Maximum Likelihood Estimation in Panels with Incidental Trends*

Hyungsik R. Moon
Department of Economics
University of California, Santa Barbara
&
Peter C. B. Phillips
Cowles Foundation for Research in Economics
Yale University

16 July, 1999

Abstract

It is shown that the maximum likelihood estimator of a local to unity parameter can be consistently estimated with panel data when the cross section observations are independent. Consistency applies when there are no deterministic trends or when there is a homogeneous deterministic trend in the panel model. When there are heterogeneous deterministic trends the panel MLE of the local to unity parameter is inconsistent. This outcome provides a new instance of inconsistent ML estimation in dynamic panels, and, unlike earlier results of this type, applies when both $T \to \infty$ and $N \to \infty$.

Keywords: Deterministic trends, dynamic panels, incidental parameters, inconsistent maximum likelihood estimator, local to unity, nonstationary panel data.

JEL Classification Numbers: C32 Time Series Models, C33 Panel Data.

1 Introduction

In recent nonstationary time series applications, it has been extremely common to model time series with roots near unity using the device of an autoregressive root that is local to unity. Some early studies of near unit root nonstationary time series include developments of local alternatives to unit root specifications (Bobkoski, 1983; Phillips, 1987), derivations of power functions and power envelopes of unit root tests (*e.g.*, Cavanagh, 1985, Phillips, 1987, Johansen, 1991), and the construction of confidence intervals for the long run autoregressive coefficient (Stock, 1991). More

*Moon thanks the Academic Research Committee of UCSB for research support and Phillips thanks the NSF for research support under Grant No. SBR97-30295.

recent research on near unit root nonstationary time series investigates the efficient extraction of deterministic trends (Phillips and Lee, 1996, and Canjels and Watson, 1997), and the construction of point optimal invariant tests for a unit root (Elliot *et al*, 1996) and cointegrating rank (Xiao and Phillips, 1999). For further examples, readers can refer to recent surveys on unit root processes (e.g., Stock, 1994, and Phillips and Xiao, 1998).

Like other parameters in econometric models, localizing parameters in near integrated processes are not usually observable. But, implementation of some methods in the forementioned studies requires knowledge of the localizing parameter or a consistent estimate of it. For example, it is well known that efficiency gains in the estimation of deterministic trends can be obtained by quasi-differencing the data using the unknown localizing parameter (*e.g.* Phillips and Lee, 1996, and Canjels and Watson, 1997). However, if we implement this procedure using inconsistent estimates of the localizing parameter, then the limit distribution of the resulting trend coefficient estimator is highly nonstandard, which introduces new difficulties, e.g, in constructing confidence intervals for the trend coefficient. Largely because of this problem, Cavanagh *et al* (1995) and Canjels and Watson (1997) suggested the use of Bonferroni-type confidence intervals, which are often very conservative.

Finding a consistent estimate of the localizing parameter is not straightforward. Obvious procedures like the use of least squares are well known to be inconsistent (Phillips, 1987); and, even in the simplest framework, consistent estimation inevitably involves the introduction of additional information. In view of its potential applications in both estimation and inference, the problem of consistent estimation of the localizing parameter in local to unity models poses an interesting problem with important implications. Two recent studies that consider the subject are Moon and Phillips (1998) and Phillips *et al* (1998).

The main purpose of this paper is to investigate the asymptotic properties of the Gaussian maximum likelihood estimators (MLE) of the localizing parameter in local to unity dynamic panel regression models. The model we consider here allows for the panel to be generated with deterministic and stochastic trends, and a common localizing parameter is assumed to apply across individuals. Commonality of the localizing parameter is restrictive, but is no more restrictive than the conventional assumption of common AR parameters in stationary dynamic panels (e.g., Nickell, 1981). Two different models are considered: a homogeneous trend model in which the deterministic trends are homogenous across the individuals in the panel; and a heterogeneous trend model where the deterministic trends may vary across individuals, much like fixed individual effects. In the homogeneous trend model we show that the Gaussian MLE of the common localizing parameter is $\sqrt{N}-$ consistent and has a limiting normal distribution that is the same as that in the case where the trends are known. In the heterogeneous trends model it is shown that the Gaussian MLE of the localizing parameter is inconsistent.

The inconsistency of the MLE of the localizing parameter in the heterogeneous trend model is an instance of the so-called incidental parameter problem originally explored by Neyman and Scott (1948). In this model, the heterogenous trend coefficients correspond to incidental parameters whose number goes to infinity as the cross section dimension $N \to \infty$. Such problems frequently appear in panel data models with fixed effects, a well known example being the dynamic panel regression model with fixed effects. In this case, the MLE of the lagged dependent variable coefficient that is common over individuals is inconsistent if $N \to \infty$ while the sample size dimension, T, is fixed (Nickell, 1981). In most panel data situations this incidental parameter problem disappears when T passes to infinity also (*e.g.*, Alvarez and Arellano, 1998, and Hahn, 1998). A particularly interesting aspect of the incidental parameter problem discovered in this paper is that the inconsistency of the MLE of the localizing parameter does not disappear even when both N and T tend to infinity.

The paper is organized as follows. Section 2 lays out the model and assumptions, and shows that when the deterministic components are known, the Gaussian MLE of the localizing parameter is consistent. Section 3 studies asymptotic properties of the Gaussian MLE of the panel regression model with unknown deterministic trends. Section 4 reports some Monte-Carlo simulations that investigate the magnitude of the inconsistency. Section 5 concludes and offers some suggestions for dealing with the inconsistency. Proofs and technical derivations are collected in the Appendix in Section 6.

Our notation is mostly standard. We use "$\to_p$ " and "$\Rightarrow$ " to denote convergence in probability and convergence in distribution, respectively. The notation $(N, T \to \infty)$ implies that N and T tend to infinity together, while $(N, T \to \infty)_{\text{seq}}$ means that the indices pass to infinity sequentially (first T and then N). Standard Browian Motion is denoted by $W(r)$.

2 Near Integrated Panels - Preliminary Theory

We start by introducing a panel regression model where data $z_{i,t}$ are generated by deterministic trends $G_i(t)$ and near integrated stochastic trends $y_{i,t}$ as follows:

$$\begin{aligned} z_{i,t} &= G_i(t) + y_{i,t}, \qquad t = 1, ..., T; i = 1, ..., N, \\ y_{i,t} &= a y_{i,t-1} + \varepsilon_{i,t}, \qquad a = \exp(\frac{c}{T}) \sim \left(1 + \frac{c}{T}\right). \end{aligned} \tag{1}$$

The parameter c in (1) is a local to unity parameter that is common to all individuals in the panel. The main purpose of this paper is to investigate asymptotic properties of the MLE of the localizing parameter c.

To provide some intuition, we first consider the simple case where $y_{i,t} = z_{i,t} - G_i(t)$ is observable, abstracting from the problem of fitting the deterministic component in (1). Assume that the errors $\varepsilon_{i,t}$ are iid $N(0, \sigma^2)$, and, for simplicity in this section, that σ^2 is known and that the initial observations $y_{i,0} = 0$ for all i. Under these assumptions the standardized log-likelihood function of the panel data $y^{N,T} = (y_{1,1}, ..., y_{N,T})'$ is

$$L_{N,T}\left(y^{N,T}; c\right) = -\frac{1}{2\sigma^2 N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\Delta y_{i,t} - \frac{c}{T} y_{i,t-1}\right)^2 + \text{ constant.} \tag{2}$$

Let c_0 denote the true localizing parameter, and assume that c_0 is an element of the interior of a convex set of $\mathbb{R}$. Define $\varepsilon_{i,t}(c_0) = \Delta y_{i,t} - \frac{c_0}{T} y_{i,t-1}$. Then, the MLE of c is obtained by maximizing the standardized log-likelihood

$$
\begin{aligned}
& L_{N,T}\left(y^{N,T}; c\right) - L_{N,T}\left(y^{N,T}; c_0\right) \\
= \; & -\frac{1}{2\sigma^2 N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\Delta y_{i,t} - \frac{c}{T} y_{i,t-1}\right)^2 + \frac{1}{2\sigma^2 N} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{i,t}(c_0)^2 \\
= \; & -\frac{1}{2}(c - c_0)^2 \frac{1}{\sigma^2 N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} y_{i,t-1}^2 - (c - c_0) \frac{1}{\sigma^2 N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} y_{i,t-1} \varepsilon_{i,t}(c_0),
\end{aligned}
$$

which is quadratic in c.

According to Lemma 6 (c) and (d) in the Appendix, as $(N, T \to \infty)$, we have

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

and

$$
\frac{1}{\sigma^2 N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} y_{i,t-1} \varepsilon_{i,t}(c_0) \to_p 0.
$$

It follows that

$$
L_{N,T}\left(y^{N,T}; c\right) - L_{N,T}\left(y^{N,T}; c_0\right) \to_p -\frac{1}{2}(c - c_0)^2 \left(\int_0^1 \int_0^r e^{2c_0(r-s)} ds dr\right) = l(c, c_0), \text{ say}
$$

for each c, as $(N, T \to \infty)$. Note that the objective function $L_{N,T}\left(y^{N,T}; c\right) - L_{N,T}\left(y^{N,T}; c_0\right)$ is concave in c over $\mathbb{R}$ and the limit function $l(c, c_0)$ has a unique maximum at c_0 and is continuous and concave in c over $\mathbb{R}$. Thus, the MLE $\hat{c}$ is consistent for c_0 by standard theory for extremum estimator (*e.g.*, Theorem 2.7 in Newey and McFadden, 1994).

In this particular case, the MLE has the closed form

$$
\hat{c} = T(\hat{a} - 1),
$$

where

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

Using Lemma 6(a) and Lemma 7(c) in the Appendix, we can show that as $(N, T \to \infty)$

$$
\sqrt{N}(\hat{c} - c_0) \Rightarrow N\left(0, \frac{\sigma^2}{\left(\int_0^1 \int_0^r e^{2c_0(r-s)} ds dr\right)}\right). \tag{3}
$$

Therefore, when $y_{i,t}$ is observable (*i.e.*, when $G_i(t)$ in model (1) is known), the Gaussian MLE $\hat{c}$ of the common localizing parameter c is $\sqrt{N}$−consistent and weakly convergent to the normal distribution (3).

The question to be explored in the present paper is whether these asymptotic properties (particularly, the consistency and asymptotic normality, of the Gaussian MLE of c) continue to hold in panel models with unknown deterministic trends. It is known from Moon and Phillips (1998) that the OLS estimator of c is inconsistent under these circumstances, viz. when the deterministic trends are estimated and eliminated by prior regression.

Before proceeding further, we introduce the following three assumptions which will be maintained throughout the paper.

Assumption 1 ***(Error Normality)*** *The $\varepsilon_{i,t}$ are iid $N\left(0,\sigma_0^2\right)$ across i and over t.*

Assumption 2 ***(Parameter Set)***

(a) The localizing parameter c and the variance parameter σ^2 of $\varepsilon_{i,t}$ take values in a compact subset $\mathbb{C}\times\mathbb{V}$ of $\mathbb{R}^2$.

(b) The true localizing parameter c_0 and the true variance parameter σ_0^2 are in interior of the parameter subsets $\mathbb{C}$ and $\mathbb{V}$, respectively.

