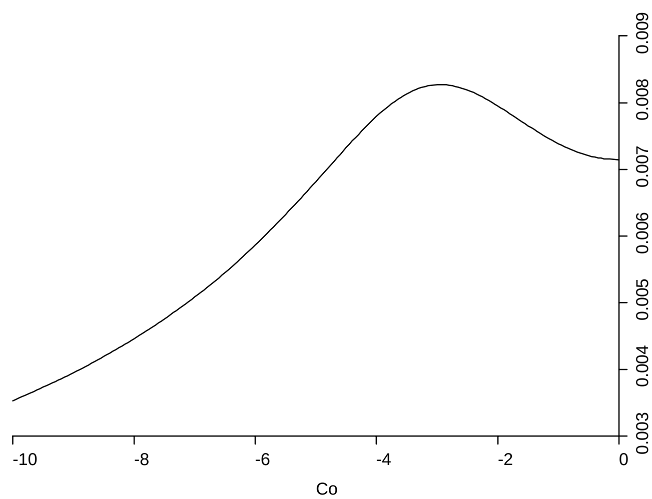

Fig. 6. Graph of the Minimum Eigenvalue of $J'\Phi(c_0)J$ when $g_{2t} = \left(t, t^2\right)'$.

4.3 Limiting Distribution of the GMM Estimator When $c_0 = 0$

An important special case of model 1 occurs when $c_0 = 0$. The time series components of y_{it} in (1) then have a unit root, and the deterministic trend is linear, so

$$\begin{aligned} z_{it} &= \beta_{i0}t + y_{it} \\ y_{it} &= \rho_0 y_{it-1} + \varepsilon_{it}, \end{aligned} \tag{24}$$

where $\rho_0 = 1$, *i.e*, $c_0 = 0$. According to Remark (c) below Lemma 5, in this case, the information from the moment conditions is zero because $\mathcal{H}_{nT} \to_p 0$. We cannot then use a conventional quadratic approximation approach, as in the previous section, and need instead to employ a higher order approximation to extract the limit theory (c.f., Sargan, 1983).

This section develops asymptotics for the GMM estimator when the true localizing parameter is zero, so throughout this section we set $c_0 = 0$. The following lemmas find the limits of the first and the second moment conditions and their higher order derivatives at $c = 0$.

Lemma 7 *Suppose that the panel data is generated by* (24). *Under Assumptions 2 and 3, the following hold as* $(n, T \to \infty)$ *following Assumption 5.*

(a) $\sqrt{n}M_{1nT}(0) \Rightarrow N\left(0, \frac{\sigma^4}{60}\right) \equiv \sqrt{\frac{\sigma^4}{60}}\mathcal{Z}$, *where* $\mathcal{Z} \equiv N(0,1)$,

(b) $\sqrt{n}dM_{1nT}(0) = N\left(0, \sigma^4 \frac{11}{6300}\right)$,

(c) $\sqrt{n}d^2M_{1nT}(0) \Rightarrow o_p(1)$,

(d) $d^3M_{1nT}(c) \to_p \sigma^2 d^3M_1(c, 0)$ *uniformly in* c *with* $d^3M_1(0) = -\frac{1}{70}$,

where $d^kM_{1nT}(c)$ *is the* k^{th} *left derivative of* $M_{1nT}(c)$, *and* $d^3M_1(c)$ *is the third left derivative of* $M_1(c)$, *the probability limit of* $M_{1nT}(c)$.

Lemma 8 *Suppose that the assumptions in Lemma 7 hold. Then, when* $(n, T \to \infty)$ *following Assumption 5,*

(a) $\sqrt{n}M_{2nT}(0) = o_p(1)$,

(b) $\sqrt{n}dM_{2nT}(0) \Rightarrow N\left(0, \frac{\sigma^4}{45}\right)$,

(c) $\sqrt{n}d^2M_{2nT}(0) = o_p(1)$,

(d) $d^3M_{2nT}(c) \to_p \sigma^2 d^3M_2(c)$ *uniformly in* c *with* $d^3M_2(0) = -\frac{1}{15}$,

where $d^kM_{2nT}(0)$ *is the* k^{th} *left derivative of* $M_{2nT}(c)$ *at* $c = 0$, *and* $d^3M_2(0)$ *is the third left derivative of* $d^3M_2(c)$ *at* $c = 0$.

Remarks. Since the higher order derivatives of $M_{2nT}(0)$ are complicated and involve extremely lengthy expressions, we omit the details of their derivation in the appendix. Instead, we give a sketch of the proof in the appendix and here provide some simulation evidence relating to the various parts of Lemmas 7 and 8. Using simulated data for z_{it} in (24) with $\varepsilon_{it} \sim iid\ N(0,1)$ and $y_{i0} = 0$, we estimate the means and the variances of $\sqrt{n}d^kM_{jnT}(0)$, $k = 0, ..., 2$; $j = 1, 2$ and the means of $d^3M_{jnT}(0)$, $j = 1, 2$. Table 1 reports the results. The numbers in the table are consistent with the theoretical results in the lemmas. Noticeably, the variance estimates of $\sqrt{n}M_{1nT}(0)$, $\sqrt{n}dM_{1nT}(0)$, and $\sqrt{n}dM_{2nT}(0)$ are all small. This is because their theoretical limit variances are small but not zero. In fact, a long calculation shows that the theoretical limit variances of $\sqrt{n}M_{1nT}(0)$, $\sqrt{n}dM_{1nT}(0)$, and $\sqrt{n}dM_{2nT}(0)$ are $\frac{1}{60}(\simeq 0.01667)$, $\frac{11}{6300}(\simeq 0.00175)$, and $\frac{1}{45}(\simeq 0.0222)$, respectively when $\varepsilon_{it} \sim iid\ N(0,1)$.

Table 1[9]

[9]Notice that the second and the third derivatives of $M_{1nT}(c)$ are deterministic.

	$\sqrt{n}M_{1nT}(0)$	$\sqrt{n}dM_{1nT}(0)$	$\sqrt{n}d^2M_{1nT}(0)$	$d^3M_{1nT}(c)$
Mean Variance	-0.0019 0.018	-0.0003 0.0017	7.96×10^{-7} 0	-0.0169 N/A
	$\sqrt{n}M_{2nT}(0)$	$\sqrt{n}dM_{2nT}(0)$	$\sqrt{n}d^2M_{2nT}(0)$	$d^3M_{2nT}(0)$
Mean Variance	9.4×10^{-5} 0.0012	-0.0001 0.022	-2.88×10^{-6} 4.85×10^{-6}	-0.06 N/A

Using the left derivatives of the moment condition $M_{nT}(c)$ at $c=0$, we approximate $M_{nT}(c)$ around the true parameter $c_0=0$ with a third order polynomial as follows,

$$M_{nT}(c)=M_{nT}(0)+c\left(dM_{nT}(0)\right)+\frac{1}{2}c^2\left(d^2M_{nT}(0)\right)+\frac{1}{6}c^3\left(d^3M_{nT}(0)\right)+c^3\tilde{r}_{nT}(c,0),$$

where

$$\begin{aligned}\tilde{r}_{nT}(c,0) &= \left(\tilde{r}_{1nT}(c,0),\tilde{r}_{2nT}(c,0)\right)',\\ \tilde{r}_{knT}(c,0) &= d^3M_{knT}\left(c_k^+\right)-d^3M_{knT}(0),\ k=1\text{ and }2.\end{aligned}$$

Then,

$$\begin{aligned}Z_{nT}(c) &= M_{nT}(c)'\hat{W}M_{nT}(c)\\ &= \sum_{k=0}^{6}c^k\mathcal{A}_{k,nT}+\mathcal{N}_{nT}(c,0),\end{aligned}$$

where

$$\begin{aligned}\mathcal{A}_{0,nT} &= M_{nT}(0)'\hat{W}M_{nT}(0),\\ \mathcal{A}_{1,nT} &= 2M_{nT}(0)'\hat{W}dM_{nT}(0),\\ \mathcal{A}_{2,nT} &= M_{nT}(0)'\hat{W}d^2M_{nT}(0)+dM_{nT}(0)'\hat{W}dM_{nT}(0),\\ \mathcal{A}_{3,nT} &= \frac{1}{3}M_{nT}(0)'\hat{W}d^3M_{nT}(0)+dM_{nT}(0)'\hat{W}d^2M_{nT}(0),\\ \mathcal{A}_{4,nT} &= \frac{1}{3}dM_{nT}(0)'\hat{W}d^3M_{nT}(0)+\frac{1}{4}d^2M_{nT}(0)'\hat{W}d^2M_{nT}(0),\\ \mathcal{A}_{5,nT} &= \frac{1}{6}d^2M_{nT}(0)'\hat{W}d^3M_{nT}(0),\\ \mathcal{A}_{6,nT} &= \frac{1}{36}d^3M_{nT}(0)'\hat{W}d^3M_{nT}(0),\end{aligned}$$

and

$$\begin{aligned}\mathcal{N}_{nT}(c,0) &= \sum_{k=3}^{6}c^k\mathcal{N}_{k,nT}(c,0),\\ \mathcal{N}_{k,nT}(c,0) &= \alpha_k d^{(k-3)}M_{nT}(0)'\hat{W}\tilde{r}_{nT}(c,0)\ \text{ for } k=3,4,5,\\ \mathcal{N}_{6,nT}(c,0) &= \alpha_6 d^3M_{nT}(0)'\hat{W}\tilde{r}_{nT}(c,0)+\tilde{r}_{nT}(c,0)'\hat{W}\tilde{r}_{nT}(c,0),\\ \alpha_3,\alpha_4 &= 2,\ \alpha_5=1,\ \alpha_6=\frac{1}{3},\end{aligned}$$

where $d^0 M_{nT}(0)$ denotes $M_{nT}(0)$.

In view of Lemmas 7 and 8, it is easy to find that as $(n, T \to \infty)$ with $\frac{n}{T} \to \kappa < \infty$,

$$n^{5/6}\mathcal{A}_{1,nT} = o_p(1), \tag{25}$$
$$n^{2/3}\mathcal{A}_{2,nT} = o_p(1), \tag{26}$$
$$n^{1/3}\mathcal{A}_{4,nT} = o_p(1), \tag{27}$$
$$n^{1/6}\mathcal{A}_{5,nT} = o_p(1), \tag{28}$$

and

$$\mathcal{A}_{6,nT} \underset{p}{\to} \frac{\sigma^4}{36}\left(\frac{W_{11}}{4900} + \frac{2W_{12}}{1050} + \frac{W_{22}}{225}\right) > 0, \tag{29}$$

$$n^{1/2}\mathcal{A}_{3,nT} \Rightarrow \mathcal{A}_3\mathcal{Z}, \tag{30}$$
$$n\mathcal{A}_{0,nT} \Rightarrow \mathcal{A}_0\mathcal{Z}^2, \tag{31}$$

where $\mathcal{Z} \equiv N(0,1)$ and $\mathcal{A}_3 = -\frac{\sigma^2}{3}\left(\frac{W_{11}}{70} + \frac{W_{12}}{15}\right)\sqrt{\frac{\sigma^4}{60}}$ and $\mathcal{A}_0 = W_{11}\frac{\sigma^4}{60}$.

Also, using Lemmas 7 and 8 and following similar lines of proof to Lemma 6, we can show that

$$\sup_{c \in \mathbb{C}: |c| \leq \gamma_{nT}} \left| n^{(6-k)/6}\mathcal{N}_{k,nT}(c, 0)\right| = o_p(1), \tag{32}$$

for any sequence γ_{nT} tending to zero as $(n, T \to \infty)$. Then, we have the following limit theory for $\hat{c}$ at the origin.

Theorem 4 *Under the assumptions in Lemmas 7 or 8, and as $(n, T \to \infty)$ following Assumption 5,*

$$n^{1/6}(\hat{c} - c_0) = O_p(1),$$

where $c_0 = 0$.

So, when the true localizing parameter is $c_0 = 0$, the GMM estimator $\hat{c}$ is $n^{1/6}-$ consistent, which is slower than the regular case of $\sqrt{n}$ that applies for $c_0 < 0$ as shown in Section 4. To find the limiting distribution of $\hat{c}$, we use an argument similar to that of the previous section. Consequently, we sketch the derivation and give the final result in Theorem 5 below.

In view of $(25) - (31)$ and (32), the standardized objective function $nZ_{nT}(c)$ is approximated by

$$Z_{q,nT}(c) = n\mathcal{A}_{0,nT} + \left(n^{1/6}c\right)^3\sqrt{n}\mathcal{A}_{3,nT} + \left(n^{1/6}c\right)^6\mathcal{A}_{6,nT}.$$

Notice that the probability limit of $\mathcal{A}_{6,nT}$ is positive, as shown in (29). Then, it is easy to see that the approximate objective function $Z_{q,nT}(c)$ is minimized at

$$\begin{aligned} n^{1/6}\hat{c}_q &= -\left(\frac{\sqrt{n}\mathcal{A}_{3,nT}}{2\mathcal{A}_{6,nT}}\right)^{1/3} \text{ if } \left\{n^{1/6}\bar{c} \leq -\frac{\sqrt{n}\mathcal{A}_{3,nT}}{2\mathcal{A}_{6,nT}} \leq 0\right\} \\ &= 0 \text{ if } \left\{-\frac{\sqrt{n}\mathcal{A}_{3,nT}}{2\mathcal{A}_{6,nT}} > 0\right\} \\ &= -\left(n^{1/6}(-\bar{c})\right)^{1/3} \text{ if } \left\{n^{1/6}\bar{c} > -\frac{\sqrt{n}\mathcal{A}_{3,nT}}{2\mathcal{A}_{6,nT}}\right\}. \end{aligned}$$

Using arguments similar to those in the proof of Theorem 3, we can prove that the standardized GMM estimator $n^{1/6}\hat{c}$ is approximated by $n^{1/6}\hat{c}_q$, the minimizer of $Z_{q,nT}(c)$, that is,

$$n^{1/6}\hat{c} = n^{1/6}\hat{c}_q + o_p(1),$$

and the estimator $n^{1/6}\hat{c}_q$ is approximated by

$$\hat{\phi}_{nT} = -\left(\frac{\sqrt{n}\mathcal{A}_{3,nT}}{2\mathcal{A}_{6,nT}}\right)^{1/3} \mathbf{1}\left\{-\frac{\sqrt{n}\mathcal{A}_{3,nT}}{2\mathcal{A}_{6,nT}} \leq 0\right\},$$

where $\mathbf{1}\{\mathbf{A}\}$ is the indicator of A. In view of (30) and (29), as $(n, T \to \infty)$ following Assumption 5, it follows by the continuous mapping theorem that

$$\hat{\phi}_{nT} \Rightarrow \mathcal{Z}_0^{1/3}\mathbf{1}\{\mathcal{Z}_0 \leq 0\},$$

where

$$\mathcal{Z}_0 = V_0\mathcal{Z}, \tag{33}$$

$$V_0 = \left|\frac{\sqrt{\frac{1}{15}}\left(\frac{W_{11}}{70} + \frac{W_{12}}{15}\right)}{\frac{1}{3}\left(\frac{W_{11}}{4900} + \frac{2W_{12}}{1050} + \frac{W_{22}}{225}\right)}\right|, \tag{34}$$

and W_{ij} are the $(i,j)^{th}$ element of the weight matrix W. Thus, we have the following theorem.

Theorem 5 *Under the assumptions in Lemmas 7 and 8, as $(n, T \to \infty)$ following Assumption 5,*

$$n^{1/6}\hat{c} \Rightarrow \mathcal{Z}_0^{1/3}\mathbf{1}\{\mathcal{Z}_0 \leq 0\},$$

where $\mathcal{Z}_0$ is defined in (33).

Remarks

(a) Theorem 4 shows that when the true parameter $c_0 = 0$, *i.e.*, in the case of a panel unit root, the GMM estimator is $n^{1/6}$-consistent and that its limit distribution is nonstandard, involving the cube root of a truncated normal. The truncation in the limiting distribution arises because the true parameter is on the boundary of the parameter set.

(b) The reason for the slower convergence rate in the panel unit root case is that first order information in the moment condition (from the first derivative of the moment condition) is aymptotically zero at the true parameter. In order to obtain nonneglible information from the moment condition, we need to pass to third order derivatives of the moment condition. Taking the higher order approximation slows down the convergence rate because the rate at which information in the moment condition is passed to the estimator is slowed down at the origin because of the zero lower derivatives.