Assumption 3 ***(Initial Conditions)*** $y_{i,0}=0$ *for all i.*

Assumption 3 on the initial condition is made mainly to simplify the arguments that follow. When the initial errors $y_{i,0}$ are random, the corresponding log-likelihood is obtained by conditioning on the initial errors. Some changes in the limit theory are to be expected in the case of distant initial conditions, as in Phillips and Lee (1986) and Canjels and Watson (1997), but otherwise this assumption has little bearing on the main results.

3 Estimation when the Trends are Unknown

This section studies the realistic situation of the panel model (1) when the trend functions are unknown. The following two subsections investigate the two cases of homogeneous deterministic trends and heterogeneous deterministic trends.

3.1 Homogeneous Trends

Suppose $G_i(t)$ in (1) is linear and homogeneous across i. Specifically, let us impose the following condition.

Assumption 4 ***(Homogeneous Trends)*** $G_i(t)=\delta t$.

The linear trend assumption is relevant for much empirical work and it simplifies formulae and derivations. However, the main thrust of the theory in this section continues to hold for general polynomial trends.

Let δ_0 denote the true value of δ. Then the data $z_{i,t}$ are generated by

$$\begin{aligned} z_{i,t} &= \delta_0 t + y_{i,t} \\ y_{i,t} &= \left(1+\frac{c_0}{T}\right) y_{i,t-1}+\varepsilon_{i,t}. \end{aligned}$$

Let $z^{N,T} = (z_{1,1}, ..., z_{N,T})'$, and define $y_{i,t}(\delta) = z_{i,t} - \delta t$, and $\varepsilon_{i,t}(\delta, c) = y_{i,t}(\delta) - (1 + \frac{c}{T})y_{i,t-1}(\delta)$. Let Δ_c be the quasi-differencing operator, $\Delta_c = 1 - aL$, where L is the lag operator and $a = 1 + \frac{c}{T}$.

Under the Gaussian assumption, the log-likelihood function of the panel data $z^{N,T}$ is

$$L_{N,T}\left(c, \delta, \sigma_\varepsilon^2; z^{N,T}\right) = -\frac{NT}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta_c z_{i,t} - \delta\left(1 - c\frac{t-1}{T}\right)\right)^2.$$

Since the parameter c is our main interest, we focus on the concentrated log-likelihood. For fixed c and σ^2, the log-likelihood $L_{N,T}(c, \delta, \sigma^2; z^{N,T})$ is maximized by $\hat{\delta}(c)$, where

$$\begin{aligned}\hat{\delta}(c) &= \frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(1 - c\frac{t-1}{T}\right)\left(\Delta z_{i,t} - \frac{c}{T}z_{i,t-1}\right)}{\sum_{t=1}^{T}\left(1 - c\frac{t-1}{T}\right)^2} \\ &= \delta_0 + \frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(1 - c\frac{t-1}{T}\right)\left(\varepsilon_{i,t}(\delta_0, c_0) - (c - c_0)\frac{y_{i,t-1}(\delta_0)}{T}\right)}{\sum_{t=1}^{T}\left(1 - c\frac{t-1}{T}\right)^2}. \qquad (4)\end{aligned}$$

Substituting $\hat{\delta}(c)$ in $L_{N,T}\left(c, \delta, \sigma^2; z^{N,T}\right)$ gives the following concentrated log-likelihood function;

$$\begin{aligned}&L_{N,T}\left(c, \hat{\delta}(c), \sigma^2; z^{N,T}\right) \\ &= -\frac{NT}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta_c z_{i,t} - \hat{\delta}(c)\left(1 - c\frac{t-1}{T}\right)\right)^2.\end{aligned}$$

Maximizing $L_{N,T}\left(c, \hat{\delta}(c), \sigma^2; z^{N,T}\right)$, we find the MLE of σ^2 as

$$\hat{\sigma}^2(c) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left(\Delta_c z_{i,t} - \hat{\delta}(c)\left(1 - c\frac{t-1}{T}\right)\right)^2.$$

Plugging $\hat{\sigma}^2(c)$ into $L_{N,T}\left(c, \hat{\delta}(c), \sigma^2; z^{N,T}\right)$ leads to the following concentrated log-likelihood,

$$\begin{aligned}&L_{N,T}\left(c, \hat{\delta}(c), \hat{\sigma}^2(c); z^{N,T}\right) \\ &= -\frac{NT}{2}\log\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left(\Delta_c z_{i,t} - \hat{\delta}(c)\left(1 - c\frac{t-1}{T}\right)\right)^2\right) - \frac{NT}{2} \\ &= -\frac{NT}{2}\log\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta_c z_{i,t} - \hat{\delta}(c)\left(1 - c\frac{t-1}{T}\right)\right)^2\right) - \frac{NT}{2} + \frac{NT}{2}\log T.\end{aligned}$$

The MLE $\hat{c}$ is obtained by maximizing the concentrated log-likelihood $L_{N,T}\left(c, \hat{\delta}(c), \hat{\sigma}_\varepsilon^2(c); z^{N,T}\right)$, so that

$$L_{N,T}\left(\hat{c}, \hat{\delta}(\hat{c}), \hat{\sigma}_\varepsilon^2(\hat{c}); z^{N,T}\right) = \max_{c\in\mathbb{C}} L_{N,T}\left(c, \hat{\delta}(c), \hat{\sigma}_\varepsilon^2(c); z^{N,T}\right), \qquad (5)$$

which is equivalent to maximizing[1]

$$\max_{c\in\mathbb{C}} l_{N,T}\left(c,\hat{\delta}\left(c\right),\hat{\sigma}_{\varepsilon}^{2}\left(c\right);z^{N,T}\right),$$

where

$$l_{N,T}\left(c,\hat{\delta}\left(c\right),\hat{\sigma}_{\varepsilon}^{2}\left(c\right);z^{N,T}\right) = -\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta_{c}z_{i,t}-\hat{\delta}\left(c\right)\left(1-c\frac{t-1}{T}\right)\right)^{2}$$
$$+\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta_{c_0}z_{i,t}-\hat{\delta}\left(c_0\right)\left(1-c_0\frac{t-1}{T}\right)\right)^{2}. \tag{6}$$

To investigate the consistency of the MLE $\hat{c}$ as $(N,T\to\infty)$, we write

$$\begin{aligned} & l_{N,T}\left(c,\hat{\delta}\left(c\right),\hat{\sigma}_{\varepsilon}^{2}\left(c\right);z^{N,T}\right) \\ = & -\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{i,t}\left(\delta_0,c_0\right)-\left(c-c_0\right)\frac{y_{i,t-1}\left(\delta_0\right)}{T}-\left(\hat{\delta}\left(c\right)-\delta_0\right)\left(1-c\frac{t-1}{T}\right)\right)^{2} \\ & +\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\varepsilon_{i,t}\left(\delta_0,c_0\right)-\left(\hat{\delta}\left(c_0\right)-\delta_0\right)\left(1-c_0\frac{t-1}{T}\right)\right\}^{2}. \end{aligned} \tag{7}$$

It then follows that as $(N,T\to\infty)$

$$l_{N,T}\left(c,\hat{\delta}\left(c\right),\hat{\sigma}_{\varepsilon}^{2}\left(c\right);z^{N,T}\right)\to_{p}-\frac{1}{2}\left(c-c_0\right)^{2}\int_{0}^{1}\int_{0}^{r}e^{2c_0(r-s)}dsdr \tag{8}$$

uniformly in c. The proof of (8) is given in the Appendix. Note that the limit function $-\frac{1}{2}\left(c-c_0\right)^{2}\int_{0}^{1}\int_{0}^{r}e^{2c_0(r-s)}dsdr$, is continuous and concave over $\mathbb{R}$ and is uniquely maximized at the true parameter $c=c_0$. Therefore, the MLE $\hat{c}$ that maximizes the objective function $l_{N,T}\left(c,\hat{\delta}\left(c\right),\hat{\sigma}_{\varepsilon}^{2}\left(c\right);z^{N,T}\right)$ is consistent for the localizing parameter c_0 as $(N,T\to\infty)$ by standard asymptotic theory (*e.g.*, Theorem 2.1 in Newey and McFadden, 1994). Summarizing, we have the following result.

Theorem 1 *Under Assumptions 1–4,* $\hat{c}\to_{p}c_0$ *as* $(N,T\to\infty)$.

Next, we derive the limit distribution of $\hat{c}$. Since the log-likelihood function $L_{N,T}\left(c,\hat{\delta}\left(c\right),\hat{\sigma}_{\varepsilon}^{2}\left(c\right);z^{N,T}\right)$ is differentiable with respect to c and since $\hat{c}$ is consistent for c_0, a point in an interior of the parameter set $\mathbb{C}$, the MLE $\hat{c}$ solves the following first order condition with probability one;

$$0 = \frac{dL_{N,T}\left(\hat{c},\hat{\delta}\left(\hat{c}\right),\hat{\sigma}_{\varepsilon}^{2}\left(\hat{c}\right);z^{N,T}\right)}{dc}=\frac{\partial L_{N,T}\left(\hat{c},\hat{\delta}\left(\hat{c}\right),\hat{\sigma}_{\varepsilon}^{2}\left(\hat{c}\right);z^{N,T}\right)}{\partial c}$$

[1] Notice that the second term, $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta_{c^0}z_{i,t}-\hat{\delta}\left(c^0\right)\left(1-c^0\frac{t-1}{T}\right)\right)^2$, in (6) below is not a function of c.

$$= \left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\Delta z_{i,t}-\hat{\delta}(\hat{c})-\hat{c}\left(\frac{z_{i,t-1}}{T}-\hat{\delta}(\hat{c})\frac{t-1}{T}\right)\right\}^{2}\right]^{-1}$$
$$\times\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\Delta z_{i,t}-\hat{\delta}(\hat{c})-\hat{c}\left(\frac{z_{i,t-1}}{T}-\hat{\delta}(\hat{c})\frac{t-1}{T}\right)\right)\right.$$
$$\left.\times\left(\frac{z_{i,t-1}}{T}-\hat{\delta}(\hat{c})\frac{t-1}{T}\right)\right\}, \tag{9}$$

where the second equality holds by the Envelope Function Theorem.

Theorem 2 *Under Assumptions 1–4,* $\sqrt{N}(\hat{c}-c_0)\Rightarrow N\left(0,\frac{1}{\sigma_0^2\int_0^1\int_0^r e^{2c_0(r-s)}dsdr}\right)$ *as* $(N,T\to\infty)$.

In view of (4), the MLE of the homogeneous trend coefficient $\hat{\delta}$ is found to be

$$\hat{\delta}(\hat{c})=\frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(1-\hat{c}\frac{t-1}{T}\right)\left(\Delta z_{i,t}-\frac{\hat{c}}{T}z_{i,t-1}\right)}{\sum_{t=1}^{T}\left(1-\hat{c}\frac{t-1}{T}\right)^{2}}, \tag{10}$$

and as $(N,T\to\infty)$, it is possible to show that

$$\sqrt{NT}\left(\hat{\delta}(\hat{c})-\delta_0\right)\Rightarrow N\left(0,\frac{\sigma_0^2}{\int_0^1(1-c_0r)^2\,dr}\right). \tag{11}$$

The proof of (11) is straightforward using the results in Lemmas 6 and 7 and the consistency of $\hat{c}$ and is therefore omitted. Summarizing, we have:

Theorem 3 *Under Assumptions 1–4,* $\sqrt{NT}\left(\hat{\delta}(\hat{c})-\delta_0\right)\Rightarrow N\left(0,\frac{\sigma_0^2}{\int_0^1(1-c_0r)^2dr}\right)$ *as* $(N,T\to\infty)$.

Remarks

(a) When the trends in the panel regression model (1) are homogeneous, the Gaussian MLE $\hat{c}$ is $\sqrt{N}-$consistent and has an asymptotic normal limit distribution that is equivalent to the normal limit distribution in (3), a result that continues to hold in a model with general polynomial deterministic trends.

(b) Since $\hat{\delta}(c)$ is a nonlinear function of c in general, it is not easy to find a closed form solution of the first order condition (9). In this case, to solve the first order condition (9), it would be common to employ an iteration involving the use of a preliminary $\sqrt{N}$-consistent estimator, $\tilde{c}$, say, which leads to a second stage estimator via suitable numerical optimization, such as Newton-Raphson. In the model (1), a natural candidate for the preliminary estimator would be

$$\tilde{c}=\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{z_{i,t-1}}{T}-\hat{\delta}(c)\frac{t-1}{T}\right)^{2}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta z_{i,t}-\hat{\delta}(c)\right)\left(\frac{z_{i,t-1}}{T}-\hat{\delta}(c)\frac{t-1}{T}\right)\right), \tag{12}$$

where c is arbitrarily chosen. Then, using the first step estimator $\tilde{c}$, we may construct the following second step estimator;

$$\breve{c} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \left(\frac{z_{i,t-1}}{T} - \hat{\delta}(\tilde{c}) \frac{t-1}{T} \right)^2 \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \left(\Delta z_{i,t} - \hat{\delta}(\tilde{c}) \right) \left(\frac{z_{i,t-1}}{T} - \hat{\delta}(\tilde{c}) \frac{t-1}{T} \right) \right). \tag{13}$$

An important feature of the first step estimator $\tilde{c}$ is that it is asymptotically as efficient as the MLE $\hat{c}$, because

$$\sqrt{N}(\tilde{c} - c_0) \Rightarrow N\left(0, \frac{1}{\sigma_0^2 \int_0^1 \int_0^r e^{2c_0(r-s)} ds dr}\right), \tag{14}$$

the proof of which is provided in the Appendix.

(c) From Theorem 2 we see that the asymptotic variance of $\sqrt{N}(\hat{c} - c_0)$ depends on the true parameter c_0. Figure 2 below graphs the asymptotic variance of $\sqrt{N}(\hat{c} - c_0)$. As is apparent in the graph, the asymptotic variance of $\sqrt{N}(\hat{c} - c_0)$ decreases rather rapidly to zero as c_0 increases.

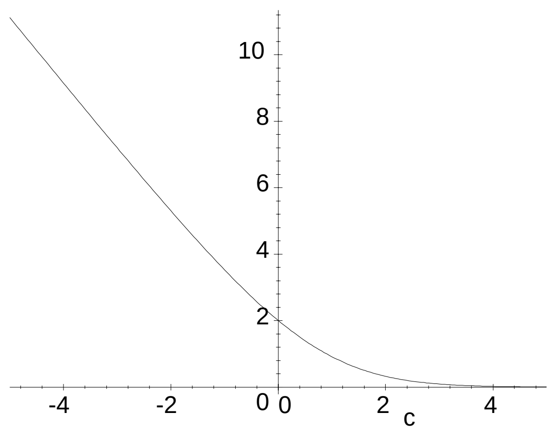

Figure 1. Graph of the Asymptotic Variance of the MLE $\hat{c}$

3.2 Heterogeneous Trends

Here we study the asymptotic properties of the MLE of the panel regression model (1) with heterogeneous deterministic trends specified as follows.

Assumption 5 ***(Heterogeneous Trends)*** $G_i(t) = \delta_i t$.

Suppose that the true trend coefficients are $\{\delta_{0,i} : i = 1, ..., N\}$. Then, the data $z_{i,t}$ are generated by the following parametric model:

$$\begin{aligned} z_{i,t} &= \delta_{0,i} t + y_{i,t} \\ y_{i,t} &= \left(1 + \frac{c_0}{T}\right) y_{i,t-1} + \varepsilon_{i,t}. \end{aligned} \tag{15}$$

Let $\delta^{N,0} = (\delta_{0,1}, ..., \delta_{0,N})'$ and $z^{N,T} = (z_{1,1}, ..., z_{N,T})'$. Define $y_{i,t}(\delta_i) = z_{i,t} - \delta_i t$ and $\varepsilon_{i,t}(\delta_i, c) = y_{i,t}(\delta_i) - (1 + \frac{c}{T})y_{i,t-1}(\delta_i)$.

Under Gaussianity, the standardized log-likelihood function is

$$L_{N,T}\left(c, \delta^N, \sigma^2; z^{N,T}\right) = -\frac{NT}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta_c z_{i,t} - \delta_i\left(1 - c\frac{t-1}{T}\right)\right)^2.$$

Given c, the MLE for δ_i is

$$\hat{\delta}_i(c) = \left(\sum_{t=1}^{T}\left(1 - c\frac{t-1}{T}\right)^2\right)^{-1}\left(\sum_{t=1}^{T}\Delta_c z_{i,t}\left(1 - c\frac{t-1}{T}\right)\right),$$

leading to the concentrated log-likelihood function

$$L_{N,T}\left(c, \hat{\delta}(c)^N, \sigma^2; z^{N,T}\right) = -\frac{NT}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta_c z_{i,t} - \hat{\delta}_i(c)\left(1 - c\frac{t-1}{T}\right)\right)^2, \tag{16}$$

where $\hat{\delta}(c)^N = (\hat{\delta}_1(c), ..., \hat{\delta}_N(c))'$.

We seek to show that $\hat{c}$, the MLE of c maximizing $L_{N,T}(c, \hat{\delta}(c)^N, \sigma^2; z^{N,T})$, is inconsistent. To do so, it is simplest to assume that the variance of $\varepsilon_{i,t}$, σ^2, is known. By definition

$$\begin{aligned}\Delta_c z_{i,t} &= \delta_{0,i}\Delta_c t + \Delta_c y_{i,t}(\delta_{0,i}) = \delta_{0,i}\Delta_c t + \Delta_{c_0} y_{i,t}(\delta_{0,i}) + (\Delta_c - \Delta_{c_0}) y_{i,t}(\delta_{0,i}) \\ &= \delta_{0,i}\left(1 - c\frac{t-1}{T}\right) + \varepsilon_{i,t}(\delta_{0,i}, c_0) - (c - c_0)\frac{y_{i,t-1}(\delta_{0,i})}{T},\end{aligned}$$

so we can write

$$\begin{aligned}&\frac{1}{N}\left[L(z^{N,T}; \hat{\delta}^N(c), c) - L(z^{N,T}; \hat{\delta}^N(c_0), c_0)\right] \\ = \; & -\frac{1}{2\sigma^2 N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\varepsilon_{i,t}(\delta_{0,i}, c_0) - (c - c_0)\frac{y_{i,t-1}(\delta_{0,i})}{T} - \left(\hat{\delta}_i(c) - \delta_{0,i}\right)\left(1 - c\frac{t-1}{T}\right)\right\}^2 \\ & + \frac{1}{2\sigma^2 N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\varepsilon_{i,t}(\delta_{0,i}, c_0) - \left(\hat{\delta}_i(c_0) - \delta_{0,i}\right)\left(1 - c_0\frac{t-1}{T}\right)\right\}^2.\end{aligned} \tag{17}$$

Lemma 4 *Suppose Assumptions 1–3, and 5 hold and that the variance of $\varepsilon_{i,t}$, σ^2 is known. Then, as $(N, T \to \infty)$,*

$$\begin{aligned}&\frac{1}{N}\left[L(z^{N,T}; \hat{\delta}^N(c), c) - L(z^{N,T}; \delta^{0,N}, c_0)\right] \\ \to_p \; & \frac{(c - c_0)^2\int_0^1\int_0^1(1 - cr)(1 - cs)\int_0^{r\wedge s} e^{c_0(r+s-2p)}dp\,ds\,dr}{2\int_0^1(1 - cr)^2 dr} \\ & - \frac{2(c - c_0)\int_0^1\int_0^r e^{c_0(r-s)}(1 - cr)(1 - cs)\,ds\,dr}{2\int_0^1(1 - cr)^2 dr} \\ & - \frac{1}{2}(c - c_0)^2\int_0^1\int_0^r e^{2c_0(r-s)}ds\,dr \\ = \; & G(c; c_0), \text{ say, uniformly in } c.\end{aligned}$$

According to this lemma, the standardized concentrated log-likelihood function, $\frac{1}{N}\left[L(z^{N,T};\hat{\delta}^{N}(c),c)-L(z^{N,T};\delta^{0,N},c_0)\right]$, has the uniform limit $G(c;c_0)$, a function that is continuous on the parameter set $\mathbb{C}$. Hence, the MLE $\hat{c}$ that maximizes $\frac{1}{N}\left[L(z^{N,T};\hat{\delta}^{N}(c),c)-L(z^{N,T};\delta^{0,N},c_0)\right]$ converges in probability to the point that maximizes the limit function $G(c;c_0)$. For the MLE $\hat{c}$ to be consistent, the true parameter c_0 must maximize $G(c;c_0)$; and, conversely, if some point $\overline{c}\neq c_0$ maximizes $G(c;c_0)$, then the MLE $\hat{c}$ is not consistent.

We proceed to differentiate $G(c;c_0)$ with respect to c and evaluate the derivative at $c=c_0$. Since the true parameter c_0 is in an interior of the parameter set $\mathbb{C}$ and the limit function $G(c;c_0)$ is differentiable, if c_0 maximizes $G(c;c_0)$, its first derivative at $c=c_0$ must necessarily be zero. However, direct calculation shows that

$$
\begin{aligned}
&\frac{dG(c;c_0)}{dc}\Big|_{c=c_0}\\
&= \frac{\left\{-\int_0^1\int_0^r e^{c_0(r-s)}(1-cr)(1-cs)\,dsdr\right\}\left\{\int_0^1(1-cr)^2\,dr\right\}}{\left(\int_0^1(1-cr)^2\,dr\right)^2}\Big|_{c=c_0}\\
&= \frac{-3+2c_0}{6\left(1-c_0+\frac{1}{3}(c_0)^2\right)}\neq 0 \qquad (18)
\end{aligned}
$$

if $c_0\neq\frac{3}{2}$. Therefore, for all $c_0\neq\frac{3}{2}$ the limit function $G(c;c_0)$ cannot attain a maximum at $c=c_0$. For $c_0=\frac{3}{2}$, we graph of the function $G(c;\frac{3}{2})$ in Figure 1. As the figure shows, the limit function $G(c;\frac{3}{2})$ has a local minimum at $c=\frac{3}{2}$, and so $G(c;c_0)$ does not attain a maximum at $c=c_0$ for any value of c_0.

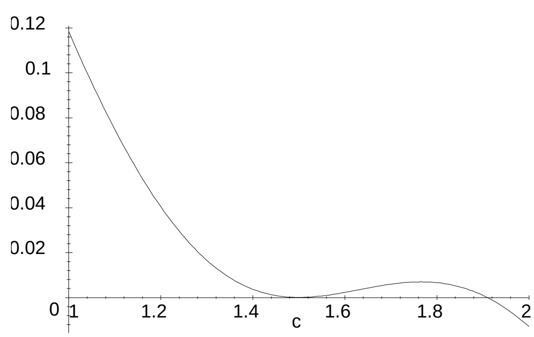

Figure 2. Graph of $G\left(c,\frac{2}{3}\right)$

In summary, we have the following result.