(c) In view of Lemmas 7(a) and 8(a), we find that $\sqrt{n}M_{2nT}(0) = o_p(1)$, while $\sqrt{n}M_{1nT}(0)$ converges in distribution to a normal random variable with positive variance. Because of the convergence rate difference between $\sqrt{n}M_{2nT}(0)$ and $\sqrt{n}M_{1nT}(0)$, we have only W_{11} and W_{12} but not W_{22} in the limiting scale V_0 of (34). In this case, setting $W_{11} = W_{12} = 0$, i.e. not considering the first moment condition, causes the variance of the limit variate $\mathcal{Z}_0$ in (33) to vanish, from which one might expect that the GMM estimator from the second moment condition alone would have a faster convergence rate than $n^{1/6}$. The reason for using the first moment condition is to identify the true parameter c_0 when $c_0 < 0$. As we discuss in Appendix D, the second moment condition cannot identify the true parameter c_0 unless $c_0 = 0$.

4.4 On Testing for a Unit Root

This section briefly considers how the asymptotic results for the localizing coefficient given in the previous section may be used to test for a unit root in the panel. Suppose the null hypothesis is $\mathbb{H}_0 : c_0 = 0$ and the alternative hypothesis is $\mathbb{H}_1 : c_0 < 0$. We discuss two types of panel unit root tests, one involving a $t-$ test and the other an LM test.

First, Theorem 5 shows that to test $\mathbb{H}_0$ we can use a suitably constructed $t-$ statistic. Specifically, let $\hat{V}_0$ be a consistent estimator of V_0 and define

$$t_{gmm} = \frac{\sqrt{n}\hat{c}^3}{\hat{V}_0}.$$

Then, since $\hat{V}_0 \to_p V_0$ and from Theorem 5 with $(n, T \to \infty)$ as in Assumption 5, we get

$$t_{gmm} \Rightarrow \mathcal{Z}\mathbf{1}\left\{\mathcal{Z} \leq 0\right\},$$

where $\mathcal{Z} \equiv N(0,1)$. Under the alternative hypothesis $c = c_A < 0$, we have

$$\begin{aligned} t_{gmm} &= \frac{\sqrt{n}\left(\hat{c}^3 - c_A^3\right)}{\hat{V}_0} + \frac{\sqrt{n}c_A^3}{\hat{V}_0} \\ &= O_p(1) + \frac{\sqrt{n}c_A^3}{\hat{V}_0} \end{aligned}$$

by Theorem 3 and the delta method. So, under the alternative hypothesis, $t_{gmm} \to -\infty$ and the test is consistent.

Another type of test is to use the asymptotic properties of the moment conditions in Lemmas 7 and 8 in conjunction with the restricted parameter estimator, which is zero in this case. For example, in Lemma 7 we observe that $\sqrt{n}dM_{1nT}(0) = N\left(0, \sigma^4\frac{11}{6300}\right)$. Thus, a simple test can be based on

$$LM_{gmm} = \left(\sqrt{\frac{6300}{11}}\frac{\sqrt{n}dM_{1nT}(0)}{\hat{\sigma}^2}\right)^2.$$

Then, as $(n, T \to \infty)$ as in Assumption 5, we have

$$LM_{gmm} \Rightarrow \chi^2(1).$$

Under the alternative $c = c_A < 0$, it is easy to show that $\left(\sqrt{n}dM_{2nT}(0)\right)^2 = nO(1)^2 \to \infty$, while $d^3M_{2nT}(0) \to_p -\sigma^2\frac{1}{15} < 0$. Thus, under the alternative hypothesis, $LM_{gmm} \to \infty$. The same principle can be applied to the second moment condition $M_{2nT}(0)$.

5 Conclusion

Part of the richness of panel data is that it can provide information about features of a model on which time series and cross section data are uninformative when they are used on their own. In the context of nonstationary panels with near unit roots, an interesting new example of this 'added information' feature of panel data is that consistent estimation of the common local to unity coefficient becomes possible. This means that panel data help to sharpen our capacity to learn from data about the precise form of nonstationarity where time series data alone are insufficient to do so. However, as the authors have shown in earlier work, the presence of individual deterministic trends in a panel model introduces a serious complication in this nice result on the consistent estimation of a root local to unity. The complication is that individual trends produce an incidental parameter problem as $n \to \infty$ that does not disappear as $T \to \infty$. The outcome is that common procedures like pooled least squares and maximum likelihood are inconsistent. Thus, the presence of deterministic trends continues to confabulate inference about stochastic trends even in the panel data case.

One option is to adjust procedures like maximum likelihood to deal with the bias. The present paper shows how to make these adjustments. The theory is cast in the context of moment formulae that lead naturally to GMM based estimation. The paper has two important findings.

The first is that bias correction in the moment formulae arising from GLS estimation of the trend coefficients corresponds to taking the projected score (under Gaussian assumptions) on the Bhattacharya basis. This correspondence relates the approach we take here to recent work on projected score methods by Waterman and Lindsay (1998) that deals with models that have infinite numbers of nuisance parameters like the original incidental parameters problem.

The second is that our limit theory validates GMM-based inference about the localizing coefficient in near unit root panels. A notable new result is that the GMM estimator has a convergence rate slower than $\sqrt{n}$ when the true localizing parameter is zero (i.e., when there is a panel unit root) and the deterministic trends in the panel are linear. The asymptotic theory in this case provides a new example of limit theory on the boundary of a parameter space. The results point to the continued difficulty of distinguishing unit roots from local alternatives when there are deterministic trends in the data even when time series data is coupled with an infinity of additional data from a cross section.

6 Appendix

6.1 Proof of the Equivalence Lemma

Before we start the proof of Lemma 1, we give some useful background results.

Lemma 9 *Let K_m denote the $(m \times m)$ commutation matrix, D_m denote the $m^2 \times \frac{1}{2}m\,(m+1)$ duplication matrix, and set $D_m^+ = \left(D_m' D_m\right)^{-1} D_m'$. Also, assume that x and y are $m-$vectors and A is an $(m \times m)$ invertible matrix. Then the following hold.*

(a) $xy' \otimes yx' = K_m \left(yy' \otimes xx'\right).$

(b) $\left(I_m + K_m\right)\left((x \otimes y) + (y \otimes x)\right) = 2\,(x \otimes y) + 2\,(y \otimes x).$

(c) $D_p^+ D_p = I_{\frac{1}{2}p(1+p)}.$

(d) $D_p D_p^+ = \frac{1}{2}\left(I_p + K_p\right).$

(e) $\left(D_p^+ \left(A \otimes A\right) D_p\right)^{-1} = D_p^+ \left(A^{-1} \otimes A^{-1}\right) D_p.$

Proof

Parts (c), (d), and (e) are standard results (e.g., Magnus and Neudecker, 1988, pp. 49-50). Part (a) holds because

$$\begin{aligned} xy' \otimes yx' &= (x \otimes y)(y' \otimes x') = vec(yx')(vec(xy'))' \\ &= (K_m vec(xy'))(vec(xy'))' = K_m (y \otimes x)(y \otimes x)' \\ &= K_m (yy' \otimes xx'). \end{aligned}$$

Part (b) holds because

$$\begin{aligned} & (I_m + K_m)((x \otimes y) + (y \otimes x)) \\ = & (x \otimes y) + (y \otimes x) + K_m vec(yx') + K_m vec(xy') \\ = & (x \otimes y) + (y \otimes x) + vec(xy') + vec(yx') \\ = & 2(x \otimes y) + 2(y \otimes x). \blacksquare \end{aligned}$$

Proof of Lemma 1

In this proof we omit the subscript p that denotes the order of the polynomial trends for notational simplicity. To complete the proof, it is enough to show that $\lambda_T(c)$ in $m_{2,iT}(c)$ is equivalent to $\xi_2' D_p^+ \sum_{t=1}^T (\Delta_c g_t \otimes \Delta_c g_t)$ in $U_{2i}\left(c, \hat{\beta}_i(c)\right)$. First, we define

$$\begin{aligned} \tilde{A}_{1T} &= \frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1} D_p^+ \left[\widehat{\Delta_c g_t} \otimes \widehat{\Delta_c g_s} + \widehat{\Delta_c g_s} \otimes \widehat{\Delta_c g_t}\right]\left[\left(1+\frac{c}{T}\right)^T\right]^{\frac{t-s-1}{T}}, \\ \tilde{A}_{2T} &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{T}\sum_{s=1}^{T} D_p^+ \left\{\left(\widehat{\Delta_c g_t}\widehat{\Delta_c g_t}' \otimes \widehat{\Delta_c g_s}\widehat{\Delta_c g_s}'\right) + \left(\widehat{\Delta_c g_t}\widehat{\Delta_c g_s}' \otimes \widehat{\Delta_c g_s}\widehat{\Delta_c g_t}'\right)\right\}\left(D_p^+\right)', \\ \tilde{A}_{3T} &= D_p^+ \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\Delta_c g_t} \otimes \widehat{\Delta_c g_t}\right). \end{aligned}$$

Then, by definition, we write

$$\xi_2' D_p^+ \sum_{t=1}^T (\Delta_c g_t \otimes \Delta_c g_t) = \tilde{A}_{1T}' \tilde{A}_{2T}^{-1} \tilde{A}_{3T}.$$

Notice by Lemma 9(a), (d), and (c) that

$$\begin{aligned} & \tilde{A}_{2T} \\ = & D_p^+ (I_p + K_p)\frac{1}{T}\sum_{t=1}^{T}\frac{1}{T}\sum_{s=1}^{T}\left(\widehat{\Delta_c g_t}\widehat{\Delta_c g_t}' \otimes \widehat{\Delta_c g_s}\widehat{\Delta_c g_s}'\right)\left(D_p^+\right)' \\ = & 2D_p^+ D_p D_p^+ \left[\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta_c g_t}\widehat{\Delta_c g_t}'\right) \otimes \left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\Delta_c g_s}\widehat{\Delta_c g_s}'\right)\right]\left(D_p^+\right)' \\ = & 2D_p^+ \left[\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta_c g_t}\widehat{\Delta_c g_t}'\right) \otimes \left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\Delta_c g_s}\widehat{\Delta_c g_s}'\right)\right]\left(D_p^+\right)' \\ = & 2\left[D_p^+ \left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta_c g_t}\widehat{\Delta_c g_t}'\right) \otimes \left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\Delta_c g_s}\widehat{\Delta_c g_s}'\right) D_p\right]\left(D_p' D_p\right)^{-1}. \end{aligned}$$

By Lemma 9(e),

$$
\begin{aligned}
& \tilde{A}_{2T}^{-1} \\
= \; & \frac{1}{2}\left(D_p' D_p\right) D_p^{+}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta_c g_t}\widehat{\Delta_c g_t}'\right)^{-1}\otimes\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\Delta_c g_s}\widehat{\Delta_c g_s}'\right)^{-1} D_p\right] \\
= \; & \frac{1}{2} D_p'\left[\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta_c g_t}\widehat{\Delta_c g_t}'\right)^{-1}\otimes\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\Delta_c g_s}\widehat{\Delta_c g_s}'\right)^{-1}\right] D_p .
\end{aligned}
$$

Again, from Lemma 9(d) and (b), we have

$$
\begin{aligned}
& \tilde{A}_{1T}' \tilde{A}_{2T}^{-1} \tilde{A}_{3T} \\
= \; & \frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1}\left[\widehat{\Delta_c g_t}\otimes\widehat{\Delta_c g_s}+\widehat{\Delta_c g_s}\otimes\widehat{\Delta_c g_t}\right]'\left[\left(1+\frac{c}{T}\right)^{T}\right]^{\frac{t-s-1}{T}}\left(D_p^{+}\right)' \\
& \times\frac{1}{2} D_p'\left[\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta_c g_t}\widehat{\Delta_c g_t}\right)^{-1}\otimes\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\Delta_c g_s}\widehat{\Delta_c g_s}'\right)^{-1}\right] D_p \\
& \times D_p^{+}\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\Delta_c g_t}\otimes\widehat{\Delta_c g_t}\right) \\
= \; & \frac{1}{8}\left[\frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1}\left[\widehat{\Delta_c g_t}\otimes\widehat{\Delta_c g_s}+\widehat{\Delta_c g_s}\otimes\widehat{\Delta_c g_t}\right]'\left[\left(1+\frac{c}{T}\right)^{T}\right]^{\frac{t-s-1}{T}}\right]\left(I_p+K_p\right)' \\
& \times\left[\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta_c g_t}\widehat{\Delta_c g_t}\right)^{-1}\otimes\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\Delta_c g_s}\widehat{\Delta_c g_s}'\right)^{-1}\right] \\
& \times\left(I_p+K_p\right)\left[\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\Delta_c g_t}\otimes\widehat{\Delta_c g_t}\right)\right] \\
= \; & \frac{1}{2}\left[\frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1}\left[\widehat{\Delta_c g_t}\otimes\widehat{\Delta_c g_s}+\widehat{\Delta_c g_s}\otimes\widehat{\Delta_c g_t}\right]'\left[\left(1+\frac{c}{T}\right)^{T}\right]^{\frac{t-s-1}{T}}\right] \\
& \times\left[\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta_c g_t}\widehat{\Delta_c g_t}\right)^{-1}\otimes\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\Delta_c g_s}\widehat{\Delta_c g_s}'\right)^{-1}\right] \\
& \times\left[\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\Delta_c g_t}\otimes\widehat{\Delta_c g_t}\right)\right]. \qquad (35)
\end{aligned}
$$

Expanding (35) yields

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1}\frac{1}{T}\sum_{p=1}^{T}\left[\left(1+\frac{c}{T}\right)^{T}\right]^{\frac{t-s-1}{T}}\left[\widehat{\Delta_c g_s}' A_{pT}^{-1}\widehat{\Delta_c g_p}\right]\left[\widehat{\Delta_c g_p}' A_{pT}^{-1}\widehat{\Delta_c g_t}\right] \\
= \; & \frac{1}{T}\sum_{t=2}^{T}\frac{1}{T}\sum_{s=1}^{t-1}\left[\left(1+\frac{c}{T}\right)^{T}\right]^{\frac{t-s-1}{T}}\widehat{\Delta_c g_s}' A_{pT}^{-1}\widehat{\Delta_c g_t} \\
= \; & \lambda_{pT}(c). \; \blacksquare
\end{aligned}
$$

6.2 Appendix A: Useful Results for Joint Asymptotics

This section consists of two subsections. The first subsection introduces some useful results for joint asymptotic theories. Many of these are modified versions of results developed in Phillips and Moon (1999) so we report them only briefly here. The second subsection introduces some useful results which will be used repeatedly in the following sections of the proofs for the results in the main text.

6.2.1 Appendix A1

The following two theorems provide convenient conditions to find the joint probability limit of double indexed processes.

Theorem 6 ***(Joint Probability Limits)*** *Suppose the* $(m \times 1)$ *random vectors* Y_{iT} *are independent across* $i = 1, ..., n$ *for all* T *and integrable. Assume that* $Y_{iT} \Rightarrow Y_i$ *as* $T \to \infty$ *for all* i. *Let* $X_{nT} = \frac{1}{n}\sum_{i=1}^{n} Y_{iT}$ *and* $X_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

(a) *Let the following hold:*

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

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

Theorem 7 *Suppose that* $Y_{iT} = C_i Q_{iT}$, *where the* $(m \times 1)$ *random vectors* Q_{iT} *are* iid *across* $i = 1, ..., n$ *for all* T, *and the* C_i *are* $(m \times m)$ *nonrandom matrices for all* i. *Assume that*

- (i) $Q_{iT} \Rightarrow Q_i$ *as* $T \longrightarrow \infty$ *for all* i,
- (ii) $||Q_{iT}||$ *is uniformly integrable in* T *for all* i.
- (iii) $\sup_i ||C_i|| < \infty$, $\inf_i ||C_i|| > 0$, *and* $C = \lim_n \frac{1}{n}\sum_{i=1}^{n} C_i$.
  Then $\frac{1}{n}\sum_{i=1}^{n} Y_{iT} \to_p CE(Q_i)$ *as* $(n, T \to \infty)$.

Theorem 8 ***(Joint Limit CLT for Scaled Variates)*** *Suppose that* $Y_{iT} = C_i Q_{iT}$, *where the* $(m \times 1)$ *random vectors* Q_{iT} *are* $\text{iid}(0, \Sigma_T)$ *across* $i = 1, ..., n$ *for all* T *and the* C_i *are* $(m \times m)$ *nonzero and nonrandom matrices. Assume the following conditions hold:*

- (i) *Let* $\sigma_T^2 = \lambda_{\min}(\Sigma_T)$ *and* $\liminf_T \sigma_T^2 > 0$,
- (ii) $\frac{\max_{i\leq n}\|C_i\|^2}{\lambda_{\min}\left(\sum_{i=1}^{n} C_i C_i'\right)} = O\left(\frac{1}{n}\right)$ *as* $n \to \infty$,
- (iii) $||Q_{iT}||^2$ *are uniformly integrable in* T,
- (iv) $\lim_{n,T} \frac{1}{n}\sum_{i=1}^{n} C_i \sum_T C_i' = \Omega > 0$.