Theorem 5 ***(Inconsistency)*** *Suppose Assumptions 1–3 and 5 hold. Then, the MLE $\hat{c}$ is inconsistent when $(N,T\to\infty)$.*

Remarks

(a) From (18), it is apparent that $\frac{dG(c;c_0)}{dc}|_{c=c_0}$ tends to zero as $|c_0|$ increases to infinity. So when the absolute value of c_0 is large, we may expect the limit function to be maximized at a value close to c_0. In such cases, the probability limit of the MLE can be expected to be close to the true parameter c_0, even though the MLE is inconsistent. To investigate, we present graphs of the limit functions $G(c, 4)$ and $G(c, -8)$ in Figure 3 and Figure 4, respectively. When the true parameter $c_0 = 4$, the limit of the standardized concentrated log-likelihood $G(c, 4)$ is maximized around $c = 4.057$, which is close to the true parameter value, involving only a 1% bias. On the other hand when the true parameter $c_0 = -8$, $G(c, -8)$ is maximized around $c = -10.27$, giving a 28% asymptotic bias. These results indicate that we can expect the inconsistency of the MLE to be greater when c_0 is negative.

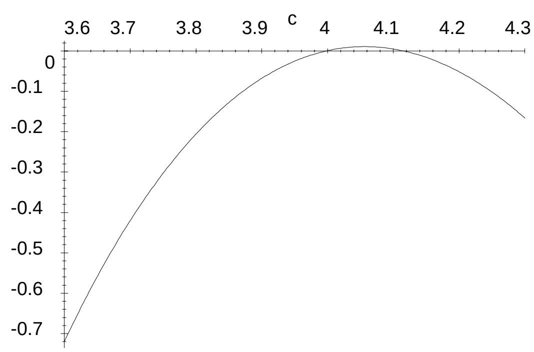

Figure 3. Graph of $G(c, 4)$

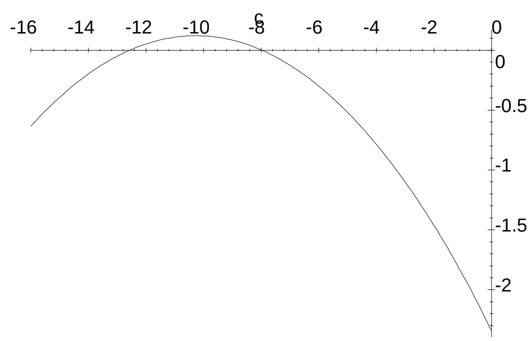

Figure 4. Graph of $G(c, -8)$

(b) The inconsistency of the MLE $\hat{c}$ in the above theorem is an instance of the so-called incidental parameter problem (Neyman and Scott, 1948). Incidental parameter problems are known to arise in other panel data regression models, the celebrated example being the dynamic panel regression model with fixed effects. In that case, the panel data $z_{i,t}$ are generated by the autoregression

$$z_{i,t} = \delta_i + az_{i,t-1} + \varepsilon_{i,t},$$

where $|a| < 1$ and the $\varepsilon_{i,t}$ are iid $N\left(0, \sigma^2\right)$. The individual intercept terms δ_i enter the model to account for individual effects in the panel data $z_{i,t}$. The main focus of interest in this model is the estimation of the common parameter a, and the individual effects δ_i are incidental parameters. For simplicity, assume that $z_{i,0} = 0$ for all i. Then, the MLE of a is equivalent to the within estimator, defined as:

$$\begin{aligned} \hat{a} &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{i,t-1} - \bar{z}_{i,-}\right)\left(z_{i,t} - \bar{z}_{i,.}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{i,t-1} - \bar{z}_{i,-}\right)^2} \\ &= a + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{i,t-1} - \bar{z}_{i,-}\right)\left(\varepsilon_{i,t} - \bar{\varepsilon}_{i,.}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{i,t-1} - \bar{z}_{i,-}\right)^2}, \end{aligned} \tag{19}$$

where $\bar{z}_{i,-} = \frac{1}{T}\sum_{t=1}^{T} z_{i,t-1}$, $\bar{z}_{i,.} = \frac{1}{T}\sum_{t=1}^{T} z_{i,t}$, and $\bar{\varepsilon}_{i,.} = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t}$. In this case, when $N \to \infty$ for fixed T, we know that $\hat{a} \nrightarrow_p a$, due to the correlation between $z_{i,t-1} - \bar{z}_{i,-}$ and $\varepsilon_{i,t} - \bar{\varepsilon}_{i,.}$. So, in this case with $N \to \infty$ and T fixed, the MLE $\hat{a}$ is inconsistent (Nickell, 1981).

(c) An especially interesting aspect of the model (15) is that the incidental parameter problem leading to the inconsistency of the MLE $\hat{c}$ continues to be present even though $T \to \infty$ as well as $N \to \infty$. In contrast, the incidental parameter problem that gives rise to the inconsistency of $\hat{a}$ in (19) disappears if $T \to \infty$ fast enough when $N \to \infty$.

4 Monte-Carlo Simulations

This section reports some simulations designed to explore the finite sample properties the maximum likelihood estimators studied in the previous section. First, to investigate the homogeneous trend model, data $z_{i,t}$ was generated by the system

$$\begin{aligned} z_{i,t} &= \delta_0 t + y_{i,t}, \qquad \delta_0 = 3, \\ y_{i,t} &= (1 + \frac{c_0}{T})y_{i,t-1} + \varepsilon_{i,t}, \qquad c_0 \in \{-4, -2, 0, 2, 4\}, \end{aligned} \tag{20}$$

where the $\varepsilon_{i,t}$ are *iid* $N(0,1)$ across i and over t, and the initial values of $y_{i,0}$ are zeros. Following the notation used in the previous section, we let $\hat{c}$ denote the MLE of the localizing parameter and $\hat{\delta}(\hat{c})$ to be the MLE of the homogeneous trend coefficient in (10). Also, let $\tilde{c}$ denote the first step estimator in (12), and $\check{c}$ denote the second step estimator in (13).

The main goals of the simulation experiment with model (20) are as follows: (i) to examine the finite sample properties of the MLE's $\hat{\delta}(\hat{c})$ and $\hat{c}$ by comparing their mean squared errors for various parameter configurations; and (ii) to compare the asymptotic efficiencies of the three estimators considered in Section 3.1 – the MLE $\hat{c}$, the first step estimator $\tilde{c}$, and the second step estimator $\check{c}$. From the DGP (20), we generate panels of 16 different sizes, with $N \in \{25, 50, 75, 100\}$ and $T \in \{25, 50, 75, 100\}$. The estimates $\hat{\delta}(\hat{c})$, $\hat{c}$, $\tilde{c}$, and $\check{c}$ are computed and 1000 replications used to calculate their mean squared errors. Table 1 reports the mean squared errors of $\hat{\delta}(\hat{c})$ and $\hat{c}$. The first column of the table contains the sample size, the top element of each column contains the true parameter value, and the first and the second elements in the table are the MSE of $\hat{\delta}(\hat{c})$ and the MSE of $\hat{c}$, respectively.

Table 1. MSE of $\hat{\delta}(\hat{c})$ and $\hat{c}$

	(δ_0, c_0)				
(N, T)	$(3, -4)$	$(3, -2)$	$(3, 0)$	$(3, 2)$	$(3, 4)$
$(25, 25)$	0.010, 8.486	0.010, 2.084	0.011, 0.125	0.012, 2.583	0.014, 9.411
$(25, 50)$	0.009, 3.779	0.009, 0.968	0.010, 0.117	0.012, 1.160	0.012, 4.042
$(25, 75)$	0.009, 1.064	0.009, 0.344	0.010, 0.114	0.011, 0.318	0.010, 1.032
$(25, 100)$	0.009, 0.421	0.009, 0.249	0.009, 0.106	0.010, 0.020	0.010, 0.019
$(50, 25)$	0.009, 8.803	0.010, 2.189	0.010, 0.052	0.010, 2.380	0.012, 9.146
$(50, 50)$	0.009, 3.887	0.009, 0.971	0.009, 0.046	0.010, 1.079	0.010, 4.038
$(50, 75)$	0.009, 1.034	0.009, 0.290	0.009, 0.049	0.010, 0.282	0.009, 1.019
$(50, 100)$	0.009, 0.204	0.009, 0.116	0.009, 0.047	0.010, 0.012	0.009, 0.017
$(75, 25)$	0.009, 8.817	0.009, 2.184	0.010, 0.036	0.010, 2.361	0.011, 9.142
$(75, 50)$	0.009, 3.911	0.009, 0.974	0.009, 0.034	0.010, 1.061	0.010, 4.034
$(75, 75)$	0.009, 1.021	0.009, 0.273	0.009, 0.030	0.010, 0.273	0.009, 1.017
$(75, 100)$	0.009, 0.145	0.009, 0.081	0.009, 0.032	0.009, 0.009	0.009, 0.016
$(100, 25)$	0.009, 8.920	0.009, 2.224	0.009, 0.023	0.010, 2.312	0.010, 9.078
$(100, 50)$	0.009, 3.981	0.009, 0.999	0.009, 0.022	0.010, 1.033	0.010, 4.014
$(100, 75)$	0.009, 1.047	0.009, 0.280	0.009, 0.023	0.010, 0.264	0.009, 1.013
$(100, 100)$	0.009, 0.107	0.009, 0.059	0.009, 0.022	0.009, 0.008	0.009, 0.016

Several features of the results are notable. First, the MSE of $\hat{c}$ is much more sensitive to the sample size than the MSE of $\hat{\delta}(\hat{c})$. Second, the MSE of $\hat{c}$ decreases more as T increases than when N increases. For example, when $(\delta_0, c_0) = (3, -4)$ and the sample size changes from $(N, T) = (50, 75)$ to $(N, T) = (50, 100)$, the MSE of $\hat{c}$ decreases from 1.034 to 0.204. On the other hand, when the sample size changes from $(N, T) = (50, 75)$ to $(N, T) = (75, 75)$, the MSE of $\hat{c}$ decreases from 1.034 to 1.021. A more interesting feature is that when the sample size is small, increases in N sometimes lead to a deterioration in the finite sample properties of $\hat{c}$. For example, when $(\delta_0, c_0) = (3, -4)$ again, and the sample size changes from $(N, T) = (50, 50)$ to $(N, T) = (75, 50)$, the MSE of $\hat{c}$ increases from 3.887 to 3.911. Third, when $c_0 = 0$, the finite sample performance of $\hat{c}$ is apparently far better than it is for $c_0 < 0$. Also, as implied by the form of the asymptotic variance (see Theorem 2 and Remark(c) following Theorem 3), the MSE of $\hat{c}$ decreases as c_0 increases.

Table 2. MSE of $\tilde{c}$ and $\breve{c}$

(N,T)	(δ_0, c_0)				
	$(3,-4)$	$(3,-2)$	$(3,0)$	$(3,2)$	$(3,4)$
$(25,25)$	8.505, 8.486	2.086, 2.084	0.125, 0.125	2.585, 2.583	9.449, 9.412
$(25,50)$	3.825, 3.780	0.972, 0.968	0.117, 0.117	1.173, 1.161	4.157, 4.042
$(25,75)$	1.098, 1.064	0.346, 0.344	0.113, 0.114	0.336, 0.318	1.124, 1.034
$(25,100)$	0.398, 0.412	0.242, 0.249	0.105, 0.106	0.022, 0.020	0.056, 0.012
$(50,25)$	8.812, 8.803	2.190, 2.189	0.052, 0.052	2.381, 2.380	9.169, 9.147
$(50,50)$	3.906, 3.887	0.973, 0.971	0.046, 0.046	1.085, 1.079	4.090, 4.038
$(50,75)$	1.054, 1.034	0.292, 0.290	0.049, 0.049	0.291, 0.282	1.062, 1.019
$(50,100)$	0.200, 0.204	0.115, 0.116	0.047, 0.047	0.012, 0.012	0.025, 0.017
$(75,25)$	8.822, 8.817	2.184, 2.184	0.036, 0.036	2.362, 2.361	9.156, 9.142
$(75,50)$	3.926, 3.911	0.975, 0.974	0.034, 0.034	1.065, 1.061	4.068, 4.034
$(75,75)$	1.034, 1.021	0.274, 0.273	0.030, 0.030	0.279, 0.273	1.046, 1.017
$(75,100)$	0.143, 0.145	0.080, 0.081	0.032, 0.032	0.010, 0.009	0.020, 0.016
$(100,25)$	8.924, 8.920	2.224, 2.224	0.023, 0.023	2.313, 2.312	9.087, 9.078
$(100,50)$	3.991, 3.981	1.000, 0.999	0.022, 0.022	1.036, 1.033	4.039, 4.014
$(100,75)$	1.057, 1.047	0.281, 0.280	0.023, 0.023	0.268, 0.264	1.034, 1.013
$(100,100)$	0.106, 0.107	0.058, 0.059	0.022, 0.022	0.008, 0.008	0.018, 0.016

*$c = 0$ is used in estimating $\tilde{c}$.

Table 2 reports the mean squared errors of the first step estimator $\tilde{c}$ and the second step estimator $\breve{c}$. The simulations cover the same16 panel data sizes and use the same number of replications as before. The layout of the table is the same as Table 1. To calculate $\tilde{c}$ we use $c = 0$ for quasi-differencing the data. This experiment focuses on comparing the finite sample properties of three asymptotically equivalent estimators, the MLE $\hat{c}$, the first step estimator $\tilde{c}$, and the second step estimator $\breve{c}$. As is apparent from comparison of Table 1 and Table 2, there are apparently no major differences in the mean squared errors of the three asymptotically equivalent estimators. So, finite sample effects are not important in this case.

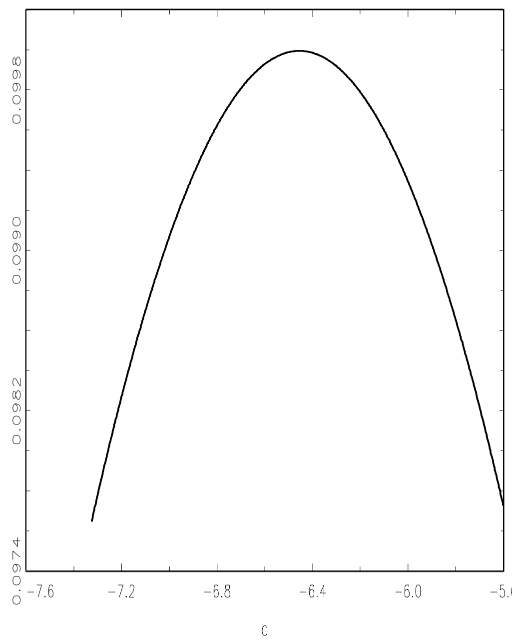

Figure. 5: Density of $\hat{c}$ when the true $c_0 = -4$

The next simulation experiment involves the heterogeneous trend model, for which the generating process is taken to be

$$\begin{aligned} z_{i,t} &= \delta_{0,i}t + y_{i,t}, \qquad \delta_{0,i} \sim iid\ Uniform[0,4], \\ y_{i,t} &= (1 + \frac{c_0}{T})y_{i,t-1} + \varepsilon_{i,t}, \qquad c_0 \in \{-4, 0, 4\}, \end{aligned} \tag{21}$$

where $\varepsilon_{i,t}$ are iid $N(0,1)$ across i and over t, and $y_{i,0} = 0$ for all i. The main purpose of this simulation is to explore the finite sample manifestation of the inconsistency of the MLE $\hat{c}$. For this, we generated a panel data set with size dimensions $N = 300$, $T = 300$, and found the Gaussian MLE $\hat{c}$ by a grid search method. The grid used in the simulation is 0.075. 1,000 replications were employed. Estimated densities functions of the Gaussian MLE $\hat{c}$ of the panel models with $c_0 \in \{-4, 0, 4\}$ are shown in Figures 5-7.

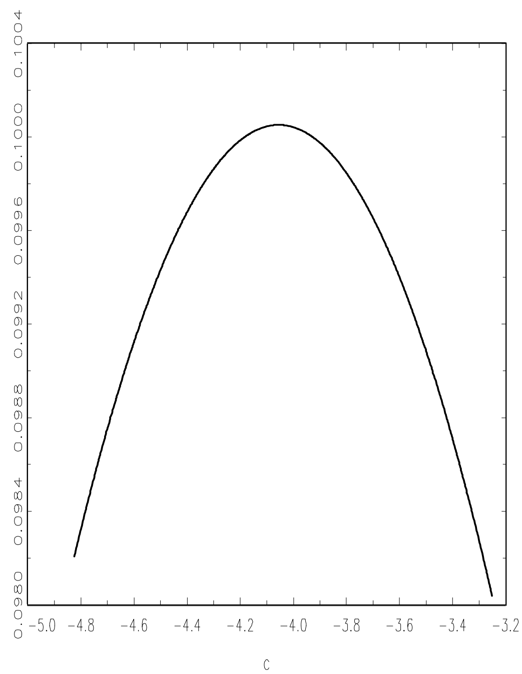

Figure 6. Density of $\hat{c}$ when the true $c_0 = 0$

As is apparent in Figures 5 and 6, the density of $\hat{c}$ is concentrated in a region substantially removed from the true parameter value when $c = -4$ and $c = 0$. On the other hand, in Figure 7, when $c_0 = 4$, the density of the $\hat{c}$ appears to be concentrated around 4.16, a value that is quite near the true value. This outcome corroborates the asymptotic analysis of the previous section, where it was shown that when $c_0 = 4$, the standardized Gaussian log-likelihood converges in probability to the limit function $G(c, 4)$ whose maximum is close to the true value $c_0 = 4$.

5 Conclusion

This paper explores the asymptotic properties of the Gaussian maximum likelihood estimator of the localizing parameter in a panel model with deterministic and stochastic trends. Several new findings emerge. First, when the trends are homogenous across individuals in the panel, the Gaussian MLE of the common localizing parameter is $\sqrt{N}-$ consistent and has a limiting normal distribution that is equivalent to the asymptotic distribution of the Gaussian MLE of the model in which the deterministic trends are known. So, in this case, trend elimination carries no cost in the limit, just as in the case of a stationary autoregression with trend. However, when the trends are heterogenous across individuals, the Gaussian MLE of the localizing parameter is shown to be inconsistent. The inconsistency is due to the presence of an infinite number of incidental parameters for the individual trends. Procedures for resolving this manifestation of the incidental parameter problem in panel regression are now being explored by the authors and will be reported in later work.

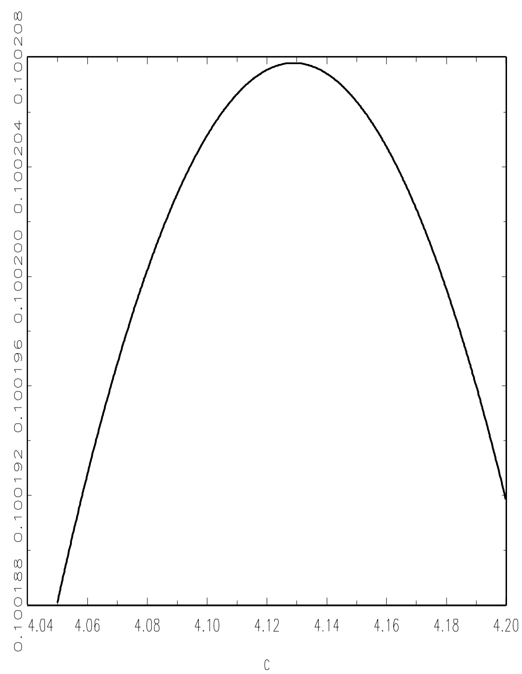

Figure 7. Density of $\hat{c}$ when the true $c_0 = 4$

6 Appendix

Lemma 6 *Suppose that* $\mathbb{C}$ *be a compact subset of* $\mathbb{R}$. *Assume that, for* $k = 1, ..., K$, $h_k(c, \tilde{c})$ *is a real-valued continuous function on* $\mathbb{C} \times \mathbb{C}$ *with* $h_k(c, c) = 0$, *and* $l_k(x, y)$ *is a real-valued continuous function on* $[0, 1] \times [0, 1]$. *Also, assume that* $f(x, c)$ *and* $g(x, c)$ *are continuous functions from* $[0, 1] \times \mathbb{C}$ *to* $\mathbb{R}$ *such that* $f(x, c) g(y, c) - f(x, \tilde{c}) g(y, \tilde{c}) = \sum_{k=1}^{K} h_k(c, \tilde{c}) l_k(x, y)$. *Suppose that* $y_{i,t} = \exp\left(\frac{c_0}{T}\right) y_{i,t-1} + \varepsilon_{i,t}$, *where* $\varepsilon_{i,t}$ *are iid* $\left(0, \sigma_0^2\right)$ *across* i *and over* t *and* $y_{i,0} = 0$. *Then, as* $(N, T \to \infty)$, *the following hold.*

(a) $\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} y_{i,t-1}^2 \to_p \sigma_0^2 \int_0^1 \int_0^r e^{2c_0(r-s)} ds dr.$

(b) $\frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{i,t} f\left(\frac{t}{T}, c\right)\right) \left(\frac{1}{T\sqrt{T}} \sum_{t=1}^{T} y_{i,t-1} g\left(\frac{t}{T}, c\right)\right) \to_p \sigma_0^2 \int_0^1 \int_0^r e^{c_0(r-s)} g(r, c) f(s, c) ds dr$ *uniformly in* c.