Then,

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

6.2.2 Appendix A2

Recall that

$$\tilde{h}_{pT}(t,s)=\tilde{g}'_{pt}\tilde{D}_{p,T}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{D}_{p,T}\tilde{g}_{pt}\tilde{g}'_{pt}\tilde{D}_{p,T}\right)^{-1}\tilde{D}_{p,T}\tilde{g}_{ps}$$

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

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

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

$$\sup_{1\le t,s\le T}\tilde{h}_{pT}(t,s)\to\sup_{0\le r,v\le 1}\tilde{h}_p(r,v).$$

Next, define

$$x_{it}=\sum_{s=1}^{t}\left(1+\frac{c_0}{T}\right)^{(t-s)}\varepsilon_{is} \tag{36}$$

and $x_{i0}=0$. Then, we can write y_{it} as

$$y_{it}=x_{it}+R_{it}, \tag{37}$$

where $R_{it}=\left(1+\frac{c_0}{T}\right)^{t}y_{i0}$.

When $t=[Tr]$, as $T\to\infty$,

$$\begin{aligned}E\left(\frac{x_{it}^2}{T}\right) &= \sigma^2\left(\frac{1}{T}\sum_{s=1}^{t}\left[\left(1+\frac{c_0}{T}\right)^{T}\right]^{\frac{2(t-s)}{T}}\right)\\ &\to \sigma^2\int_0^r\exp\left((r-s)2c_0\right)ds<\bar{K}, \qquad (38)\\ \frac{1}{T}\sum_{t=1}^{T}\sqrt{E\left(\frac{x_{it}^2}{T}\right)} &= \frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{T}\sum_{s=1}^{t}\left[\left(1+\frac{c_0}{T}\right)^{T}\right]^{\frac{2(t-s)}{T}}}\\ &\to \sigma\int_0^1\left(\int_0^r e^{(r-s)2c_0}ds\right)^{\frac{1}{2}}dr<\bar{K}, \qquad (39)\end{aligned}$$

where $\bar{K}$ is a finite generic constant.

Lemma 10 *Assume that $\{F_T(c)\}$ is a sequence of real valued functions on a compact set $\mathbb{C}$ in $\mathbb{R}$ with*

$$F_T(c)\to 0,\ \text{for all } c\in\mathbb{C},$$

as $T\to\infty$. Suppose that for any given $\varepsilon>0$ and $c\in\mathbb{C}$, there exist $T_0(c,\varepsilon)$ and $\delta(c,\varepsilon)>0$ such that $T\ge T_0(c,\varepsilon)$ implies that

$$\sup_{|c-\bar{c}|<2\delta(c,\varepsilon)}|F_T(c)-F_T(\bar{c})|<\varepsilon. \tag{40}$$

Then,

$$\sup_{c\in\mathbb{C}}|F_T(c)|\to 0.$$

Proof

Let $\mathbb{B}(c,\delta)$ be a $\delta-$ ball around c. From (40), for the given $\varepsilon > 0$ and c, one may choose $T_0(c,\varepsilon)$ and $\delta(c,\varepsilon)$ such that if $T \geq T_0(c,\varepsilon)$, then

$$|c-\bar{c}| < 2\delta(c,\varepsilon)$$

implies

$$|F_T(c) - F_T(\bar{c})| < \frac{\varepsilon}{2}.$$

Also, since $F_T(c) \to 0$ pointwise in $c \in \mathbb{C}$, we may choose $T_1(c,\varepsilon)$ such that $T \geq T_1(c,\varepsilon)$ implies that

$$|F_T(c)| < \frac{\varepsilon}{2}.$$

Since $\mathbb{C}$ is compact, among the open covers $\{\mathbb{B}(c,\delta(c,\varepsilon))\}_{c\in\mathbb{C}}$, we can choose a finite number of open covers of $\mathbb{C}$, $\{\mathbb{B}(c_l,\delta(c_l))\}_{l=1,\ldots,L}$. Set $T_0(\varepsilon) = \max_{l=1,\ldots,L}\max\{T_0(c_l,\varepsilon), T_1(c_l,\varepsilon)\}$. Then, if $T \geq T_0(\varepsilon)$, then,

$$\begin{aligned}
\sup_{c\in\mathbb{C}}|F_T(c)| &= \sup_{l=1,\ldots,L}\ \sup_{c\in\mathbb{B}(c_l,\delta(c_l,\varepsilon))}|F_T(c)| \\
&\leq \sup_{l=1,\ldots,L}\ \sup_{c\in\mathbb{B}(c_l,\delta(c_l,\varepsilon))}|F_T(c) - F_T(c_l)| + \sup_{l=1,\ldots,L}|F_T(c_l)| \\
&\leq \varepsilon,
\end{aligned}$$

and we have the required result. ■

Lemma 11 *Let* $f_T(c) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\left(1+\frac{c}{T}\right)^T\right]^{h_T\left(\frac{t}{T},\frac{s}{T}\right)} g_T\left(\frac{t}{T},\frac{s}{T},c\right)$, *where* $c \in \mathbb{C}$, *a compact subset in* $\mathbb{R}$, $g_T\left(\frac{t}{T},\frac{s}{T},c\right)$ *is continuously differentiable in* c, *and* $\sup_{(r,p,c)\in[0,1]^2\times\mathbb{C}}|g_T(r,p,c)|$, $\sup_{(r,p,c)\in[0,1]^2\times\mathbb{C}}\left|\frac{\partial g_T(r,p,c)}{\partial c}\right|$, $\sup_{(r,p)\in[0,1]^2}|h_T(r,p)| < \bar{K}$. *Suppose that* $g_T\left(\frac{[Tr]}{T},\frac{[Tp]}{T},c\right) \to g(r,p,c)$ *and* $h_T\left(\frac{[Tr]}{T},\frac{[Tp]}{T}\right) \to h(r,p)$ *unifomrly in* $(r,p) \in [0,1]^2$, *where* $g(r,p,c)$ *and* $h(r,p)$ *are continuous functions on* $[0,1]^2\times\mathbb{C}$ *and* $[0,1]^2$, *respectively, satisfying*

$$\int_0^1\int_0^1 e^{\bar{c}|h(r,p)|}dpdr < \infty,$$

where $\bar{c} = \max_{c\in\mathbb{C}}|c|$. *Then*

$$f_T(c) \to f(c) = \int_0^1\int_0^1 e^{ch(r,p)}g(r,p,c)\,drdp \text{ uniformly in } c$$

as $T \to \infty$.

Proof

Let $F_T(c) = f_T(c) - f(c)$. Under the restrictions in the lemma,

$$F_T(c) \to 0$$

for all $c \in \mathbb{C}$. Then, by Lemma 10, the desired result follows if we verify that $F_T(c)$ satisfies condition (40). For this, it is enough to show that $f_T(c)$ satisfies condition (40) because

$f(c)$ is uniformly continuous in c. Fix $c_0 \in \mathbb{C}$. Observe that $f_T(c)$ is a differentiable function. Now, by the mean value theorem,

$$f_T(c) - f_T(c_0) = \begin{pmatrix} \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} h_T\left(\frac{t}{T},\frac{s}{T}\right)\left[\left(1+\frac{c^*}{T}\right)^T\right]^{h_T\left(\frac{t}{T},\frac{s}{T}\right)-\frac{1}{T}} g_T\left(\frac{t}{T},\frac{s}{T},c^*\right) \\ +\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\left(1+\frac{c^*}{T}\right)^T\right]^{h_T\left(\frac{t}{T},\frac{s}{T}\right)} \frac{\partial g_T\left(\frac{t}{T},\frac{s}{T},c^*\right)}{\partial c} \end{pmatrix}(c-c_0),$$

where c^* is located between c and c_0. Let $\bar{c} = \max_{c\in\mathbb{C}}|c|$. Then,

$$\begin{aligned} & \left| \begin{array}{c} \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} h_T\left(\frac{t}{T},\frac{s}{T}\right)\left[\left(1+\frac{c^*}{T}\right)^T\right]^{h_T\left(\frac{t}{T},\frac{s}{T}\right)-\frac{1}{T}} g_T\left(\frac{t}{T},\frac{s}{T},c^*\right) \\ +\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\left(1+\frac{c^*}{T}\right)^T\right]^{h_T\left(\frac{t}{T},\frac{s}{T}\right)} \frac{\partial g_T\left(\frac{t}{T},\frac{s}{T},c^*\right)}{\partial c} \end{array} \right| \\ \leq \quad & \frac{\bar{K}}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\left(1+\frac{\bar{c}}{T}\right)^T\right]^{\left|h_T\left(\frac{t}{T},\frac{s}{T}\right)\right|} \\ \to \quad & \bar{K}\int_0^1\int_0^1 e^{\bar{c}|h(r,p)|}dpdr = K, \text{ say.} \end{aligned} \tag{41}$$

Choose $2\delta(c_0,\varepsilon) = \frac{\varepsilon}{K+\varepsilon}$. From (41), we can choose a $T_0(c_0,\varepsilon)$ such that $T \geq T_0(c_0,\varepsilon)$ implies

$$\left| \begin{array}{c} \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} h_T\left(\frac{t}{T},\frac{s}{T}\right)\left[\left(1+\frac{c^*}{T}\right)^T\right]^{h_T\left(\frac{t}{T},\frac{s}{T}\right)-\frac{1}{T}} g_T\left(\frac{t}{T},\frac{s}{T},c^*\right) \\ +\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\left(1+\frac{c^*}{T}\right)^T\right]^{h_T\left(\frac{t}{T},\frac{s}{T}\right)} \frac{\partial g_T\left(\frac{t}{T},\frac{s}{T},c^*\right)}{\partial c} \end{array} \right| \leq K+\varepsilon.$$

Then, $T \geq T_0(c_0,\varepsilon)$ and $|c-c_0| < 2\delta(c_0,\varepsilon)$ imply that

$$|f_T(c) - f_T(c_0)| \leq (K+\varepsilon)|c-c_0| \leq \varepsilon,$$

and the proof is completed. ■

Corollary 1 *(a) $A_{pT}(c) \to A_p(c)$ uniformly in $c \in \mathbb{C}$,*
(b) $B_{pT}(c) \to B_p(c)$ uniformly in $c \in \mathbb{C}$,
(c) $A_{pT}(c)^{-1} \to A_p^{-1}(c)$ unifomly in $c \in \mathbb{C}$,
(d) $\omega_{pT}(c) \to \omega_p(c)$ unifomly in $c \in \mathbb{C}$,
(e) $\lambda_{pT}(c) \to \lambda_p(c)$ unifomly in $c \in \mathbb{C}$, where $p = 1, 2$.

Proof
Part (a). Notice that each element in $A_{pT}(c)$ is of the form

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{t^p-(t-1)^p}{T^{p-1}} - c\left(\frac{t-1}{T}\right)^p\right]\left[\frac{t^p-(t-1)^q}{T^{p-1}} - c\left(\frac{t-1}{T}\right)^q\right], \; p = 1,2 \text{ and } q = 1,2.$$

Apply Lemma 11 with $h_T(\cdot,\cdot) = 0$ and

$$g_T\left(\frac{t}{T},\frac{s}{T},c\right) = \left[\frac{t^p-(t-1)^p}{T^{p-1}} - c\left(\frac{t-1}{T}\right)^p\right]\left[\frac{t^p-(t-1)^q}{T^{p-1}} - c\left(\frac{t-1}{T}\right)^q\right].$$

Then, we have the required result. ■

Part (b). The proof of Part (b) is similar to that of Part (a) and is omitted. ■

Part (c). Let mineig(A) denote the minimum eigenvalue of matrix A. A direct calculation shows that $\inf_{c\in\mathbb{C}}[\text{mineig}(A_{2T}(c))] > 0$ and $\inf_{c\in\mathbb{C}}[\text{mineig}(A_2(c))] > 0$. So, by Part (a), we have

$$\begin{aligned}
& \sup_{c\in\mathbb{C}} \left\| A_{2T}(c)^{-1} - A_2^{-1}(c) \right\|^2 \\
\leq \; & \left[\sup_{c\in\mathbb{C}} \left\| A_{2T}(c)^{-1} \right\|^2 \right] \left[\sup_{c\in\mathbb{C}} \left\| A_{2T}(c) - A_2(c) \right\|^2 \right] \left[\sup_{c\in\mathbb{C}} \left\| A_2(c)^{-1} \right\|^2 \right] \\
\leq \; & \frac{2}{\inf_{c\in\mathbb{C}}[\text{mineig}(A_{2T}(c))]} \frac{2}{\inf_{c\in\mathbb{C}}[\text{mineig}(A_2(c))]} \left[\sup_{c\in\mathbb{C}} \left\| A_{2T}(c) - A_2(c) \right\|^2 \right] \\
= \; & o(1),
\end{aligned}$$

as required. The proof for the case of $p = 1$ is similar, and is omitted. ■

Parts (d) and (e). Parts (d) and (e) hold by Lemma 11 and Parts (b) and (c). ■

Lemma 12 *For $j = 1, ..., J$, assume that $h_j(c, \tilde{c})$ are real-valued continuous functions on the product of the compact parameter set $\mathbb{C} \times \mathbb{C}$ with $h_j(c, c) = 0$. Also, for $j = 1, ..., J$, assume that $l_{jT}(x, y)$ are real-valued continuous functions on $[0, 1] \times [0, 1]$. Let $f_T(x, c)$ and $g_T(x, c)$ be continuously differentiable functions from $[0, 1] \times \mathbb{C}$ to $\mathbb{R}$ such that $f_T(x, c)\, g_T(y, c) - f_T(x, \tilde{c})\, g_T(y, \tilde{c}) = \sum_{j=1}^{J} h_j(c, \tilde{c})\, l_{jT}(x, y)$. Assume that $f_T(x, c) \to f(x, c)$, $g_T(y, c) \to g(y, c)$, and $\sup_T \sup_{(r,c)\in[0,1]\times\mathbb{C}} |g_T(r, c)| < \infty$. Suppose that $y_{it} = \left(1 + \frac{c_0}{T}\right) y_{it-1} + \varepsilon_{it}$, where ε_{it} follows Assumption 3. Let Assumption 2 holds for the initial condition y_{i0}. Then, as $(n, T \to \infty)$, the following hold.*

(a) $\frac{1}{n}\sum_{i=1}^{n} \frac{1}{T^2}\sum_{t=1}^{T} y_{it-1}^2 \to_p \sigma^2 \int_0^1 \int_0^r e^{2c_0(r-s)} ds dr$.

(b) $\frac{1}{n}\sum_{i=1}^{n} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{it} f_T\left(\frac{t}{T}, c\right) \right) \left(\frac{1}{T\sqrt{T}} \sum_{t=1}^{T} y_{it-1} g_T\left(\frac{t}{T}, c\right) \right) \to_p \sigma^2 \int_0^1 \int_0^r e^{c_0(r-s)} g(r, c) f(s, c) ds dr$ *uniformly in c.*

(c) $\frac{1}{n}\sum_{i=1}^{n} \left(\frac{1}{T\sqrt{T}} \sum_{t=1}^{T} y_{it-1} f_T\left(\frac{t}{T}, c\right) \right) \left(\frac{1}{T\sqrt{T}} \sum_{t=1}^{T} y_{it-1} g_T\left(\frac{t}{T}, c\right) \right)$
$\to_p \sigma^2 \int_0^1 \int_0^1 f(r, c) g(s, c) \int_0^{r\wedge s} e^{c_0(r+s-2v)} dv ds dr$ *uniformly in c.*

(d) $\frac{1}{n}\sum_{i=1}^{n} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{it} f_T\left(\frac{t}{T}, c\right) \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{it} g_T\left(\frac{t}{T}, c\right) \right) \to_p \sigma^2 \int_0^1 f(r, c) g(r, c) dr$ *uniformly in c.*

Proof

Part (a) From the decomposition (37), we write

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n} \frac{1}{T^2}\sum_{t=1}^{T} y_{it-1}^2 \\
= \; & \frac{1}{n}\sum_{i=1}^{n} \frac{1}{T^2}\sum_{t=1}^{T} x_{it-1}^2 + 2\frac{1}{n}\sum_{i=1}^{n} \frac{1}{T^2}\sum_{t=1}^{T} x_{it-1} R_{it-1} + \frac{1}{n}\sum_{i=1}^{n} \frac{1}{T^2}\sum_{t=1}^{T} R_{it-1}^2 \\
= \; & I_a + 2II_a + III_a, \text{ say.}
\end{aligned}$$

In what follows we show that $I_a \to_p \sigma^2 \int_0^1 \int_0^r e^{2c_0(r-s)} ds dr$ and $II_a, III_a \to_p 0$ as $(n, T \to \infty)$.

For I_a, define $Q_{iT} = \frac{1}{T^2}\sum_{t=1}^{T} x_{it-1}^2$. Note that $\{Q_{iT}\}_{i=1,\dots,n}$ are iid across i. Since

$$T^{-\frac{1}{2}} x_{it} \Rightarrow J_{c_0,i}(r) = \sigma^2 \int_0^r e^{c_0(r-s)} dW_i(s) \tag{42}$$

as $T \to \infty$ (see Phillips, 1987), where W_i is standard Brownian motion, we have by the continuous mapping theorem as $(n, T \to \infty)$,

$$Q_{iT} \Rightarrow Q_i = \sigma^2 \int_0^1 J_{c_0,i}^2(r) dr. \tag{43}$$

Notice that $EQ_i = \sigma^2 \int_0^1 \int_0^r e^{2c_0(r-s)} ds dr$.

We claim $I_a \to_p \sigma^2 \int_0^1 \int_0^r e^{2c_0(r-s)} ds dr$ in joint limits as $(n, T \to \infty)$ by verifying conditions (i) - (iii) in Theorem 7. Condition (iii) is trivial because $C_i = 1$. Condition (i) is obvious from (43). For condition (ii), observe that

$$\begin{aligned} EQ_{iT} &= \sigma^2 \frac{1}{T} \sum_{t=2}^{T} \frac{1}{T} \sum_{s=1}^{t-1} \left[\left(1 + \frac{c_0}{T}\right)^T \right]^{2\frac{t-1-s}{T}} \\ &\to \sigma^2 \int_0^1 \int_0^r e^{(r-s)2c_0} ds dr = EQ_i \text{ as } (n, T \to \infty). \end{aligned}$$

Since $Q_{iT} (\geq 0) \Rightarrow Q_i$ with $EQ_{iT} \to EQ_i$ as $(n, T \to \infty)$, $\{Q_{iT}\}_T$ are uniformly integrable in T by Theorem 5.4 in Billingsley (1968).

Next, we prove that

$$II_a = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} x_{it-1} R_{it-1} \to_p 0,$$

and

$$III_a = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} R_{it-1}^2 \to_p 0 \text{ as } n, T \to \infty,$$

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

Since $\left|\left(1 + \frac{c_0}{T}\right)^{t-1}\right| \leq 1$ and by the Cauchy-Schwarz inequality,

$$\begin{aligned} E|II_a| &= E\left| \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} x_{it-1} R_{it-1} \right| \\ &\leq \frac{1}{n} \sum_{i=1}^{n} \left\{ E\left| \frac{1}{T^2} \sum_{t=2}^{T} x_{it-1} \left(1 + \frac{c_0}{T}\right)^{t-1} y_{i0} \right| \right\} \\ &\leq \sup_i \sqrt{E y_{i0}^2} \frac{1}{\sqrt{T}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T} \sum_{t=1}^{T} \sqrt{E\left(\frac{x_{it-1}}{\sqrt{T}}\right)^2} \\ &= O\left(\frac{1}{\sqrt{T}}\right), \end{aligned}$$

where the last inequality holds by (39) and by Assumption 2. Thus,

$$II_a = O_p\left(\frac{1}{\sqrt{T}}\right) = o_p(1).$$

Similarly,

$$
\begin{aligned}
E\left|III_a\right| &= E\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}R_{it-1}^2\right| \leq \left(\sup_i Ey_{i0}^2\right)\left(\frac{1}{T^2}\sum_{t=2}^{T}\left(1+\frac{c_0}{T}\right)^{2(t-1)}\right) \\
&\leq \frac{1}{T}\left(\sup_i Ey_{i0}^2\right) = O\left(\frac{1}{T}\right),
\end{aligned}
$$

and so

$$
III_a = O_p\left(\frac{1}{T}\right) = o_p(1).
$$

Therefore we have all the required results to complete the proof of part (a). ■

Part (b) Using (37), we write

$$
\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}f_T\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y_{it-1}g_T\left(\frac{t}{T},c\right)\right) = I_b + II_b,
$$

where

$$
\begin{aligned}
I_b &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}f_T\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}x_{it-1}g_T\left(\frac{t}{T},c\right)\right), \\
II_b &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}f_T\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}R_{it-1}g_T\left(\frac{t}{T},c\right)\right).
\end{aligned}
$$

We will show that

$$
\text{Part (b}_1\text{): } I_b \to_p \sigma^2\int_0^1\int_0^r e^{c_0(r-s)}g(r,c)f(s,c)dsdr \text{ uniformly in } c
$$

and

$$
\text{Part (b}_2\text{): } II_b \to_p 0 \text{ uniformly in } c
$$

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

First, we establish Part (b$_1$) for fixed c (pointwise convergence). Now, as in Part (a), we apply Theorem 7. Let

$$
\begin{aligned}
Q_{iT}(c) &= \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}f_T\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}x_{it-1}g_T\left(\frac{t}{T},c\right)\right), \\
\text{and } Q_i(c) &= \sigma^2\left(\int_0^1 f(r,c)dW_i(r)\right)\left(\int_0^1 g(r,c)J_{c_0,i}(r)dr\right).
\end{aligned}
$$

Using (42) and the extended-continuous mapping theorem (see Theorem 1.11.1 in van der Vaart and Wellner, 1996), we can show that

$$
Q_{iT}(c) \Rightarrow Q_i(c) \tag{44}
$$

as $T \to \infty$ for fixed c, which verifies condition (i) in Theorem 7. Condition (iii) is trivial because $C_i = 1$. Condition (ii) holds for fixed c if

$$
Q_{1iT}(c) = \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}f_T\left(\frac{t}{T},c\right)\right)^2
$$

and

$$Q_{2iT}(c)=\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}x_{it-1}g_T\left(\frac{t}{T},c\right)\right)^2$$

are uniformly integrable in T for fixed c. Notice that $Q_{1iT}(c)\Rightarrow Q_{1i}(c)=\sigma^2\left(\int_0^1 f(r,c)dW_i(r)\right)^2>0$, and $EQ_{1iT}(c)=\sigma^2\frac{1}{T}\sum_{t=1}^{T}f_T\left(\frac{t}{T},c\right)^2\to\sigma^2\int_0^1 f(r,c)^2dr=EQ_{1i}(c)$ as $T\to\infty$ for all i. By Theorem 5.4 in Billingsley (1968), it follows that $Q_{1iT}(c)$ are uniformly integrable in T for fixed c. In a similar fashion, $Q_{2iT}(c)$ is also uniformly integrable in T for fixed c. Therfore, as $(n,T\to\infty)$,

$$I_b\to_p \Omega\int_0^1\int_0^r e^{c_0(r-s)}g(r,c)f(s,c)dsdr \text{ for fixed } c.$$

Next, define $X_{nT}(c)=\frac{1}{n}\sum_{i=1}^{n}Q_{iT}(c)$. To complete the proof, we need to show that $X_{nT}(c)$ is stochastically equicontinuous. That is, for given $\varepsilon>0$ and $\eta>0$, there exists $\delta>0$ such that

$$\limsup_{(n,T\to\infty)}P\left\{\sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}}|X_{nT}(c)-X_{nT}(\tilde{c})|>\varepsilon\right\}<\eta.$$

Then, since the parameter set $\mathbb{C}$ is compact, the pointwise convergence of $X_{nT}(c)$ and the stochastic equicontinuity of $X_{nT}(c)$ imply uniform convergence.

Now we show the stochastic equicontinuity of $X_{nT}(c)$. First, notice that

$$\begin{aligned}
&\sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}}|X_{nT}(c)-X_{nT}(\tilde{c})|\\
=&\sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}x_{is-1}\left\{f_T\left(\frac{t}{T},c\right)g_T\left(\frac{s}{T},c\right)-f_T\left(\frac{t}{T},\tilde{c}\right)g_T\left(\frac{s}{T},\tilde{c}\right)\right\}\right|\\
=&\sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}x_{is-1}\left\{\sum_{j=1}^{J}h_j(c,\tilde{c})\,l_{jT}\left(\frac{t}{T},\frac{s}{T}\right)\right\}\right|\\
=&\sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}}\left|\sum_{j=1}^{J}h_j(c,\tilde{c})\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}x_{is-1}l_{jT}\left(\frac{t}{T},\frac{s}{T}\right)\right|\\
\leq&\left[\sup_{1\leq j\leq J}\sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}}|h_j(c,\tilde{c})|\right]\left[\sum_{j=1}^{J}\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}x_{is-1}l_{jT}\left(\frac{t}{T},\frac{s}{T}\right)\right|\right].
\end{aligned}$$

Since $h_j(c,\tilde{c})$ is continuous on the compact set with $h_j(c,c)=0$ for all $j=1,...,J$, we can make $\sup_{1\leq j\leq J}\sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}}|h_j(c,\tilde{c})|$ arbitrarily small by choosing a small $\delta>0$. Also, under the assumptions in the lemma, it is not difficult to show that $\sum_{j=1}^{J}\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\varepsilon_{it}x_{is-1}l_{jT}\left(\frac{t}{T},\frac{s}{T}\right)\right|=O_p(1)$. Therefore, $X_{nT}(c)$ is stochastically equicontinuous, and $I_b\to_p\sigma^2\int_0^1\int_0^r e^{c_0(r-s)}g(r,c)f(s,c)dsdr$ uniformly in c.

Next, for Part (b_2), notice that

$$
\begin{aligned}
|II_b| &\leq \frac{1}{n\sqrt{T}}\sum_{i=1}^{n}\left|y_{i0}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}f_T\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T}\sum_{t=1}^{T}\left(1+\frac{c_0}{T}\right)^{t-1}g_T\left(\frac{t}{T},c\right)\right)\right| \\
&\leq \frac{1}{\sqrt{T}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}y_{i0}^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}f_T\left(\frac{t}{T},c\right)\right)^2\left(\frac{1}{T}\sum_{t=1}^{T}\left(1+\frac{c_0}{T}\right)^{t-1}g_T\left(\frac{t}{T},c\right)\right)^2} \\
&\leq \frac{1}{\sqrt{T}}\left(\frac{1}{T}\sum_{t=1}^{T}\left(1+\frac{c_0}{T}\right)^{t-1}\left|g_T\left(\frac{t}{T},c\right)\right|\right)\sqrt{\frac{1}{n}\sum_{i=1}^{n}y_{i0}^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}f_T\left(\frac{t}{T},c\right)\right)^2} \\
&\leq \frac{1}{\sqrt{T}}\sup_T\sup_{(r,c)\in[0,1]\times\mathbb{C}}|g_T(r,c)|\sqrt{\frac{1}{n}\sum_{i=1}^{n}y_{i0}^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}f_T\left(\frac{t}{T},c\right)\right)^2}
\end{aligned}
$$

Using similar arguments in the proof of the limit of I_a, it is possible to prove that as $(n,T\to\infty)$,

$$
\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}f_T\left(\frac{t}{T},c\right)\right)^2 \to_p \sigma^2\left(\int_0^1 f(r,c)^2\,dr\right)^2 \text{ uniformly in } c.
$$

Also, $\frac{1}{n}\sum_{i=1}^{n}y_{i0}^2 = O_p(1)$ by Assumption 2, and $\sup_T\sup_{(r,c)\in[0,1]\times\mathbb{C}}|g_T(r,c)| < \bar{K}$. Thus,

$$
II_b = O_p\left(\frac{1}{\sqrt{T}}\right) = o_p(1) \text{ uniformly in } c,
$$

and we complete the proof of Part (b_2). ■

Part (c) and Part (d) The proofs of Parts (c) and (d) are similar to that of Part (b) and they are omitted. ■

The following lemma is important in establishing asymptotic normality of the GMM estimator $\hat{c}$. To simplify notation, let

$$
\begin{aligned}
l_{1pT}(t,s,c) &= \widehat{\Delta_c g_{pt}}' A_{pT}(c)^{-1}\widehat{\Delta_c g_{ps}} \\
l_{2pT}(t,s,c) &= \widehat{\Delta_c g_{pt}}' A_{pT}(c)^{-1} g_{ps-1}D_{p,T}^{-1} \\
l_{3pT}(t,s,c) &= \widehat{\Delta_c g_{pt}}' A_{pT}(c)^{-1} B_{pT}(c)A_{pT}(c)^{-1}\widehat{\Delta_c g_{ps}},
\end{aligned}
$$

and

$$
\begin{aligned}
l_{1p}(r,s,c) &= \dot{g}_{pc}(r)' A_p(c)^{-1}\dot{g}_{pc}(s) \\
l_{2p}(r,s,c) &= \dot{g}_{pc}(r)' A_p(c)^{-1} g_p(s) \\
l_{3p}(r,s,c) &= \dot{g}_{pc}(r)' A_p(c)^{-1} B_p(c)A_p(c)^{-1}\dot{g}_{pc}(s) \\
l_{4p} &= \int_0^1 g_p(r)\,g_p(r)'\,dr.
\end{aligned}
$$

Lemma 13 *Suppose that* $x_{it} = \exp\left(\frac{c_0}{T}\right) x_{it-1} + \varepsilon_{it}$, *where* ε_{it} *are iid* $\left(0, \sigma^2\right)$ *with finite fourth moments and* $x_{i0} = 0$ *for all* i. *Let*

$$Q_{1iT} = \frac{1}{T} \sum_{t=1}^{T} x_{it-1} \varepsilon_{it}$$

$$Q_{2iT} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{T\sqrt{T}} \sum_{s=1}^{T} \varepsilon_{it} x_{is-1} \tilde{h}_{pT}(t,s) - \sigma^2 \omega_{pT}(c_0)$$

$$Q_{3iT} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{T\sqrt{T}} \sum_{s=1}^{T} \varepsilon_{it} x_{is-1} l_{1pT}(t,s,c_0) - \sigma^2 \lambda_{pT}(c_0)$$

$$Q_{4iT} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \varepsilon_{it} \varepsilon_{is} l_{2pT}(t,s,c_0) - \sigma^2 tr\left(A_{pT}(c_0)^{-1} B_{pT}(c_0)\right)$$

$$Q_{5iT} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \varepsilon_{it} \varepsilon_{is} l_{3pT}(t,s,c_0) - \sigma^2 tr\left(A_{pT}(c_0)^{-1} B_{pT}(c_0)\right)$$

$$\text{and } Q_{iT} = \left(Q_{1iT}, Q_{2iT}, Q_{3iT}, Q_{4iT}, Q_{5iT}\right)'. \quad (45)$$