(c) $\frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{T\sqrt{T}} \sum_{t=1}^{T} y_{i,t-1} f\left(\frac{t}{T}, c\right)\right) \left(\frac{1}{T\sqrt{T}} \sum_{t=1}^{T} y_{i,t-1} g\left(\frac{t}{T}, c\right)\right)$

$\to_p \sigma_0^2 \int_0^1 \int_0^1 f(r,c) g(s,c) \int_0^{r \wedge s} e^{c_0(r+s-2p)} dp ds dr$ *uniformly in c.*

(d) $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{i,t} f\left(\frac{t}{T},c\right)\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{i,t} g\left(\frac{t}{T},c\right)\right) \to_p \sigma_0^2 \int_0^1 \int_0^1 f(r,c) g(s,c) ds dr$ *uniformly in c.*

Proof

Part (a) This holds by Lemma 9(a) in Moon and Phillips (1998). ■

Part (b) First, using Corollary 1 in Phillips and Moon (1999), we establish Part (b) for fixed c (pointwise convergence). Note that

$$\begin{aligned} & \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{i,t} f\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{i,t-1} g\left(\frac{t}{T},c\right)\right) \\ \Rightarrow\ & \frac{1}{N}\sum_{i=1}^{N}\sigma_0^2\left(\int_0^1 f(s,c) dW_i(s)\right)\left(\int_0^1 g(r,c) J_{c,i}(r) dr\right) \text{ as } T\to\infty \text{ for fixed } N \text{ and } c \\ \to\ & _p\sigma_0^2 \int_0^1 \int_0^r e^{c_0(r-s)} g(r,c) f(s,c) ds dr \text{ as } N \to \infty \text{ for fixed } c. \end{aligned}$$

According to Corollary 1 in Phillips and Moon (1998), this sequential limit becomes the joint limit if $Q_{i,T}(c) = \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{i,t} f\left(\frac{t}{T},c\right)\right)\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{i,t-1} g\left(\frac{t}{T},c\right)\right)$ is uniformly integrable in T for fixed c, which holds if

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

and

$$Q_{2,i,T}(c) = \left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{i,t-1} g\left(\frac{t}{T},c\right)\right)^2$$

are uniformly integrable in T for fixed c. Notice that $Q_{1,i,T}(c) \Rightarrow Q_{1,i}(c) = \sigma_0^2\left(\int_0^1 f(r,c) dW_i(r)\right)^2$, and $EQ_{1,i,T}(c) = \sigma_0^2 \frac{1}{T}\sum_{t=1}^{T} f\left(\frac{t}{T},c\right)^2 \to \sigma_0^2 \int_0^1 f(r,c)^2 dr = EQ_{1,i}(c)$ as $T \to \infty$ for all i. Then, by Theorem 5.4 in Billingsley (1968), $Q_{1,i,T}(c)$ are uniformly integrable in T for fixed c. By similar fashion, $Q_{2,i,T}(c)$ is also uniformly integrable in T for fixed c. Therefore, Part (b) is just established for fixed c.

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

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

Then, since the index set $\mathbb{C}$ is hypothesized to be compact, the pointwise convergence of $R_{N,T}(c)$ and the stochastic equicontinuity of $R_{N,T}(c)$ imply uniform convergence.

To show the stochastic equicontinuity of $R_{N,T}(c)$, first observe that

$$\begin{aligned}
&\sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}} |R_{N,T}(c) - R_{N,T}(\tilde{c})| \\
= &\sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}} \left| \frac{1}{N}\sum_{i=1}^{N} \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} \varepsilon_{i,t} y_{i,s-1} \left\{ f\left(\frac{t}{T},c\right) g\left(\frac{s}{T},c\right) - f\left(\frac{t}{T},\tilde{c}\right) g\left(\frac{s}{T},\tilde{c}\right) \right\} \right| \\
= &\sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}} \left| \frac{1}{N}\sum_{i=1}^{N} \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} \varepsilon_{i,t} y_{i,s-1} \left\{ \sum_{k=1}^{K} h_k(c,\tilde{c})\, l_k\left(\frac{t}{T},\frac{s}{T}\right) \right\} \right| \\
\leq &\sup_{1\leq k\leq K}\ \sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}} |h_k(c,\tilde{c})| \frac{1}{N}\sum_{i=1}^{N} \left| \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} \varepsilon_{i,t} y_{i,s-1} \left\{ \sum_{k=1}^{K} l_k\left(\frac{t}{T},\frac{s}{T}\right) \right\} \right|.
\end{aligned}$$

Since $h_k(c,\tilde{c})$ is continuous on a compact set with $h_k(c,c) = 0$ for all $k = 1,..,K$, we can make $\sup_{1\leq k\leq K} \sup_{|c-\tilde{c}|<\delta,c,\tilde{c}\in\mathbb{C}} |h_k(c,\tilde{c})|$ arbitrarily small by choosing a small $\delta > 0$. Also, under the assumptions in the lemma, it is not difficult to show that $\frac{1}{N}\sum_{i=1}^{N} \left| \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} \varepsilon_{i,t} y_{i,s-1} \left\{ \sum_{k=1}^{K} l_k\left(\frac{t}{T},\frac{s}{T}\right) \right\} \right| = O_p(1)$. Therefore, $R_{N,T}(c)$ is stochastically equicontinuous. ■

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

Lemma 7 *Suppose that $f(x,c)$ and $g(x,c)$ are continuous functions from $[0,1]\times\mathbb{C}$ to $\mathbb{R}$. Assume that $y_{i,t} = \exp\left(\frac{c_0}{T}\right) y_{i,t-1} + \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ are iid $\left(0,\sigma_0^2\right)$ across i and over t and $y_{i,0} = 0$.*

Then, as $(N,T\to\infty)$, the following hold.

(a) $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \varepsilon_{i,t} f\left(\frac{t}{T},c\right) \Rightarrow N\left(0,\sigma_0^2\int_0^1 f(r)^2\,dr\right)$.

(b) $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{i,t-1} f\left(\frac{t}{T},c\right) \Rightarrow N\left(0,\sigma_0^2\int_0^1\int_0^1\int_0^{r\wedge s} e^{c_0(r+s-2p)}dp f(r)f(s)drds\right)$.

(c) $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{i,t} \Rightarrow N\left(0,\sigma_0^2\int_0^1\int_0^r e^{2c_0(r-s)}dsdr\right)$

Proof The proofs verify the conditions of Theorem 3 in Phillips and Moon (1999).

Part (a) Following the notation in Phillips and Moon (1999), we let $Q_{i,T} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{i,t} f\left(\frac{t}{T}\right)$. Then, $Q_{i,T}$ are iid $(0,\Sigma_T)$ across i with $\Sigma_T = \sigma_0^2\frac{1}{T}\sum_{t=1}^{T} f\left(\frac{t}{T}\right)^2$. Since $Q_{i,T}\Rightarrow Q_i = \sigma_0\int_0^1 f(r)\,dW_i(r)$ and $E\left(Q_{i,T}^2\right) = \Sigma_T \to \Sigma = E\left(Q_i^2\right) = \sigma_0^2\int_0^1 f(r)^2\,dr > 0$ as $T\to\infty$ for fixed i, it follows that $Q_{i,T}^2$ are uniformly integrable in T. Then, by Theorem 3 in Phillips and Moon (1999), we have the desired result. ■

Part (b) By similar fashion, we let $Q_{i,T} = \frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{i,t-1} f\left(\frac{t}{T}\right)$ and $Q_i = \sigma_0\int_0^1 f(r)J_{c_0,i}(r)\,dr$. Then, we know that $Q_{i,T}\Rightarrow Q_i$ and

$$\begin{aligned}
&E\left(Q_{i,T}^2\right) = E\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{i,t-1} f\left(\frac{t}{T}\right)\right)^2 = \frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left(y_{i,t-1}y_{i,s-1}\right) f\left(\frac{t}{T}\right) f\left(\frac{s}{T}\right) \\
\to\ &\sigma_0^2\int_0^1\int_0^1\int_0^{r\wedge s} e^{c_0(r+s-2p)}dp f(r)f(s)dsdr = E\left(Q_i^2\right),
\end{aligned}$$

as $T \to \infty$ for all i. Therefore, $Q_{i,T}^2$ are uniformly integrable in T, and by Theorem 3 in Phillips and Moon (1999), we have the desired result. ■
Part (c) holds by the similar fashion, and we omit the proof. ■

Lemma 8 $\sqrt{N}\sqrt{T}\left(\hat{\delta}(c)-\delta_0\right) = O_p(1)$ *uniformly in* c *as* $(N,T \to \infty)$, *where* $\hat{\delta}(c)$ *is defined in* (4).

Proof. By definition,

$$\begin{aligned} & \sqrt{N}\sqrt{T}\left(\hat{\delta}(c)-\delta_0\right) \\ = & \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)\left(\varepsilon_{i,t}(\delta_0,c_0)-(c-c_0)\frac{y_{i,t-1}(\delta_0)}{T}\right)}{\frac{1}{T}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)^2}. \end{aligned}$$

Then, we have

$$\begin{aligned} & \sup_{c\in\mathbb{C}}\left|\sqrt{N}\sqrt{T}\left(\hat{\delta}(c)-\delta_0\right)\right| \\ \leq & \frac{\sup_{c\in\mathbb{C}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)\left(\varepsilon_{i,t}(\delta_0,c_0)-(c-c_0)\frac{y_{i,t-1}(\delta_0)}{T}\right)\right|}{\inf_{c\in\mathbb{C}}\frac{1}{T}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)^2}. \end{aligned}$$

First, note that

$$\begin{aligned} & \inf_{c\in\mathbb{C}}\frac{1}{T}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)^2 \\ = & 1-2\left(\inf_{c\in\mathbb{C}}c\right)\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}+\left(\inf_{c\in\mathbb{C}}c\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t-1}{T}\right)^2 \\ = & \frac{1}{T}\sum_{t=1}^{T}\left(\frac{t-1}{T}\right)^2\left(\inf_{c\in\mathbb{C}}c-\frac{\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}}{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t-1}{T}\right)^2}\right)^2+1-\frac{\left(\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t-1}{T}\right)^2} \\ \geq & 1-\frac{\left(\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t-1}{T}\right)^2}\to 1-\frac{3}{4}=\frac{1}{4}>0. \end{aligned}$$

Next, note that

$$\begin{aligned} & \sup_{c\in\mathbb{C}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)\left(\varepsilon_{i,t}(\delta_0,c_0)-(c-c_0)\frac{y_{i,t-1}(\delta_0)}{T}\right)\right| \\ \leq & \left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{i,t}(\delta_0,c_0)\right| \\ & +\sup_{c\in\mathbb{C}}|c|\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{t-1}{T}\varepsilon_{i,t}(\delta_0,c_0)\right| \end{aligned}$$

$$+\sup_{c\in\mathbb{C}}|c-c_0|\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y_{i,t-1}(\delta_0)\right|$$
$$+\sup_{c\in\mathbb{C}}|c(c-c_0)|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}\frac{t-1}{T}y_{i,t-1}(\delta_0).$$

Recall that $\sup_{c\in\mathbb{C}}|c|$ is finite. In view of Lemma 7(a) and (b), each term in the above display is $O_p(1)$ as $(N,T\to\infty)$. Therefore, we have $\sup_{c\in\mathbb{C}}\left|\sqrt{N}\sqrt{T}\left(\hat{\delta}(c)-\delta_0\right)\right|=O_p(1)$ as $(N,T\to\infty)$. ■

Derivation of (8)

Recall that

$$\begin{aligned}&l_{N,T}\left(c,\hat{\delta}(c),\hat{\sigma}^2(c);z^{N,T}\right)\\=&-\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{i,t}(\delta_0,c_0)-(c-c_0)\frac{y_{i,t-1}(\delta_0)}{T}-\left(\hat{\delta}(c)-\delta_0\right)\left(1-c\frac{t-1}{T}\right)\right)^2\\&+\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\varepsilon_{i,t}(\delta_0,c_0)-\left(\hat{\delta}(c_0)-\delta_0\right)\left(1-c_0\frac{t-1}{T}\right)\right\}^2.\end{aligned}$$