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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q_{iT} \Rightarrow N\left(0, \sigma^4 \Phi(c_0)\right),$$

where

$$\Phi(c_0) = \begin{bmatrix} \Phi_{11}(c_0) & \Phi_{12}(c_0) & \Phi_{13}(c_0) & \Phi_{14}(c_0) & \Phi_{15}(c_0) \\ \Phi_{12}(c_0) & \Phi_{22}(c_0) & \Phi_{23}(c_0) & \Phi_{24}(c_0) & \Phi_{25}(c_0) \\ \Phi_{13}(c_0) & \Phi_{23}(c_0) & \Phi_{33}(c_0) & \Phi_{34}(c_0) & \Phi_{35}(c_0) \\ \Phi_{14}(c_0) & \Phi_{24}(c_0) & \Phi_{34}(c_0) & \Phi_{44}(c_0) & \Phi_{45}(c_0) \\ \Phi_{15}(c_0) & \Phi_{25}(c_0) & \Phi_{35}(c_0) & \Phi_{45}(c_0) & \Phi_{55}(c_0) \end{bmatrix} \quad (46)$$

and

$$\Phi_{11}(c_0) = \int_0^1 \int_0^r e^{2c_0(r-s)} ds dr,$$

$$\Phi_{12}(c_0) = \int_0^1 \int_0^1 \int_0^{r \wedge s} e^{c_0(r+s-2v)} \tilde{h}_p(r,s) dv ds dr + \int_0^1 \int_0^r \int_0^s e^{c_0(r-v)} \tilde{h}_p(v,r) dv ds dr,$$

$$\Phi_{13}(c_0) = \int_0^1 \int_0^r \int_0^s e^{c_0(r-v)} l_{1p}(r,v,c_0) dv ds dr + \int_0^1 \int_0^1 \int_0^{r \wedge s} e^{c_0(r+s-2v)} l_{1p}(r,s,c_0) dv ds dr,$$

$$\Phi_{14}(c_0) = \int_0^1 \int_0^r e^{c_0(r-s)} l_{2p}(r,s,c_0) ds dr + \int_0^1 \int_0^r e^{c_0(r-s)} l_{2p}(s,r,c_0) ds dr,$$

$$\Phi_{15}(c_0) = \int_0^1 \int_0^r e^{c_0(r-s)} l_{3p}(r,s,c_0) ds dr + \int_0^1 \int_0^r e^{c_0(r-s)} l_{3p}(s,r,c_0) ds dr,$$

$$\Phi_{22}(c_0) = \int_0^1 \int_0^1 \int_0^{r \wedge s} e^{c_0(r+s-2v)} \tilde{h}_p(r,s)\, dvdsdr$$
$$+ \int_0^1 \int_0^1 \int_0^r \int_0^s e^{c_0(r-v)} e^{c_0(s-q)} \tilde{h}_p(r,q)\, \tilde{h}_p(s,v)\, dqdvdsdr,$$

$$\Phi_{23}(c_0) = \int_0^1 \int_0^1 \int_0^1 \int_0^{s \wedge v} e^{c_0(s+v-2q)} \tilde{h}_p(r,s)\, l_{1p}(r,v,c_0)\, dqdvdsdr$$
$$+ \int_0^1 \int_0^1 \int_0^r \int_0^s e^{c_0(r-v)} e^{c_0(s-q)} \tilde{h}_p(r,q)\, l_{1p}(s,v,c_0)\, dqdvdsdr,$$

$$\Phi_{24}(c_0) = \int_0^1 \int_0^r \int_0^1 e^{c_0(r-s)} \tilde{h}_p(r,v)\, l_{2p}(v,s,c_0)\, dvdsdr$$
$$+ \int_0^1 \int_0^r \int_0^1 e^{c_0(r-s)} \tilde{h}_p(r,v)\, l_{2p}(s,v,c_0)\, dvdsdr,$$

$$\Phi_{25}(c_0) = \int_0^1 \int_0^r \int_0^1 e^{c_0(r-s)} \tilde{h}_p(r,v)\, l_{3p}(v,s,c_0)\, dvdsdr$$
$$+ \int_0^1 \int_0^r \int_0^1 e^{c_0(r-s)} \tilde{h}_p(r,v)\, l_{3p}(s,v,c_0)\, dvdsdr,$$

$$\Phi_{33}(c_0) = \int_0^1 \int_0^1 \int_0^{r \wedge s} e^{c_0(r+s-2v)} l_{1p}(r,s,c_0)\, dvdsdr$$
$$+ \int_0^1 \int_0^1 \int_0^r \int_0^s e^{c_0(r-v)} e^{c_0(s-q)} l_{1p}(r,q,c_0)\, l_{1p}(s,v,c_0)\, dqdvdsdr,$$

$$\Phi_{34}(c_0) = \int_0^1 \int_0^r e^{c_0(r-s)} l_{2p}(r,s,c_0)\, dsdr + \int_0^1 \int_0^r e^{c_0(r-s)} l_{3p}(r,s,c_0)\, dsdr,$$

$$\Phi_{35}(c_0) = \int_0^1 \int_0^r e^{c_0(r-s)} l_{3p}(r,s,c_0)\, dsdr + \int_0^1 \int_0^r e^{c_0(r-s)} l_{3p}(s,r,c_0)\, dsdr,$$

$$\Phi_{44}(c_0) = \left(vecA_p(c_0)^{-1}\right)' vecl_{4p}(c_0) + tr\left(A_p(c_0)^{-1} B_p(c_0)' A_p(c_0)^{-1} B_p(c_0)\right),$$

$$\Phi_{45}(c_0) = tr\left(A_p(c_0)^{-1} B_p(c_0) A_p(c_0)^{-1} B_p(c_0)'\right) + tr\left(A_p(c_0)^{-1} B_p(c_0) A_p(c_0)^{-1} B_p(c_0)\right),$$

$$\Phi_{55}(c_0) = tr\left(A_p(c_0)^{-1} B_p(c_0) A_p(c_0)^{-1} B_p(c_0)'\right) + tr\left(A_p(c_0)^{-1} B_p(c_0) A_p(c_0)^{-1} B_p(c_0)\right).$$

Proof

The proof uses Theorem 8. First, a direct calculation shows that $EQ_{iT} = 0$. Let $\Phi_{nT}(c_0) = EQ_{iT}Q'_{iT}$. Notice that Q_{iT} are iid $(0, \Phi_{nT}(c_0))$ across i. As $T \to \infty$,

$$Q_{iT} \Rightarrow \sigma^2 Q_i,$$

where

$$\begin{aligned}
Q_i &= (Q_{1i}, Q_{2i}, Q_{3i}, Q_{4i}, Q_{5i})' \\
Q_{1i} &= \int_0^1 J_{c_0,i}(r)\, dW_i(r) \\
Q_{2i} &= \int_0^1 \int_0^1 J_{c_0,i}(r)\, \tilde{h}_p(r,s)\, dW_i(s)\, dr - \omega_p(c_0) \\
Q_{3i} &= \int_0^1 \int_0^1 l_{1p}(r,s,c_0)\, dW_i(r)\, dW_i(s) - \lambda_p(c_0) \\
Q_{4i} &= \int_0^1 \int_0^1 l_{2p}(r,s,c_0)\, dW_i(r)\, dW_i(s) - tr\left(A_p(c_0)^{-1} B_p(c_0)\right) \\
Q_{5i} &= \int_0^1 \int_0^1 l_{3p}(r,s,c_0)\, dW_i(r)\, dW_i(s) - tr\left(A_p(c_0)^{-1} B_p(c_0)\right).
\end{aligned}$$

Also, a direct calculation shows that as $T \to \infty$,

$$\Phi_{nT}(c_0) = EQ_{iT}Q'_{iT} \to \sigma^4 EQ_iQ'_i = \sigma^4 \Phi(c_0).$$

Let l be any (5×1) vector with $\|l\| = 1$. We consider two cases.

Case 1: If $l'\Phi(c_0)\, l > 0$.

To establish the desired result with a joint limit, we apply Theorem 7. Condition (i) holds because it is assumed that $l'\Phi(c_0)\, l > 0$. Conditions (ii) is trivial. Finally condition (iii), viz.

$$(l'Q_{iT})^2 \text{ are uniformly integrable in } T,$$

holds because $(l'Q_{iT})^2 \Rightarrow (l'Q_i)^2$ as $T \to \infty$ by the continuous mapping theorem with $E(l'Q_{iT})^2 = l'\Phi_{nT}(c_0)\, l \to \sigma^4 l'\Phi(c_0)\, l = \sigma^4 E(l'Q_i)^2$, and by applying Theorem 5.4 of Billingsley (1968).

Case 2: If $l'\Phi(c_0)\, l = 0$. Since $l'\Phi_{nT}(c_0)\, l \to l'\Phi(c_0)\, l = 0$,

$$E\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} l'Q_{iT}\right)^2 = l'\Phi_{nT}(c_0)\, l \to 0,$$

which leads to $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} l'Q_{iT} \to_p 0$. By the Cramér-Wold device, it follows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q_{iT} \Rightarrow N\left(0, \sigma^4 \Phi(c_0)\right). \blacksquare$$

6.3 Appendix B: Proofs of Section 4

Proof of Lemma 2.

By definition,

$$
\begin{aligned}
& \hat{\sigma}^2 - \tilde{\sigma}^2 \\
= \quad & (\hat{\rho} - \rho_0)^2 \frac{1}{nT} \sum_{i=1}^{n} \underset{\sim}{y}'_{i,-1} \underset{\sim}{y}_{i,-1} - 2(\hat{\rho} - \rho_0) \frac{1}{nT} \sum_{i=1}^{n} \underset{\sim}{\varepsilon}'_i \underset{\sim}{y}_{i,-1}.
\end{aligned}
$$

Since $\hat{\rho} - \rho_0 = O_p(1)$, $\frac{1}{nT^2} \sum_{i=1}^{n} \underset{\sim}{y}'_{i,-1} \underset{\sim}{y}_{i,-1} = O_p(1)$, and $\frac{1}{nT} \sum_{i=1}^{n} \underset{\sim}{\varepsilon}'_i \underset{\sim}{y}_{i,-1} = O_p(1)$ by Lemma 12 and Theorem 1 of Moon and Phillips (2000),

$$
\hat{\sigma}^2 - \tilde{\sigma}^2 = O_p\left(\frac{1}{T}\right).
$$

Next, since

$$
E\left(\tilde{\sigma}^2 - \sigma^2\right)^2 = O\left(\frac{1}{\sqrt{nT}}\right),
$$

it follows that

$$
\tilde{\sigma}^2 - \sigma^2 = O_p\left(\frac{1}{\sqrt{nT}}\right). \blacksquare
$$

Proof of Lemma 3.

We show separately the following

$$
\frac{1}{n} \sum_{i=1}^{n} \left(m_{1iT}(c) - m_1(c)\right) \to_p 0, \tag{47}
$$

and

$$
\frac{1}{n} \sum_{i=1}^{n} \left(m_{2iT}(c) - m_2(c)\right) \to_p 0, \tag{48}
$$

uniformly in c.

First, by definition and the triangle inequality, we have

$$
\begin{aligned}
m_{1,iT}(c) &= \frac{1}{T}\left(\underset{\sim}{z}_i - \left(1 + \frac{c}{T}\right)\underset{\sim}{z}_{i,-1}\right)' \underset{\sim}{z}_{i,-1} + \hat{\sigma}^2 \omega_{pT}(c) \\
&= \frac{1}{T} \underset{\sim}{\varepsilon}'_i \underset{\sim}{y}_{i,-1} - (c - c_0) \frac{1}{T^2} \underset{\sim}{y}'_{i,-1} \underset{\sim}{y}_{i,-1} + \hat{\sigma}^2 \omega_{pT}(c) \\
&= \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} y_{i,t-1} - \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \varepsilon_{it} y_{i,s-1} \tilde{h}_{pT}(t,s) - (c - c_0) \frac{1}{T^2} \sum_{t=1}^{T} \left(\underset{\sim}{y}_{i,-1}\right)_t^2 + \hat{\sigma}^2 \omega_{pT}(c).
\end{aligned}
$$

So,

$$
\begin{aligned}
& \left| \frac{1}{n} \sum_{i=1}^{n} \left(m_{1iT}(c) - m_1(c) \right) \right| \\
\leq & \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ \begin{array}{c} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} y_{it-1} - \left(\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \varepsilon_{it} y_{is-1} \tilde{h}_{pT}(t,s) - \sigma^2 \omega_p(c_0) \right) \\ +\sigma^2 \left(\omega_{pT}(c) - \omega_p(c) \right) \\ -(c - c_0) \left(\frac{1}{T^2} \sum_{t=1}^{T} \left(\underset{\sim}{y}_{i,-1} \right)_t^2 - \sigma^2 \psi_p(c_0) \right) \end{array} \right\} \right| \\
& + \left| \hat{\sigma}^2 - \sigma^2 \right| \left| \omega_{pT}(c) \right| \\
\leq & \left| \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} y_{it-1} \right| \\
& + \left| \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \varepsilon_{it} y_{is-1} \tilde{h}_{pT}(t,s) - \sigma^2 \omega_p(c_0) \right) \right| \\
& + |c - c_0| \left| \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{T^2} \sum_{t=1}^{T} \left(\underset{\sim}{y}_{i,-1} \right)_t^2 - \sigma^2 \psi_p(c_0) \right) \right| \\
& + \sigma^2 \left| \omega_{pT}(c) - \omega_p(c) \right| + \left| \hat{\sigma}^2 - \sigma^2 \right| \left| \omega_{pT}(c) \right| \\
= & \; I + II + III + IV + V, \text{ say.}
\end{aligned}
$$

Notice that the two terms I and II are independent of c, and by Lemma 9 of Moon and Phillips (1999b), $I, II \to_p 0$ as $(n, T \to \infty)$. Next, $III \to_p 0$ uniformly in c because $|\frac{1}{n}\sum_{i=1}^{n}(\frac{1}{T^2}\sum_{t=1}^{T}(\underset{\sim}{y}_{i,-1})_t^2 - \sigma^2\psi_p(c_0))|$ that is independent of c converges in probability to zero as $(n, T \to \infty)$ by Lemma 9 of Moon and Phillips (1999b). Next, $IV \to 0$ uniformly in c by Corollary 1(d). Finally, since $\hat{\sigma}^2 - \sigma^2 = o_p(1)$ by Lemma 2, and $\sup_{c \in \mathbb{C}} \omega_{pT}(c) < \bar{K}$ for some finite $\bar{K}$, V converges in probability to zero uniformly in c. Therefore, $\frac{1}{n}\sum_{i=1}^{n}(m_{1iT}(c) - m_1(c)) \to_p 0$ uniformly in c as $(n, T \to \infty)$.

Next, to prove (48), noting by defintion that

$$
\Delta_c z_{it} = \beta_{i0}' \Delta_c g_{pt} - (c - c_0) \frac{y_{it-1}}{T} + \varepsilon_{it},
$$

we write

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} m_{2iT}(c) \\
=\ &\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}y_{it-1}\right)-(c-c_0)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}y_{it-1}^2 \\
&-\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}\varepsilon_{it}\right)' A_{pT}(c)^{-1}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}y_{it-1}\right)\right]+\sigma^2\lambda_{pT}(c) \\
&+(c-c_0)\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}y_{it-1}\right)' A_{pT}(c)^{-1}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}y_{it-1}\right)\right] \\
&-\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}\varepsilon_{it}\right)' A_{pT}(c)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_{pt-1}D_{p,T}^{-1}\varepsilon_{it}\right)\right] \\
&+(c-c_0)\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}y_{it-1}\right)' A_{pT}(c)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_{pt-1}D_{p,T}^{-1}\varepsilon_{it}\right)\right] \\
&+(c-c_0)\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}\varepsilon_{it}\right)' A_{pT}(c)^{-1}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}g_{pt-1}D_{p,T}^{-1}y_{it-1}\right)\right] \\
&-(c-c_0)^2\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}y_{it-1}\right)' A_{pT}(c)^{-1}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}g_{pt-1}D_{p,T}^{-1}y_{it-1}\right)\right] \\
&+\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}\varepsilon_{it}\right)' A_{pT}(c)^{-1}B_{pT}(c)A_{pT}(c)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}\varepsilon_{it}\right)\right] \\
&-(c-c_0)\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}y_{it-1}\right)' A_{pT}(c)^{-1}B_{pT}(c)A_{pT}(c)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}\varepsilon_{it}\right)\right] \\
&-(c-c_0)\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}\varepsilon_{it}\right)' A_{pT}(c)^{-1}B_{pT}(c)A_{pT}(c)^{-1}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}y_{it-1}\right)\right] \\
&+(c-c_0)^2\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}y_{it-1}\right)' A_{pT}(c)^{-1}B_{pT}(c)A_{pT}(c)^{-1}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\widehat{\Delta_c g_{pt}}y_{it-1}\right)\right] \\
&+\left(\hat{\sigma}^2-\sigma^2\right)\lambda_{pT}(c).
\end{aligned}
$$

Notice that $\widehat{\Delta_c g_{pt}}$ and $D_{p,T}^{-1}g_{pt-1}$ satisfy the conditions for $f_T(x,c)$ and $g_T(x,c)$ in Lemma 12. The desired result, then, follows by Corollary 1 and by applying Lemma 12 together with $\hat{\sigma}^2-\sigma^2=o_p(1)$ (see Lemma 2) and the boundedness of $\lambda_{pT}(c)$ on the compact parameter set $\mathbb{C}$. ■

Proof of Lemma 4.

The proof is similar to that of Lemma 3, and is omitted. ■

Proof of Lemma 5.

First, using (37) and by Lemma 2, we may write

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{1iT}\left(c_0\right) \\
= \quad & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(Q_{1iT} - Q_{2iT}\right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{1iT} + O_p\left(\frac{\sqrt{n}}{T}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \qquad (49)
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{2iT}\left(c_0\right) \\
= \quad & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(Q_{1iT} - Q_{3iT} - Q_{4iT} + Q_{5iT}\right) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{2iT} + O_p\left(\frac{\sqrt{n}}{T}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \qquad (50)
\end{aligned}
$$

where

$$
\begin{aligned}
R_{1iT} &= y_{i0} \left(\frac{1}{T} \sum_{t=1}^{T} \left[\left(1 + \frac{c_0}{T}\right)^T\right]^{\frac{t-1}{T}} \varepsilon_{it}\right) - y_{i0} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{T\sqrt{T}} \sum_{s=1}^{T} \varepsilon_{it} \left[\left(1 + \frac{c_0}{T}\right)^T\right]^{\frac{s-1}{T}} \tilde{h}_{pT}\left(t, s\right)\right) \\
&= R_{11iT} - R_{12iT},
\end{aligned}
$$

and

$$
\begin{aligned}
R_{2iT} &= y_{i0} \left(\frac{1}{T} \sum_{t=1}^{T} \left[\left(1 + \frac{c_0}{T}\right)^T\right]^{\frac{t-1}{T}} \varepsilon_{it}\right) - y_{i0} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{T\sqrt{T}} \sum_{s=1}^{T} \varepsilon_{it} \left[\left(1 + \frac{c_0}{T}\right)^T\right]^{\frac{s-1}{T}} l_{1pT}\left(t, s, c_0\right)\right) \\
&= R_{21iT} - R_{22iT}.
\end{aligned}
$$

Notice that

$$
\begin{aligned}
& E\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{11iT}\right)^2 = \frac{1}{n} \sum_{i=1}^{n} E R_{11iT}^2 \\
\leq \quad & \frac{\sigma^2}{T} \left(\sup_i E y_{i0}^2\right) \left(\frac{1}{T} \sum_{t=1}^{T} \left[\left(1 + \frac{c_0}{T}\right)^T\right]^{\frac{2(t-1)}{T}}\right) = O\left(\frac{1}{T}\right),
\end{aligned}
$$

where the first equality holds because $E R_{11iT} = 0$ and R_{11iT} is independent across i (Assumption 3) and the second equality holds by Assumptions 2. Similary, it follows that

$$
\begin{aligned}
& E\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} R_{12iT}\right)^2 = \frac{1}{n} \sum_{i=1}^{n} E R_{12iT}^2 \\
\leq \quad & \frac{\sigma^2}{T} \left(\sup_i E y_{i0}^2\right) \left(\frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{T} \sum_{s=1}^{T} \left[\left(1 + \frac{c_0}{T}\right)^T\right]^{\frac{s-1}{T}} \tilde{h}_{pT}\left(t, s\right)\right)^2\right) \\
= \quad & O\left(\frac{1}{T}\right).
\end{aligned}
$$

Therefore,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_{1iT} = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{51}$$

By using similar arguments, it is possible to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_{2iT} = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{52}$$

In view of (49) - (52), as $(n, T \to \infty)$ following Assumption 5,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} m_{1iT}(c_0) \\ m_{2iT}(c_0) \end{pmatrix} = J'\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q_{iT}\right) J + o_p(1).$$

The required result follows by Lemma 5. ■

Proof of Lemma 6.
Part (a).

By definition and by the Cauchy-Schwarz inequality,

$$\begin{aligned} & \sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left|B_{nT}\mathcal{R}_{1nT}(c, c_0)\right| \\ \le\ & 2\left\|B_{nT}M_{nT}(c_0)\right\|\left\|\hat{W}\right\| \sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left\|r_{nT}(c, c_0)\right\|. \end{aligned}$$

By Lemma 5 and Assumption 6, we have $\left\|B_{nT}M_{nT}(c_0)\right\|\left\|\hat{W}\right\| = O_p(1)$. Thus, to complete the proof, it is enough to show that $\sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left\|r_{nT}(c, c_0)\right\| = o_p(1)$. Notice by definition and the triangle inequality that

$$\begin{aligned} & \sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left\|r_{nT}(c, c_0)\right\| \\ \le\ & \sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left|r_{1nT}(c, c_0)\right| + \sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left|r_{2nT}(c, c_0)\right| \\ \le\ & \sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left|\frac{1}{n}\sum_{i=1}^{n}\left(dm_{1iT}(c) - dm_{1iT}(c_0)\right)\right| + \sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left|\frac{1}{n}\sum_{i=1}^{n}\left(dm_{2iT}(c) - dm_{2iT}(c_0)\right)\right|, \end{aligned}$$

where the last line holds because c_k^+ locates between c and c_0 for $k = 1, 2$.

Notice that

$$\begin{aligned} & \sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left|\frac{1}{n}\sum_{i=1}^{n}\left(dm_{1iT}(c) - dm_{1iT}(c_0)\right)\right| \\ \le\ & \sup_{c\in\mathbb{C}}\left|\frac{1}{n}\sum_{i=1}^{n}\left(dm_{1iT}(c) - \hat{\sigma}^2 dm_1(c)\right)\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\left(dm_{1iT}(c_0) - \hat{\sigma}^2 dm_1(c_0)\right)\right| \\ & +\hat{\sigma}^2 \sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left|dm_1(c) - dm_1(c_0)\right|. \end{aligned}$$

Then, the first term and the second term in the last line are $o_p(1)$ by Lemma 4 and the last term is also $o_p(1)$ because $dm(c)$ is continuous in c and $\hat{\sigma}^2$ has a finite limit. Therefore $\sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left|r_{1nT}(c, c_0)\right| = o_p(1)$. Similarly, it follows that $\sup_{c\in\mathbb{C}:|c-c_0|\le\gamma_{nT}} \left|r_{2nT}(c, c_0)\right| = o_p(1)$, and we complete the proof. ■

Part (b).

The proof of Part (b) is similar to that of Part (a) and is omitted. ■

Proof of Theorem 2.

We employ similar arguments to the proof of Theorem 1 of Andrews (1999). Define $\hat{\theta}_{nT} = B_{nT}(\hat{c} - c_0)$. Then,

$$\begin{aligned} o_p(1) &\leq B_{nT}^2 (Z_{nT}(c_0) - Z_{nT}(\hat{c})) \\ &= -\mathcal{H}_{nT}\hat{\theta}_{nT}^2 + 2\mathcal{H}_{nT}(B_{nT}\mathcal{S}_{nT})\hat{\theta}_{nT} \\ &\quad -\hat{\theta}_{nT}B_{nT}\mathcal{R}_{1nT}(\hat{c}, c_0) - \hat{\theta}_{nT}^2\mathcal{R}_{2nT}(\hat{c}, c_0). \end{aligned}$$

From Lemmas 4 and 5 and Assumption 6, we have $\mathcal{H}_{nT}$, $\mathcal{H}_{nT}^{-1} = O_p(1)$ and positive with probability one and $B_{nT}\mathcal{S}_{nT} = O_p(1)$. Also, by Lemma 6, $B_{nT}\mathcal{R}_{1nT}(\hat{c}, c_0) = o_p(1)$ and $\mathcal{R}_{2nT}(\hat{c}, c_0) = o_p(1)$. Then,

$$o_p(1) \leq -\left|\hat{\theta}_{nT}\right|^2 + 2O_p(1)\left|\hat{\theta}_{nT}\right| + \left|\hat{\theta}_{nT}\right| o_p(1) + \left|\hat{\theta}_{nT}\right|^2 o_p(1),$$

which is rearranged as

$$\left|\hat{\theta}_{nT}\right|^2 \leq 2O_p(1)\left|\hat{\theta}_{nT}\right| + o_p(1).$$

Then, the required result,

$$\hat{\theta}_{nT} = O_p(1),$$

follows by relation (7.4) on page 1377 of Andrews (1999). ■

Proof of Theorem 3.

To complete the proof, it is enough to show (a) $B_{nT}(\hat{c} - c_0) = B_{nT}(\hat{c}_q - c_0) + o_p(1)$ and (b) $B_{nT}(\hat{c}_q - c_0) = \hat{\phi}_{nT} + o_p(1)$.

Part (a). Recall that $\frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} = O_p(1)$ by Lemmas 4 and 5 and Assumption 6. Then, it follows by the definition of $B_{nT}(\hat{c}_q - c_0)$ that

$$\left(B_{nT}(\hat{c}_q - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2 \leq \left(\frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2 = O_p(1),$$

which leads to

$$B_{nT}(\hat{c}_q - c_0) = \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} + O_p(1) = O_p(1).$$

From this we find that $\hat{c}_q$ is also $B_{nT}\left(= \sqrt{n}\right) -$ consistent.

Notice that we have

$$\begin{aligned} o_p(1) &\leq B_{nT}^2 Z_{nT}(\hat{c}_q) - B_{nT}^2 Z_{nT}(\hat{c}) \\ &= \left(B_{nT}(\hat{c}_q - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2 - \left(B_{nT}(\hat{c} - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2 + o_p(1) \\ &\leq o_p(1), \end{aligned}$$

where the first line holds by the definition of $\hat{c}$, the second line holds since $B_{nT}(\hat{c}_q - c_0)$, $B_{nT}(\hat{c} - c_0) = O_p(1)$ and by Lemma 6, and the last $o_p(1)$ bound holds because $\left(B_{nT}(\hat{c}_q - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2 - \left(B_{nT}(\hat{c} - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2 \leq 0$ by definition of $B_{nT}(\hat{c}_q - c_0)$. So,

$$\left|\left(B_{nT}(\hat{c}_q - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2 - \left(B_{nT}(\hat{c} - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2\right| = o_p(1). \tag{53}$$

Now, for any given $\delta > 0$, set $\varepsilon = \delta^2$. Then, since $B_{nT}(\hat{c}_q - c_0)$ achieves the minimum of the quadratic function $f(\lambda) = \left(\lambda - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2$ on the closed interval $\{\lambda : B_{nT}(\bar{c} - c_0) \leq \lambda \leq -B_{nT}c_0\}$, it follows that $|B_{nT}(\hat{c} - c_0) - B_{nT}(\hat{c}_q - c_0)| > \delta$ implies

$$\left|\left(B_{nT}(\hat{c}_q - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2 - \left(B_{nT}(\hat{c} - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2\right| > \varepsilon.$$

Therefore

$$\begin{aligned} & P\{|B_{nT}(\hat{c} - c_0) - B_{nT}(\hat{c}_q - c_0)| > \delta\} \\ \leq\; & P\left\{\left|\left(B_{nT}(\hat{c}_q - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2 - \left(B_{nT}(\hat{c} - c_0) - \frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right)^2\right| > \varepsilon\right\} \\ \to\; & 0, \end{aligned}$$

where the last convergence holds by (53), and we have completed the proof of Part (a).

Part (b). Recall that $c_0 \in \mathbb{C}_0 / \{0\}$. For any $\delta > 0$,

$$\begin{aligned} & P\left\{\left|B_{nT}(\hat{c}_q - c_0) - \hat{\phi}_{nT}\right| > \delta\right\} \\ \leq\; & P\left\{\frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} < B_{nT}(\underline{c} - c_0)\right\} + P\left\{\frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} > -B_{nT}c_0\right\}. \end{aligned} \tag{54}$$

Since $\frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} = O_p(1)$, for given $\varepsilon > 0$, we can choose $\bar{K}$ and (n_0, T_0) such that

$$P\left\{\left|\frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right| > \bar{K}\right\} < \varepsilon \text{ for all } n \geq n_0 \text{ and } T \geq T_0.$$

Choose $n_1 = \max\left\{\left(\frac{\bar{K}}{c_0 - \underline{c}}\right)^2, \left(\frac{\bar{K}}{c_0}\right)^2, n_0\right\}$. Recall that $B_{nT} = \sqrt{n}$. Notice by definition that $n \geq n_1$ implies that $-B_{nT}c_0 \geq \bar{K}$ and $B_{nT}(\underline{c} - c_0) \leq -\bar{K}$. So, whenever $n \geq n_1$ and $T \geq T_0$,

$$\begin{aligned} & P\left\{\frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} < B_{nT}(\underline{c} - c_0)\right\} + P\left\{\frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}} > -B_{nT}c_0\right\} \\ \leq\; & 2P\left\{\left|\frac{B_{nT}\mathcal{S}_{nT}}{\mathcal{H}_{nT}}\right| > \bar{K}\right\} \leq 2\varepsilon. \end{aligned} \tag{55}$$

In view of (54) and (55), for any given $\delta, \varepsilon > 0$,

$$P\left\{\left|B_{nT}(\hat{c}_q - c_0) - \hat{\phi}_{nT}\right| > \delta\right\} \leq 2\varepsilon$$

if $n \geq n_1$ and $T \geq T_0$, as required. ■

6.4 Appendix C: Proofs of Section 5

Proof of Lemma 7
Part (a).
Using (37) with $c_0 = 0$, one may write

$$
\begin{aligned}
&\sqrt{n} M_{1nT}(0) \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} x_{it-1} - \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \varepsilon_{it} x_{is-1} \tilde{h}_{1T}(t,s) + \sigma^2 \omega_{1T}(0) \right] + \sqrt{n} \left(\hat{\sigma}^2 - \sigma^2 \right) \omega_{1T}(0) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_{i0} \left[\left(\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} \right) - \left(\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} \frac{1}{T} \sum_{s=1}^{T} \tilde{h}_{1T}(t,s) \right) \right].
\end{aligned}
$$

By Lemma 2,

$$
\begin{aligned}
\sqrt{n} \left(\hat{\sigma}^2 - \sigma^2 \right) \omega_{1T}(0) &= O_p\left(\frac{\sqrt{n}}{T} \right) + O_p\left(\frac{1}{\sqrt{T}} \right) \\
&= o_p(1),
\end{aligned}
$$

where the last equality holds under Assumption 5. Also, using similar arguments that yield (51), we may have

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_{i0} \left[\left(\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} \right) - \left(\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} \frac{1}{T} \sum_{s=1}^{T} \tilde{h}_{1T}(t,s) \right) \right] = O_p\left(\frac{1}{\sqrt{T}} \right).
$$

Since $(n, T \to \infty)$ with $\frac{\sqrt{n}}{T} \to 0$ under Assumption 5,

$$
\sqrt{n} M_{1nT}(0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} x_{it-1} - \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \varepsilon_{it} x_{is-1} \tilde{h}_{1T}(t,s) + \sigma^2 \omega_{1T}(0) \right] + o_p(1).
$$

The required result follows by the limit of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Q_{1iT} - Q_{2iT})$ in Lemma 13 with $c_0 = 0$ and $p = 1$. ■

Part (b).
By Lemma 2, we may have

$$
\begin{aligned}
&\sqrt{n} d M_{1nT}(0) \\
= \; & -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\begin{array}{c} \frac{1}{T^2} \sum_{t=1}^{T} y_{it-1}^2 - \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} y_{it-1} y_{is-1} \tilde{h}_{1T}(t,s) \\ -\sigma^2 \left(\frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \left(\frac{t-s-1}{T} \right) \tilde{h}_{1T}(t,s) \right) \end{array} \right] \\
& + O_p\left(\frac{\sqrt{n}}{T} \right) + O_p\left(\frac{1}{\sqrt{T}} \right).
\end{aligned} \tag{56}
$$

Using (37) with $c_0 = 0$, *i.e.*, $y_{it-1} = x_{it-1} + y_{i0}$, we write

$$
\begin{aligned}
& \left[\begin{array}{c} \frac{1}{T^2} \sum_{t=1}^{T} y_{it-1}^2 - \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} y_{it-1} y_{is-1} \tilde{h}_{1T}(t,s) \\ -\sigma^2 \left(\frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \left(\frac{t-s-1}{T} \right) \tilde{h}_{1T}(t,s) \right) \end{array} \right] \\
= \; & Q_{6iT} + R_{6iT},
\end{aligned}
$$

where

$$Q_{6iT} = \frac{1}{T^2}\sum_{t=1}^{T} x_{it-1}^2 - \frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T} x_{it-1}x_{is-1}\tilde{h}_{1T}(t,s) - \sigma^2\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left(\frac{t-s-1}{T}\right)\tilde{h}_{1T}(t,s)$$

and

$$\begin{aligned} R_{6iT} &= \frac{2y_{i0}}{\sqrt{T}}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} x_{it-1}\right) + \left(\frac{y_{i0}}{\sqrt{T}}\right)^2 - \frac{2y_{i0}}{\sqrt{T}}\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} x_{it-1}\left(\frac{1}{T}\sum_{s=1}^{T}\tilde{h}_{1T}(t,s)\right)\right) \\ &\quad - \left(\frac{y_{i0}}{\sqrt{T}}\right)^2\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{T}\sum_{s=1}^{T}\tilde{h}_{1T}(t,s)\right) \\ &= R_{61iT} + R_{62iT} + R_{63iT} + R_{64iT}, \text{ say.} \end{aligned}$$

Notice that Q_{6iT} is mean zero and independent across i with finite asymptotic variance $Var(Q_{6iT}) \to \sigma^4\frac{11}{6300}$. Then, by Theorem 8 with $C_i = 1$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q_{6iT} \Rightarrow N\left(0, \sigma^4\frac{11}{6300}\right). \tag{57}$$

Also, using similar arguments that yield (51), we can show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_{61iT},\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_{63iT} = O_p\left(\frac{1}{\sqrt{T}}\right) \tag{58}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_{62iT},\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_{63iT} = O_p\left(\frac{\sqrt{n}}{T}\right). \tag{59}$$

Then, in view of $(56) - (59)$, we deduce that

$$\sqrt{n}dM_{1nT}(0) \Rightarrow N\left(0, \sigma^4\frac{11}{6300}\right) \tag{60}$$

as $(n, T \to \infty)$ following Assumption 5. ■

Part (c).