In view of (4), we have

$$\begin{aligned}&-\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\varepsilon_{i,t}(\delta_0,c_0)-(c-c_0)\frac{y_{i,t-1}(\delta_0)}{T}-\left(\hat{\delta}(c)-\delta_0\right)\left(1-c\frac{t-1}{T}\right)\right\}^2\\=&-\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}(\delta_0,c_0)^2+2T\left(\hat{\delta}(c)-\delta_0\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)^2\\&-\frac{1}{2}(c-c_0)^2\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-1}(\delta_0)^2}{T}\right)+(c-c_0)\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}(\delta_0)\varepsilon_{i,t}(\delta_0,c_0)\right),\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\varepsilon_{i,t}(\delta_0,c_0)-\left(\hat{\delta}(c_0)-\delta_0\right)\left(1-c_0\frac{t-1}{T}\right)\right\}^2\\=&\frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}(\delta_0,c_0)^2-\frac{1}{2}T\left(\hat{\delta}(c_0)-\delta_0\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(1-c_0\frac{t-1}{T}\right)^2,\end{aligned}$$

which yields

$$\begin{aligned}&l_{N,T}\left(c,\hat{\delta}(c),\hat{\sigma}^2(c);z^{N,T}\right)\\=&\frac{1}{2}T\left(\hat{\delta}(c)-\delta_0\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)^2-\frac{1}{2}(c-c_0)^2\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-1}(\delta_0)^2}{T}\right)\\&+(c-c_0)\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}(\delta_0)\varepsilon_{i,t}(\delta_0,c_0)\right)-\frac{1}{2}T\left(\hat{\delta}(c_0)-\delta_0\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(1-c_0\frac{t-1}{T}\right)^2.\end{aligned}$$

Note by Lemma 8 that $\sqrt{T}\left(\hat{\delta}(c)-\delta_0\right) = o_p(1)$ uniformly in c, and by Lemma 6(a) and Lemma 7(c) that $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-1}(\delta_0)^2}{T} \to_p \sigma_0^2\int_0^1\int_0^r e^{2c_0(r-s)}dsdr$ and $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}(\delta_0)\varepsilon_{i,t}(\delta_0,c_0) = o_p(1)$, respectively. Also recall that parameter set $\mathbb{C}$ is compact. Therefore, as $(N,T\to\infty)$,

$$\begin{aligned}
& l_{N,T}\left(c,\hat{\delta}(c),\hat{\sigma}^2(c);z^{N,T}\right) \\
= & -\frac{1}{2}(c-c_0)^2\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-1}(\delta_0)^2}{T}\right)+o_p(1) \\
\to & _p -\sigma_0^2\frac{1}{2}(c-c_0)^2\int_0^1\int_0^r e^{2c_0(r-s)}dsdr \text{ uniformly in } c.
\end{aligned}$$

Proof of Theorem 2

By the first order Taylor expansion of the first order condition (9) around the true parameter c_0, we have

$$0=\frac{dL(c,\hat{\delta}(c),\hat{\sigma}^2(c);z^{N,T})}{dc}+\left\{\frac{d^2L(c^*,\hat{\delta}(c^*),\hat{\sigma}^2(c^*);z^{N,T})}{dc^2}\right\}(\hat{c}-c_0),$$

where c^* lies between c_0 and $\hat{c}$. From this, we write

$$\sqrt{N}(\hat{c}-c_0)=-\left(\frac{d^2L(c^*,\hat{\delta}(c^*),\hat{\sigma}^2(c^*);z^{N,T})}{dc^2}\right)^{-1}\left(\sqrt{N}\frac{dL(c_0,\hat{\delta}(c_0),\hat{\sigma}^2(c_0);z^{N,T})}{dc}\right). \quad (22)$$

Define

$$Q_{N,T}\left(c,\hat{\delta}(c)\right)=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left(\Delta_c z_{i,t}-\hat{\delta}(c)\left(1-c\frac{t-1}{T}\right)\right)^2.$$

Then, (22) is written as

$$\begin{aligned}
& \sqrt{N}(\hat{c}-c_0) \\
= & -\left(\frac{\frac{d^2Q_{N,T}\left(c^*,\hat{\delta}(c^*)\right)}{d^2c}}{Q_{N,T}\left(c^*,\hat{\delta}(c^*)\right)}-\left(\frac{\frac{dQ_{N,T}\left(c^*,\hat{\delta}(c^*)\right)}{dc}}{Q_{N,T}\left(c^*,\hat{\delta}(c^*)\right)}\right)^2\right)^{-1}\left(\frac{\sqrt{N}\frac{dQ_{N,T}\left(c_0,\hat{\delta}(c_0)\right)}{dc}}{Q_{N,T}\left(c_0,\hat{\delta}(c_0)\right)}\right) \quad (23)
\end{aligned}$$

Note that

$$\begin{aligned}
& \sqrt{N}\frac{dQ_{N,T}\left(c_0,\hat{\delta}(c_0)\right)}{dc} \\
= & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}(\delta_0)\varepsilon_{i,t}(\delta_0,c_0) \\
& -\sqrt{N}\sqrt{T}\left(\hat{\delta}(c_0)-\delta_0\right)\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y_{i,t-1}\left(1-c_0\frac{t-1}{T}\right)
\end{aligned}$$

$$
\begin{aligned}
& -\sqrt{N}\sqrt{T}\left(\hat{\delta}\left(c_0\right)-\delta_0\right)\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{i,t}\frac{t-1}{T} \\
& +NT\left(\hat{\delta}\left(c_0\right)-\delta_0\right)^2\frac{1}{\sqrt{N}T}\sum_{t=1}^{T}\frac{t-1}{T}\left(1-c_0\frac{t-1}{T}\right) \\
= \; & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}\left(\delta_0\right)\varepsilon_{i,t}\left(\delta_0,c_0\right)+o_p\left(1\right),
\end{aligned}
$$

where the last line holds because

$$
\begin{aligned}
\sqrt{N}\sqrt{T}\left(\hat{\delta}\left(c_0\right)-\delta_0\right) &= O_p\left(1\right) \text{ by Lemma 8,} \\
\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y_{i,t-1}\left(1-c_0\frac{t-1}{T}\right) &= o_p\left(1\right) \text{ by Lemma 7(b),} \\
\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{i,t}\frac{t-1}{T} &= o_p\left(1\right) \text{ by Lemma 7(a),} \\
\text{and } \frac{1}{\sqrt{N}T}\sum_{t=1}^{T}\frac{t-1}{T}\left(1-c_0\frac{t-1}{T}\right) &= O\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}
$$

>From Lemma 7(c), as $(N,T\to\infty)$ we have

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}\left(\delta_0\right)\varepsilon_{i,t}\left(\delta_0,c_0\right)\Rightarrow N\left(0,\sigma_0^2\int_0^1\int_0^r e^{2c_0(r-s)}dsdr\right),
$$

and so

$$
\sqrt{N}\frac{dQ_{N,T}\left(c_0,\hat{\delta}\left(c_0\right)\right)}{dc}\Rightarrow N\left(0,\sigma_0^2\int_0^1\int_0^r e^{2c_0(r-s)}dsdr\right). \tag{24}
$$

Also, since $\hat{\delta}\left(c_0\right)$ is consistent for δ_0, we have

$$
Q_{N,T}\left(c,\hat{\delta}\left(c\right)\right)=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left(\Delta_c z_{i,t}-\hat{\delta}\left(c\right)\left(1-c\frac{t-1}{T}\right)\right)^2\to_p\sigma_0^2. \tag{25}
$$

Combining (24) and (25), as $(N,T\to\infty)$, we have

$$
\frac{\sqrt{N}\frac{dQ_{N,T}\left(c_0,\hat{\delta}\left(c_0\right)\right)}{dc}}{Q_{N,T}\left(c_0,\hat{\delta}\left(c_0\right)\right)}\Rightarrow N\left(0,\frac{1}{\sigma_0^2}\int_0^1\int_0^r e^{2c_0(r-s)}dsdr\right). \tag{26}
$$

Next, by the envelope function theorem and the chain rule, it follows that

$$
\begin{aligned}
& \frac{d^2Q_{N,T}\left(c,\hat{\delta}\left(c\right)\right)}{d^2c} \\
= \; & \frac{\partial^2Q_{N,T}\left(c,\hat{\delta}\left(c\right)\right)}{\partial c^2}+\frac{\partial^2Q_{N,T}\left(c,\hat{\delta}\left(c\right)\right)}{\partial\delta\partial c}\frac{d\hat{\delta}\left(c\right)}{dc}.
\end{aligned} \tag{27}
$$

A short calculation yields

$$
\begin{aligned}
& -\frac{\partial^2 Q_{N,T}\left(c^*, \hat{\delta}\left(c^*\right)\right)}{\partial c^2} \\
= & \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T} y_{i,t-1}^2\left(\delta_0\right) \\
& -2\sqrt{T}\left(\hat{\delta}\left(c^*\right)-\delta_0\right)\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{i,t-1}\left(\delta_0\right)\frac{t-1}{T} \\
& +T\left(\hat{\delta}\left(c^*\right)-\delta_0\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t-1}{T}\right)^2.
\end{aligned}
$$

Since $\sqrt{T}\left(\hat{\delta}(c)-\delta_0\right)=o_p(1)$ uniformly in c by Lemma 8 and $c^* \to_p c_0$, it follows that $\sqrt{T}\left(\hat{\delta}\left(c^*\right)-\delta_0\right)=o_p(1)$ as $(N,T\to\infty)$. Also, we know $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{i,t-1}\left(\delta_0\right)\frac{t-1}{T}=o_p(1)$ as $(N,T\to\infty)$. From these and Lemma 6(a), we have

$$
\begin{aligned}
& -\frac{\partial^2 Q_{N,T}\left(c^*, \hat{\delta}\left(c^*\right)\right)}{\partial c^2} \\
= & \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T} y_{i,t-1}^2\left(\delta_0\right)+o_p(1) \to_p \sigma_0^2\int_0^1\int_0^r e^{2c_0(r-s)}dsdr. \qquad (28)
\end{aligned}
$$

By similar fashion, using the facts that $\sqrt{T}\left(\hat{\delta}\left(c^*\right)-\delta_0\right)=o_p(1)$ and $c^* \to_p c_0$, and the results in Lemmas 6 and 7, it is not difficult to show that

$$
\frac{1}{\sqrt{T}}\frac{\partial^2 Q_{N,T}\left(c^*, \hat{\delta}\left(c^*\right)\right)}{\partial\delta\partial c}, \sqrt{T}\frac{d\hat{\delta}\left(c^*\right)}{dc}=o_p(1). \qquad (29)
$$

Since

$$
Q_{N,T}\left(c^*, \hat{\delta}\left(c^*\right)\right) \to_p \sigma_0^2,
$$

the first term in the numerator of (23)

$$
\frac{\frac{d^2 Q_{N,T}\left(c^*, \hat{\delta}\left(c^*\right)\right)}{d^2 c}}{Q_{N,T}\left(c^*, \hat{\delta}\left(c^*\right)\right)} \to_p \int_0^1\int_0^r e^{2c_0(r-s)}dsdr. \qquad (30)
$$

For limit of the second term in the numerator of (23), notice that

$$
\frac{dQ_{N,T}\left(c^*, \hat{\delta}\left(c^*\right)\right)}{dc}=\frac{\partial Q_{N,T}\left(c^*, \hat{\delta}\left(c^*\right)\right)}{\partial c} \to_p 0
$$

as $(N,T\to\infty)$ since $c^*\to_p c_0$ and $\frac{\partial Q_{N,T}\left(c,\hat{\delta}(c)\right)}{\partial c}=O_p\left(\frac{1}{\sqrt{N}}\right)$ uniformly in c. Therefore,

$$\left(\frac{\frac{dQ_{N,T}\left(c^*,\hat{\delta}(c^*)\right)}{dc}}{Q_{N,T}\left(c^*,\hat{\delta}\left(c^*\right)\right)}\right)\to_p 0. \tag{31}$$