Notice that

$$\begin{aligned} \sqrt{n}\left(d^2 M_{1nT}(0)\right) &= \hat{\sigma}^2 d^2\omega_{1T}(0) \\ &= \hat{\sigma}^2\left(\frac{1}{T^2}\sum_{t=3}^{T}\sum_{s=1}^{t-2}\left(\frac{t-s-1}{T}\right)\left(\frac{t-s-2}{T}\right)\tilde{h}_{1T}(t,s)\right). \end{aligned}$$

From

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

we have

$$\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left(\frac{t-s-1}{T}\right)\left(\frac{t-s-2}{T}\right)\tilde{h}_{1T}(t,s) \to \int_0^1\int_0^r (r-s)^2\,\tilde{h}_1(r,s)\,dsdr + \frac{1}{T}O(1).$$

Also, a direct calculation shows that

$$\int_0^1 \int_0^r (r-s)^2 \tilde{h}_1(r,s)\, dsdr = 0.$$

Therefore, since $\hat{\sigma}^2 \to_p \sigma^2$ by Lemma 2 and $\frac{\sqrt{n}}{T} \to 0$ under Assumption 5, we have

$$\sqrt{n}\left(d^2 M_{1nT}(0)\right) = O_p\left(\frac{\sqrt{n}}{T}\right) = o_p(1),$$

as required. ■

Part (d).

By definition,

$$d^3 M_{1nT}(c) = \hat{\sigma}^2 d^3 \omega_{1T}(c),$$

where

$$d^3 \omega_{1T}(c) = \frac{1}{T^2} \sum_{t=4}^{T} \sum_{s=1}^{t-3} \left(1+\frac{c}{T}\right)^{t-s-4} \left(\frac{t-s-1}{T}\right)\left(\frac{t-s-2}{T}\right)\left(\frac{t-s-3}{T}\right) \tilde{h}_{1T}(t,s).$$

Then, since $\hat{\sigma}^2 \to_p \sigma^2$ and by Lemma 11,

$$\begin{aligned} d^3 M_{1nT}(c) &\to_p \sigma^2 d^3 M_1(c,0) \\ &= \sigma^2 \int_0^1 \int_0^r e^{c(r-s)} (r-s)^3 \tilde{h}_1(r,s)\, dsdr \end{aligned}$$

uniformly in $c \in \mathbb{C}$. Also, a direct calculation shows that

$$d^3 M_1(0,0) = \int_0^1 \int_0^r (r-s)^3 \tilde{h}_1(r,s)\, dsdr = -\frac{1}{70},$$

and we have the required result. ■

Proof of Lemma 8
Part (a).

By definition, we can write

$$\begin{aligned} M_{2nT}(0) = {} & \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}y_{it-1}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y_{it-1}\right) \\ & -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{t-1}{T}\varepsilon_{it}\right) \\ & +\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right)^2\left(\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\right) + \hat{\sigma}^2\lambda_{1T}(0). \end{aligned}$$

Noticing that

$$\frac{1}{T}\sum_{t=1}^{T} y_{it-1}\varepsilon_{it} = \frac{1}{2T}\left(y_{iT}^2 - y_{i0}^2\right) - \frac{1}{2T}\sum_{t=1}^{T}\varepsilon_{it}^2, \tag{61}$$

$$\frac{1}{\sqrt{T}}\sum_{t-1}^{T}\frac{t-1}{T}\varepsilon_{it}=\frac{T-1}{T}\frac{y_{iT}}{\sqrt{T}}-\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y_{it-1}+\frac{1}{T\sqrt{T}}y_{i0}, \tag{62}$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}=\frac{y_{iT}}{\sqrt{T}}-\frac{y_{i0}}{\sqrt{T}}, \tag{63}$$

one may rearrange

$$\begin{aligned}
& M_{2nT}(0) \\
= \ & \frac{1}{2n}\sum_{i=1}^{n}\left(\frac{y_{iT}}{\sqrt{T}}\right)^2-\frac{1}{2nT}\sum_{i=1}^{n}y_{i0}^2-\frac{1}{2nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}^2 \\
& -\left(\frac{T-1}{T}\right)\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{iT}}{\sqrt{T}}\right)^2+\left(\frac{T-1}{T}\right)\frac{1}{n\sqrt{T}}\sum_{i=1}^{n}\left(\frac{y_{iT}}{\sqrt{T}}\right)y_{i0} \\
& -\frac{1}{nT\sqrt{T}}\sum_{i=1}^{n}\left(\frac{y_{iT}}{\sqrt{T}}\right)y_{i0}+\frac{1}{nT^2}\sum_{i=1}^{n}y_{i0}^2 \\
& +\left(\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\right)\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{iT}}{\sqrt{T}}\right)^2-2\left(\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\right)\frac{1}{n\sqrt{T}}\sum_{i=1}^{n}\left(\frac{y_{iT}}{\sqrt{T}}\right)y_{i0} \\
& +\left(\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\right)\frac{1}{nT}\sum_{i=1}^{n}y_{i0}^2+\hat{\sigma}^2\lambda_{1T}(0) \\
= \ & I_1+...+I_{12}, \text{ say.}
\end{aligned}$$

Using Assumption 2 and the results in Lemma 12 and modifying its proof, it is possible to show that

$$\begin{aligned}
I_1+I_4+I_8 &= -\frac{1}{T}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{iT}}{\sqrt{T}}\right)^2\right)=O_p\left(\frac{1}{T}\right), \\
I_2, I_{10} &= O_p\left(\frac{1}{T}\right), \qquad I_5+I_9=0, \\
I_6 &= O_p\left(\frac{1}{T\sqrt{T}}\right), \qquad I_7=O_p\left(\frac{1}{T^2}\right).
\end{aligned}$$

Thus,

$$\begin{aligned}
& \sqrt{n}M_{2nT}(0) \\
= \ & \sqrt{n}\left(-\frac{1}{2nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}^2+\hat{\sigma}^2\lambda_{1T}(0)\right)=\sqrt{n}\left(-\frac{1}{2}\tilde{\sigma}^2+\hat{\sigma}^2\lambda_{1T}(0)\right) \\
= \ & -\frac{\sqrt{n}}{2}\left(\tilde{\sigma}^2-\hat{\sigma}^2\right)+\hat{\sigma}^2\sqrt{n}\left(\lambda_{1T}(0)-\frac{1}{2}\right).
\end{aligned}$$

By Lemma 2, $-\frac{\sqrt{n}}{2}\left(\tilde{\sigma}^2-\hat{\sigma}^2\right)=O_p\left(\frac{\sqrt{n}}{T}\right)$. Also, $\sqrt{n}\left(\lambda_{1T}(0)-\frac{1}{2}\right)=O\left(\frac{\sqrt{n}}{T}\right)$. Therefore,

$$\sqrt{n}M_{2nT}(0)=O_p\left(\frac{\sqrt{n}}{T}\right)=o_p(1)$$

as $(n, T \to \infty)$ following Assumption 5. ■

Next, we sketch proofs for Parts (b) – (d). The details of the proofs for Part (b), (c), and (d) are similar to those of Part (b) of Lemma 7, Part (a) above, and Lemma 3, respectively, and we omit the details.[10]

Parts (b)-(d)

Take the first derivative of $M_{2nT}(c)$ with respect to the parameter c and evaluating it at $c = 0$ with $c_0 = 0$, apply Lemma 2, and use the relations of (62) and (63). Then, one may find that

$$\begin{aligned}\sqrt{n} dM_{2nT}(0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\left(\begin{array}{c} -\frac{1}{T^2} \sum_{t=1}^{T} x_{it-1}^2 + \sigma^2 \frac{1}{T} \sum_{t=1}^{T} \frac{t-1}{T} \\ +2 \frac{x_{iT}}{\sqrt{T}} \left(\frac{1}{T\sqrt{T}} \sum_{t=1}^{T} \frac{t-1}{T} x_{it-1} \right) - 2\sigma^2 \frac{1}{T} \sum_{t=1}^{T} \left(\frac{t-1}{T} \right)^2 \\ -\frac{1}{3} \left(\frac{x_{iT}}{\sqrt{T}} \right)^2 + \frac{1}{3}\sigma^2 \end{array} \right) \right] + O_p\left(\frac{\sqrt{n}}{T} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q_{7iT} + o_p(1), \text{ say,}\end{aligned}$$

where x_{it} is defined in (36) and the $o_p(1)$ term holds since $(n, T \to \infty)$ with $\frac{\sqrt{n}}{T} \to 0$ under Assumption 5. Direct calculations show that $EQ_{7iT} = 0$ and $Var(Q_{7iT}) \to \frac{\sigma^4}{45}$. By applying Theorem 8 with $C_i = 1$, then one may derive

$$\sqrt{n} dM_{2nT}(0) \Rightarrow N\left(0, \frac{\sigma^4}{45}\right), \tag{64}$$

as required.

The proof of Part (c) is similar to that of Part (b). Taking the second order derivative of $M_{2nT}(c)$ with respect to the parameter c with $c_0 = 0$, considering Lemma 2, and rearranging terms using the relations of (61) and (62), it is possible to show that as $(n, T \to \infty)$ following Assumption 5,

$$\sqrt{n} d^2 M_{2nT}(0) = O_p\left(\frac{\sqrt{n}}{T}\right) = o_p(1).$$

The proof of Part (d) is similar to the proof of Lemma 3. After taking the third order derivative of $M_{2nT}(c)$ with respect to c and using the results in Lemma 12, it is possible to show the required result. ■

Proof of Theorem 4

Define $\hat{\theta}_{nT} = n^{1/6}\hat{c}$. First, we consider the case where $\left\{ \left| \hat{\theta}_{nT} \right| > 1 \right\}$. By the definition of the GMM estimator, we have

$$\begin{aligned} o_p(1) &\leq n\left(Z_{nT}(0) - Z_{nT}(\hat{c})\right) \\ &= -\sum_{k=1}^{6} \left(n^{(1-k/6)} \mathcal{A}_{k,nT} \right) \hat{\theta}_{nT}^k - \sum_{k=3}^{6} \hat{\theta}_{nT}^k \left(n^{(1-k/6)} \mathcal{N}_{k,nT}(\hat{c}, 0) \right). \end{aligned}$$

In view of (25) – (32) and from Assumption 6, $\hat{\theta}_{nT}$ satisfies

$$o_p(1) \leq -\left|\hat{\theta}_{nT}\right|^6 + \left|\hat{\theta}_{nT}\right|^5 o_p(1) + \left|\hat{\theta}_{nT}\right|^4 o_p(1) + 2O_p(1)\left|\hat{\theta}_{nT}\right|^3 + \left|\hat{\theta}_{nT}\right|^2 o_p(1) + \left|\hat{\theta}_{nT}\right| o_p(1). \tag{65}$$

[10] One can obtain detailed derivations of $dM_{2nT}(0)$, $d^2M_{2nT}(0)$ from the first author upon request.

Since, $\left|\hat{\theta}_{nT}\right| > 1,$

$$\begin{aligned} & \text{The right hand side of (65)} \\ \leq & -\left|\hat{\theta}_{nT}\right|^6 (1 + o_p(1)) + 2O_p(1)\left|\hat{\theta}_{nT}\right|^3. \end{aligned}$$

Then,

$$\left|\hat{\theta}_{nT}\right|^6 \leq 2O_p(1)\left|\hat{\theta}_{nT}\right|^3 + o_p(1).$$

Following relation (7.4) in Andrews (1999), page 1377, we can deduce that

$$\left|\hat{\theta}_{nT}\right|^3 \leq O_p(1) + o_p(1).$$

Therefore, when $\left\{\left|\hat{\theta}_{nT}\right| > 1\right\}$,

$$\left|\hat{\theta}_{nT}\right| \leq O_p(1). \tag{66}$$

Finally, let the $O_p(1)$ random variable in (66) be ξ_{nT}. Then,

$$\begin{aligned} \left|\hat{\theta}_{nT}\right| &= \left|\hat{\theta}_{nT}\right| 1\left\{\left|\hat{\theta}_{nT}\right| \leq 1\right\} + \left|\hat{\theta}_{nT}\right| 1\left\{\left|\hat{\theta}_{nT}\right| > 1\right\} \\ &\leq \left|\hat{\theta}_{nT}\right| 1\left\{\left|\hat{\theta}_{nT}\right| \leq 1\right\} + \xi_{nT} \\ &\leq 1 + \xi_{nT} = O_p(1). \blacksquare \end{aligned}$$

Proof of Theorem 5

The proof of the theorem is similar to that of Theorem 3 and is omitted. ■

6.5 Appendix D: Numerical Validation of the Identification Condition of $m(c)$ [11]

This section provides a numerical confirmation that the uniform limit of the moment condition function $m(c) = (m_1(c), m_2(c))'$has a zero only at the true parameter $c = c_0$.We restrict the parameter set to $\mathbb{C} = [-10, 0]$in this numerical exercise. The choice of the lower limit $\bar{c} = -10$is made for computational convenience, and the results hold for all finite values of $\bar{c} < 0$. All the numerical analysis in this section is done with Mathematica and with Maple using Scientific Workplace Version 3.0.

6.5.1 When $g_{1t} = t$

The procedure we apply is to find all the roots of $m_2(c) = 0$ and verify whether these roots are also the roots of $m_1(c) = 0$. We first notice that for given c_0, the function $m_2(c)$ is simply the ratio of two polynomials - the denominator and the numerator of $m_2(c)$,

[11] We are indebted to John Owens for the numerical analysis in this section.

say $m_{d2}(c)$ and $m_{n2}(c)$, respectively, are a fourth degree polynomial and a fifth degree polynomial in c, respectively.[12]

Case A: When $c_0 \neq 0$
Step 1: Numerical Calculation of the roots of $m_2(c) = 0$.

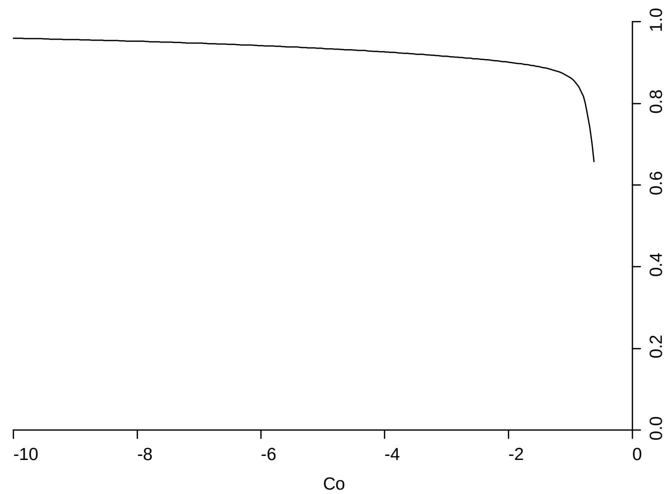

Fig. A.1. Graph of Roots of $\tilde{m}_{n2}(c)$

[12] This is verified easily by noticing that $\dot{g}_{pc}(r)$, $A_p(c)$, $B_p(c)$ are polynomials of c, except for the last term in $m_2(c)$. However, a direct calculation shows that the last term is also a ratio of two polynomials, viz.,

$$
\begin{aligned}
& \int_0^1 \int_0^r e^{c(r-s)} \dot{g}_{pc}(s)' A_p(c)^{-1} \dot{g}_{pc}(r)\, ds dr \\
= & \int_0^1 \int_0^r e^{c(r-s)} (1-cs) \left(1 - c + \frac{1}{3}c^2\right)^{-1} (1-cr)\, ds dr \\
= & -\frac{1}{2} \frac{2c-3}{3-3c+c^2}.
\end{aligned}
$$

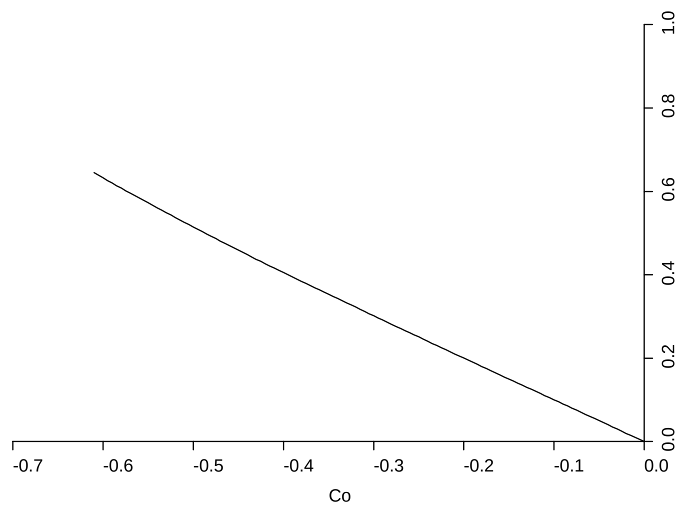

Fig. A.2. Graph of Roots of $\tilde{m}_{n2}(c)$

By a direct calculation, we find that the denominator of $m_2(c)$, $m_{d2}(c)$, equals $4c_0^5\left(c^2-3c+3\right)^2$ when $c_0 \neq 0$. Since $c^2-3c+3=\left(c-\frac{3}{2}\right)^2+\frac{3}{4}>0$, the denominator of $m_2(c)$ has no real zeros for all $c_0 \neq 0$. Thus, if we concerned with the zeros of $m_2(c)$, it suffices to consider only the numerator of $m_2(c)$, $m_{n2}(c)$. By definition of $m_2(c)$, we find that the true value $c=c_0$ is always a zero of $m_{n2}(c)$. Also, by inspection, we find that $c=0$ is always a zero of $m_{n2}(c)$. Thus, we can write

$$m_{n2}(c)=c\,(c-c_0)\,\tilde{m}_{n2}(c),$$

where $\tilde{m}_{n2}(c)$ is a third degree polynomial. Using Mathematica, we solve the third degree polynomial $\tilde{m}_{n2}(c)$ and find three roots of $\tilde{m}_{n2}(c)$ as a function of the true parameter c_0. For the numerical calculation we choose $\bar{c}=-10$, and so we assume that the parameter set $\mathbb{C}=[-10,0]$. Figs. A.1-A.2 plot the graphs of these roots on $\mathbb{C}$ only when the roots are real numbers (Fig. A2 shows a the graph on a finer scale to the left of the origin). As we see from the graphs, for $c_0<0$, the roots of $\tilde{m}_{n2}(c)$ are all positive, and so $\tilde{m}_{n2}(c)$ does not have a root in the parameter set $\mathbb{C}$.

Step 2: Plug the root $\mathbf{c=0}$ of $m_2(c)$ in $m_1(c)$

We now investigate, for given $c_0 \in \mathbb{C}/\{0\}$, whether $m_1(c)=0$ when $c=0$. By matching the given true parameter c_0 with $m_1(0)$, we can define the function $m_1\_0(c_0)$ of c_0. Using Maple, we calculate

$$m_1\_0(c_0)=\frac{1}{4c^4}\begin{pmatrix} -c_0^3+48e^{c_0}-8e^{c_0}c_0^2-8c_0^2-24 \\ +c_0^3e^{2c_0}-8e^{2c_0}c_0^2+24ce^{2c_0}-24e^{2c_0}-24c_0 \end{pmatrix},$$

and plot the graph of $m_1\_0(c_0)$. Fig. A.3 plots $m_1\_0(c_0)$ over the domain $c_0 \in [-10,0.4]$ and Fig. A.4 plots the same function on the domain $c_0 \in [0.4,0]$. Through these graphs, we can confirm that $m_1\_0(c_0)$ is positive but very close to zero when the true value c_0 is close to zero.

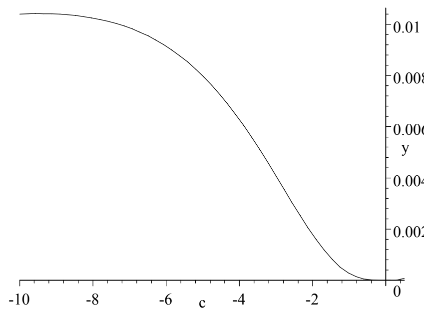

Fig. A.3 Graph of $m_1\_0\,(c_0)$

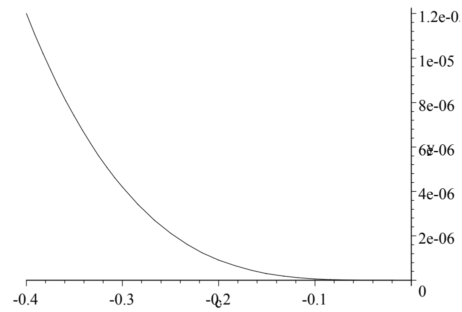

Fig. A.4 Graph of $m_1\_0\,(c_0)$

To investigate further the behavior of $m_1\_0\,(c_0)$ around $c_0 = 0$, in Fig. A.5 we plot the graph of the first derivative of the numerator of $m_1\_0\,(c_0)$ over $c_0 \in [-0.05, 0]$.

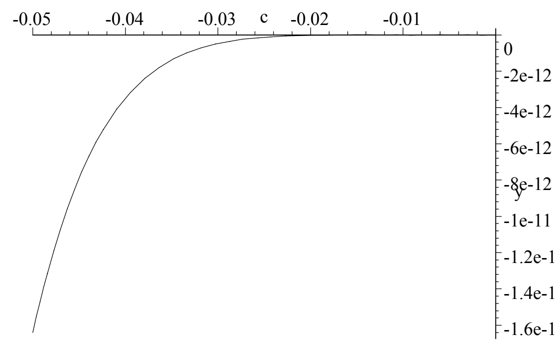

Fig. A.5. Graph of the first derivative of the Numerator of $m_1\_0\,(c_0)$

The graph shows that the first derivative of the numerator of $m_1\_0\,(c_0)$ is negative around zero, and so $m_1\_0\,(c_0)$ is strictly decreasing. Therefore, we conclude that $m_1\_0\,(c_0)$ is not zero for all $c_0 \in \mathbb{C}_0$.

Case B: When $c_0 = 0$.

Using Maple, we calculate $m_2\,(c)$ when $c_0 = 0$, and plot the graph in Figs. A.6 and A.7. These figures confirm that $m_2\,(c) = 0$ only when $c = c_0 = 0$.

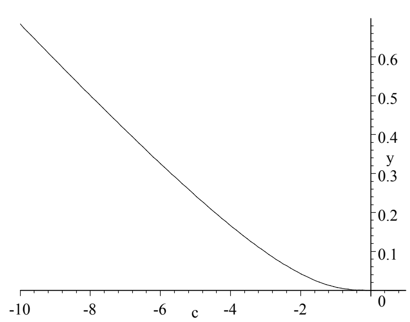

Fig. A.6 Graph of $m_2\,(c)$ when $c_0 = 0$

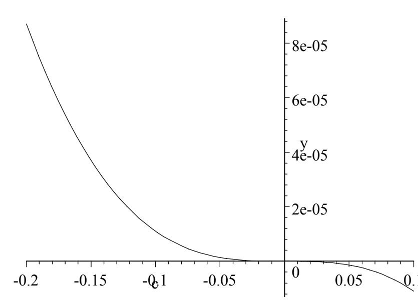

Fig. A.7 Graph of $m_2\,(c)$ when $c_0 = 0$

6.5.2 When $g_{2t} = (t, t^2)$

Although the expressions involved in $m_2(c)$ in this case are far more complex, the analysis is simpler. Like the case of $g_{1t} = t$, we find that the denominator of $m_2(c)$ does not change sign over $\mathbb{C} = [-10, 0]$, and so we focus on the numerator of $m_2(c)$. Similar to the case of $g_{1t} = t$, we numerically calculate the real roots of the numerator of $m_2(c)$ for $c_0 \in \mathbb{C} = [-10, 0]$, and we find that there exists only one root in the range of c_0, which implies that $m_2(c) = 0$ only at the true c_0. Therefore, when $g_{2t} = (t, t^2)$, the limit of the moment condition $m(c)$ identifies the true parameter c_0 in $\mathbb{C}$.

References

[1] Andews, D.W.K. (1999) : Estimation When a Parameter Is on a Boundary, *Econometrica*, 67, 1341–1384.

[2] Arellano, M. and B. Honoré (2000) : Panel Data Models: Some Recent Developments, Mimeo.

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

[4] Bhattacharyya, A. (1946) : On Some Analogues of the Amount Information and Their Uses in Statistical Estimation, *Sankhya*, 8, 1–14, 201-218, 315-328.

[5] Bernard, A. and C. Jones (1996) : Productivity Across Industries and Countries: Time Series Theory and Evidence, *Review of Economics and Statistics,* 78, 135–146.

[6] Binder, M., C. Hsiao, and H. Pesaran (1999) : Likelihood Based Inference for Panel Vector Autoregressions: Testing for Unit Roots and Cointegration in Short Panels, *Mimeo.*

[7] Canjels, E. and M. Watson (1997) : Estimating Deterministic Trends in Presence of Serially Correlated Errors, *Review of Economics and Statistics*, 79, 184–200.

[8] Choi, I. (1999) : Unit Root Tests for Panel Data, *forthcoming in Journal of International Money and Finance.*

[9] Coakely, J., F. Kulasi, and R. Smith (1996) : Current Account Solvencyand the Feldstein-Horika Puzzle, *Economic Journal*, 106, 620–627.

[10] Davidson, J. (1994) : *Stochastic Limit Theory*, Oxford University Press.

[11] Evans, D. (1987) : Tests of Alternative Theories of Firm Growth, *Journal of Political Economy*, 95, 657–674.

[12] Gibrat, R. (1931) : *Les Inegalities Economique,* Paris, Sirey.

[13] Hahn, J. (1998) : Asymptotically Unbiased Inference of Dynamic Panel Model with Fixed Effects When Both n and T are Large, *Mimeo.*

[14] Hahn, J. and G. Kuersteiner (2000) : Asymptotically Unbiased Inference of Dynamic Panel Model with Fixed Effects When Both n and T are Large, (revision of Hahn, 1998), *Mimeo.*

[15] Hahn, J., J. Hausman, and G. Kuersteiner (2001) : Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects, Mimeo.

[16] Hall, B. and J. Mairesse (2000) : Univariate Panel Data Models and GMM Estimators: An Exploration Using Real and Simulated Data, *Mimeo.*

[17] Im, K., H. Pesaran, and Y. Shin (1996) : Testing for Unit Roots Heterogeneous Panels, *Mimeo.*

[18] Jovanovic, B. (1982) : Selection and the Evolution of Industry, *Econometrica*, 50, 649–670.

[19] Kruiniger, H. (2000). : GMM Estimation of Dynamic Panel Data Models with Persistent Data, *Mimeo.*

[20] Lancaster, T. (2000) : The Incidental Parameter Problem Since 1948, *Journal of Econometrics*, 95, 391–414.

[21] Levin, A. and C. Lin (1993) : Unit Root Tests in Panel Data: New Results, UCSD Working Paper.

[22] Lucas, R. E., (1978) : On the Size Distribution of Business Firms, *Bell Journal of Economics*, 9, 508–523.

[23] MacDonald, R. (1996) : Panel Unit Root Tests and Real Exchange Rates, *Economics Letters*, 50, 7–11.

[24] Maddala, G.S. and S. Wu (1999) : A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics, 61, 631–652.*

[25] Magnus, J. and H. Neudecker (1988) : *Matrix Differential Calculus*, New York, Wiley.

[26] McCloughan, P. (1995) : Simulation of Concentration Development From Modified Gibrat Growth-Entry-Exit Processes, *Journal of Industrial Economics*, 43, 405–433.

[27] Moon, H.R. and P.C.B. Phillips (1998) : A Reinterpretation of the Feldstein-Horioka Regressions from a Nonstationary Viewpoint, *Mimeo.*

[28] Moon, H.R. and P.C.B. Phillips (1999) : Maximum Likelihood Estimation in Panels with Incidental Trends, *Oxford Bulletin of Economics and Statistics, 61, 771–748.*

[29] Moon, H.R. and P.C.B. Phillips (2000) : Estimation of Autoregressive Roots near Unity using Panel Data, *Econometric Theory*, 16, 927–997.

[30] Moon, H.R. and P.C.B. Phillips (2000) : GMM Estimation of Autoregressive Roots Near Unity with Panel Data, *Mimeo.*

[31] Newey, W. and D. McFadden (1994) : Large Sample Estimation and Hypothesis Testing, in Engle R.F. and D. McFadden ed. *Handbook of Econometrics* Vol 4., North-Holland, Amsterdam, 2111-2245.

[32] Neyman, J. and E.L. Scott (1948) : Consistent Estimates Based on Partially Consistent Observations, *Econometrica*, 16, 1–32.

[33] Nickell, S. (1981) : Biases in Dynamic Models with Fixed Effects, *Econometrica*, 49, 1417–1426.

[34] Oh, K. Y. (1996) : Purchasing Power Parity and the Unit Root Tests Using Panel Data, *Journal of International Money and Finance*, 15, 405–418.

[35] Park, J. and P.C.B. Phillips (1988) : Statistical Infencence in Regressions with Integrated Processes, Part I, *Econometric Theory*, 4, 468–497.

[36] Pedroni, P. (1996) : Fully Modified OLS for Heterogeneous Cointegrated Panels and the Case of Purchasing Power Parity, Indiana University Working Papers in Economics, No. 96-20.

[37] Pedroni, P. (1999) : Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors, *Oxford Bulletin of Economics and Statistics, 61, 653–670.*

[38] Pesaran, H. and R. Smith (1995) : Estimating Long-Run Relationships from Dynamic Heterogeneous Panels, *Journal of Econometrics*, 68, 79–113.

[39] Phillips, P. C. B. (1987) : Towards a Unified Asymptotic Theory for Autoregression, *Biometrica*, 74, 535-547.

[40] Phillips, P. C. B. (1993) : Hyper-Consistent Estimation of a Unit Root in Time Series Regression. Cowles Foundation Discussion Paper #1040.

[41] Phillips, P. C. B. (1995) : Full Modified Least Squares and Vector Autoregression. *Econometrica,* 63, 1023-1078.

[42] Phillips, P.C.B. and C. Lee (1996) : Efficiency Gains from Quasi-Differencing under Nonstationarity, in P.M. Robinson and M. Rosenblatt (eds.), *Athens Conference of Applied Probability and Time Series: Volume II Time Series Analysis in Memory of E.J. Hannan*, New York, Springer-Verlag.

[43] Phillips, P.C.B. and H.R. Moon (1999) : Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica*, 67, 1057–1111.

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

[45] Prais, S. J. (1976) : *The Evolution of Giant Firms in Britain*, Cambridge, Cambridge University Press.

[46] Quah, D. (1994) : Exploiting Cross-Section Variations for Unit Root Inference in Dynamic Data, *Economic Letters*, 44, 9–19.

[47] Schmalensee, R. (1989) : Inter-industry Studies of Structure and Performance, *Handbook of Industrial Organization*, Vol 2, Edited by R. Schmalensee and R. Willig, North Holland, Amsterdam.

[48] Stock, J. (1994) : Unit Roots, Structure Breaks and Trends, Chapter 46, *Handbook of Econometrics*, Vol 4, Edited by R. Engle and D. McFadden, Elsevier, New York.

[49] Sutton, J. (1997) : Gibrat's Legacy, *Journal of Economic Literature*, 35, 40–59.

[50] Waterman, R. and B. Lindsay (1996) : Projected Score Methods for Approximating Conditional Scores, *Biometrika*, 83, 1–13.

[51] Waterman, R. and B. Lindsay (1998) : Projected Score Methods for Nuisance Parameters: Asymptotics and Neyman-Scott Problems, *Mimeo.*

[52] Wooldridge, J. (2001) : *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.

[53] Wu, S. (1997) : Purching Power Parity Under the Currency Float: New Evidence from Panel Data Unit Root Tests, *Mimeo*.

[54] Wu, Y. (1996) : Are Real Exchange Rates Nonstationary? Evidence from A Panel Data Test, *Journal of Money, Credit, and Banking*, 28, 54–63.