>From (30) and (31) we have

$$-\left(\frac{d^2L(z^{N,T};\hat{\delta}\left(c^*\right),c^*)}{dc^2}\right)\to_p\int_0^1\int_0^r e^{2c_0(r-s)}dsdr,$$

and combining this with (24), we have the desired result,

$$\sqrt{N}\left(\hat{c}-c_0\right)\Rightarrow N\left(0,\frac{1}{\sigma_0^2\int_0^1\int_0^r e^{2c_0(r-s)}dsdr}\right).\ \blacksquare$$

Proof of (14)

By definition, we have

$$\begin{aligned}&\sqrt{N}\left(\tilde{c}-c_0\right)\\=\;&\frac{\frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T\left\{\left(\Delta z_{i,t}-\hat{\delta}\left(c\right)\right)\left(\frac{z_{i,t-1}}{T}-\hat{\delta}\left(c\right)\frac{t-1}{T}\right)-c_0\left(\frac{z_{i,t-1}}{T}-\hat{\delta}\left(c\right)\frac{t-1}{T}\right)^2\right\}}{\frac{1}{N}\sum_{i=1}^N\sum_{t=1}^T\left(\frac{z_{i,t-1}}{T}-\hat{\delta}\left(c\right)\frac{t-1}{T}\right)^2}.\end{aligned}$$

First, note that

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^N\sum_{t=1}^T\left(\frac{z_{i,t-1}}{T}-\hat{\delta}\left(c\right)\frac{t-1}{T}\right)^2\\=\;&\frac{1}{N}\sum_{i=1}^N\frac{1}{T^2}\sum_{t=1}^T y_{i,t-1}\left(\delta_0\right)^2-2\sqrt{T}\left(\hat{\delta}\left(c\right)-\delta_0\right)\frac{1}{N}\sum_{i=1}^N\frac{1}{T\sqrt{T}}\sum_{t=1}^T y_{i,t-1}\left(\delta_0\right)\frac{t-1}{T}\\&+T\left(\hat{\delta}\left(c\right)-\delta_0\right)^2\frac{1}{T}\sum_{t=1}^T\left(\frac{t-1}{T}\right)^2\\=\;&\frac{1}{N}\sum_{i=1}^N\frac{1}{T^2}\sum_{t=1}^T y_{i,t-1}\left(\delta_0\right)^2+o_p\left(1\right)\\\to\;&{}_p\sigma_0^2\int_0^1\int_0^r e^{2c_0(r-s)}dsdr.\end{aligned}$$

where the last equality holds because $\sqrt{T}\left(\hat{\delta}\left(c\right)-\delta_0\right)=o_p\left(1\right)$ and $\frac{1}{N}\sum_{i=1}^N\frac{1}{T\sqrt{T}}\sum_{t=1}^T y_{i,t-1}\left(\delta_0\right)\frac{t-1}{T}=o_p\left(1\right)$.

Next, for the numerator, we write

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\Delta z_{i,t}-\hat{\delta}\left(c\right)\right)\left(\frac{z_{i,t-1}}{T}-\hat{\delta}\left(c\right)\frac{t-1}{T}\right)-c_0\left(\frac{z_{i,t-1}}{T}-\hat{\delta}\left(c\right)\frac{t-1}{T}\right)^2\right\}\\
=\;&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\Delta z_{i,t}-\delta_0\right)\left(\frac{z_{i,t-1}}{T}-\delta_0\frac{t-1}{T}\right)-c_0\left(\frac{z_{i,t-1}}{T}-\delta_0\frac{t-1}{T}\right)^2\right\}\\
&-\sqrt{T}\left(\hat{\delta}\left(c\right)-\delta_0\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y_{i,t-1}\left(\delta_0\right)\\
&-\sqrt{T}\left(\hat{\delta}\left(c\right)-\delta_0\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(c_0\frac{y_{i,t}\left(\delta_0\right)}{T}+\varepsilon_{i,t}\left(\delta_0,c_0\right)\right)\frac{t-1}{T}\\
&+\sqrt{N}T\left(\hat{\delta}\left(c\right)-\delta_0\right)^2\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\\
&+2c_0\sqrt{T}\left(\hat{\delta}\left(c\right)-\delta_0\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y_{i,t-1}\left(\delta_0\right)\frac{t-1}{T}\\
&-c_0\sqrt{N}T\left(\hat{\delta}\left(c\right)-\delta_0\right)^2\frac{1}{T}\sum_{t=1}^{T}\frac{t-1}{T}\\
=\;&I+II+III+IV+V+VI,\ \text{say.}
\end{aligned}
$$

Recall that $\sqrt{N}\sqrt{T}\left(\hat{\delta}\left(c\right)-\delta_0\right)=O_p\left(1\right)$ by Lemma8. Then, in view of Lemmas 6 and 7, it is not difficult to find that $II,III,IV,V,VI=o_p\left(1\right)$. For I, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\Delta z_{i,t}-\delta_0\right)\left(\frac{z_{i,t-1}}{T}-\delta_0\frac{t-1}{T}\right)-c_0\left(\frac{z_{i,t-1}}{T}-\delta_0\frac{t-1}{T}\right)^2\right\}\\
=\;&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\Delta y_{i,t}\left(\delta_0\right)\frac{y_{i,t-1}\left(\delta_0\right)}{T}-c_0\frac{y_{i,t-1}\left(\delta_0\right)^2}{T^2}\right\}\\
=\;&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}\left(\delta_0\right)\varepsilon_{i,t}\left(\delta_0,c_0\right)\\
\Rightarrow\;&N\left(0,\sigma_0^2\int_0^1\int_0^r e^{2c_0(r-s)}dsdr\right)
\end{aligned}
$$

as $\left(N,T\rightarrow\infty\right)$. Then we have the desired result. ■

Proof of Lemma 4

In view of

$$
\begin{aligned}
\sqrt{T}\left(\hat{\delta}_i\left(c\right)-\delta_{i,0}\right)=&\left(\frac{1}{T}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)^2\right)^{-1}\left\{\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)\frac{\varepsilon_{i,t}\left(\delta_{0,i},c_0\right)}{\sqrt{T}}\right.\\
&\left.-\left(c-c_0\right)\frac{1}{T}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)\frac{y_{i,t-1}\left(\delta_{0,i}\right)}{\sqrt{T}}\right\},
\end{aligned}
\tag{32}
$$

we write (17) as

$$
\begin{aligned}
& L(z^{N,T};\hat{\delta}^{N}(c),c)-L(z^{N,T};\hat{\delta}^{N}(c_0),c_0) \\
= & \frac{1}{2N}\sum_{i=1}^{N}T\left(\hat{\delta}_i(c)-\delta_{0,i}\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)^2-\frac{1}{2}(c-c_0)^2\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-1}(\delta_{0,i})^2}{T}\right) \\
& +(c-c_0)\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}(\delta_{0,i})\varepsilon_{i,t}(\delta_{0,i},c_0)\right) \\
& -\frac{1}{2N}\sum_{i=1}^{N}T\left(\hat{\delta}_i(c_0)-\delta_{0,i}\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(1-c_0\frac{t-1}{T}\right)^2, \\
= & I+II+III+IV, \text{ say.}
\end{aligned}
$$

Now we find limits of I, II, III, and IV. In view of (32) and by Lemma 6, and Assumption 2, we have

$$
\begin{aligned}
I = & \left(\frac{1}{T}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)^2\right)^{-1} \\
& \times\frac{1}{2N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)\frac{\varepsilon_{i,t}(\delta_{0,i},c_0)}{\sqrt{T}}-(c-c_0)\frac{1}{T}\sum_{t=1}^{T}\left(1-c\frac{t-1}{T}\right)\frac{y_{i,t-1}(\delta_{0,i})}{\sqrt{T}}\right)^2 \\
& \underset{p}{\rightarrow}\frac{1}{2}\left(\int_0^1(1-cr)^2dr\right)^{-1} \\
& \times\left(\begin{array}{c}\int_0^1(1-cr)^2dr-2(c-c_0)\int_0^1\int_0^s e^{c_0(s-r)}(1-cr)(1-cs)\,drds \\ +(c-c_0)^2\int_0^1\int_0^1(1-cr)(1-cs)\int_0^{r\wedge s}e^{c_0(r+s-p)}dpdsdr\end{array}\right)
\end{aligned} \tag{33}
$$

uniformly in c as $(N,T\rightarrow\infty)$. Similarly, using Lemma 6(a) and Lemma 7(c), we can show that, as $(N,T\rightarrow\infty)$

$$
\begin{aligned}
II = & -\frac{1}{2}(c-c_0)^2\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\frac{y_{i,t-1}(\delta_{0,i})^2}{T}\right) \\
& \underset{p}{\rightarrow}-\frac{1}{2}(c-c_0)^2\int_0^1\int_0^r e^{2c_0(r-s)}dsdr \text{ uniformly in } c,
\end{aligned} \tag{34}
$$

and

$$
III=(c-c_0)\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}(\delta_{0,i})\varepsilon_{i,t}(\delta_{0,i},c_0)\right)\underset{p}{\rightarrow}0 \text{ uniformly in } c. \tag{35}
$$

Also, it is not difficult to derive that as $(N,T\rightarrow\infty)$

$$
IV=-\frac{1}{2N}\sum_{i=1}^{N}T\left(\hat{\delta}_i(c_0)-\delta_{0,i}\right)^2\frac{1}{T}\sum_{t=1}^{T}\left(1-c_0\frac{t-1}{T}\right)^2\underset{p}{\rightarrow}-\frac{1}{2}. \tag{36}
$$

Combining (33) – (36), we finally have the desired result.

References

[1] Alvarez, J. and M. Arellano (1998). "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators", Mimeo.

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

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

[4] Canjels, N. And M. Watson (1995). "Estimating deterministic trends in the presence of serially correlated errors," forthcoming in *Review of Economics and Statistics*.

[5] Cavanagh, C. (1985). "Roots Local to Unity", manuscript, Department of Economics, Harvard University.

[6] Cavanagh, C., G. Elliott, and J.H. Stock (1995). "Inference in Models with Nearly Integrated Regressors", *Econometric Theory,* 11, 1131–1147.

[7] Elliott, G., T.J. Rothenberg and J.H. Stock (1995). "Efficient tests of an autoregressive unit root," *Econometrica,* Vol. 64, 813 - 836.

[8] Hahn, J. (1998). "Asymptotically Unbiased Inference of Dynamic Panel Model with Fixed Effects When Both n and T are large", Mimeo.

[9] Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators," *Econometrica* 50, 1029–1054.

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

[11] Moon, H. R.and P.C.B. Phillips (1998). "Consistent Estimation of Autoregressive Roots near Unity using Panel Data" Yale University, Mimeographed.

[12] Newey, W. and D. MacFadden (1994). "Large Sample Estimation and Hypothesis Testing", Chapter 36, *Handbook of Econometrics Vol 4*, North Holland: Amsterdam.

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

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

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

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

[17] Phillips, P.C.B. and H.R. Moon (1999). "Linear Regression Limit Theory for Nonstationary Panel Data", forthcoming in *Econometrica.*

[18] Phillips, P.C.B., H. R. Moon, and Z. Xiao (1998). "How to Estimate Autoregressive Roots Near Unity", Yale University, Mimeographed

[19] Phillips, P.C.B., and Zhijie. Xiao (1998). "A Primer on Unit Root Testing", *Journal of Economic Surveys*, Vol. 12, No. 5, December, 1998, pp. 423-469.

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

[21] Xiao, Zhijie and Peter C.B. Phillips (1999). "Efficient Detrending in Cointegrating Regression," forthcoming in *Econometric Theory.*