Likelihood Based Inference for Dynamic Panel Data Models

Seung C. Ahn[a]
Arizona State University

Gareth M. Thomas
Quantitative Micro Software[b]

Abstract

This paper considers maximum likelihood (ML) based inferences for dynamic panel data models. We focus on the analysis of the panel data with a large number (N) of cross-sectional units and a small number (T) of repeated time-series observations for each cross-sectional unit. We examine several different ML estimators and their asymptotic and finite-sample properties. Our major finding is that when data follow unit-root processes without or with drifts, the ML estimators have singular information matrices. This is a case of Sargan (1983) in which the first order condition for identification fails, but parameters are identified. The ML estimators are consistent, but they have nonstandard asymptotic distributions and their convergence rates are lower than $N^{1/2}$. In addition, the sizes of usual Wald statistics based on the estimators are distorted even asymptotically, and they reject the unit-root hypothesis too often. However, following Rotnitzky, Cox, Bottai and Robins (2000), we show that likelihood ratio (LR) tests for unit root follow mixtures of chi-square distributions. Our Monte Carlo experiments show that the LR tests with the p-values from the mixed distributions are much better sized than the Wald tests, although they tend to slightly over-reject the unit root hypothesis in small samples. It is also shown that the LR tests for unit roots have good finite-sample power properties.

***JEL Classification Codes:* C23, C40**
Keywords: dynamic panel data, maximum likelihood, singular information matrix.

Acknowledgment
The first author gratefully acknowledges the financial support of the College of Business and Dean's Council of 100 at Arizona State University, the Economic Club of Phoenix, and the alumni of the College of Business. We would like to thank participants in the econometrics seminar at Rice University, the Eighth Annual Texas Econometrics Camp, the University of California at San Diego, Pennsylvania State University, and the University of California at Los Angeles.

---

[a] Corresponding Author: Seung C. Ahn, Department of Economics, W.P. Carey School of Business, Arizona State University, Tempe, AZ 85287; email: miniahn@asu.edu.
[b] The views and research contained in this paper represent those of the individual authors, and not of Quantitative Micro Software.

1. Introduction

Panel data models assume that individual cross-section units have different intercept terms due to unobservable heterogeneity. Two different models are used to control the unobservable heterogeneity. One is the random effects (RE) model in which the individual-specific intercept terms, or individual effects, are treated as random variables. The other is the fixed effects (FE) model in which the effects are treated as parameters. The FE model is more general than the RE model in that it requires weaker distributional assumptions about the effects. One difficulty in estimating the fixed effects models however is that the number of parameters increases with N. A traditional treatment for this so-called "incidental parameters problem" (Neyman and Scott, 1948) is the within estimator, i.e., least squares on data transformed into deviations from individual means, which is also a ML estimator. For models with strictly exogenous regressors, the within estimator is consistent. Unfortunately, however, when T is small, the within estimator is inconsistent for the dynamic models that use lagged dependent variables as regressors (Nickell, 1981). One way to avoid this problem is to use the random effects ML estimator that treats the effects as time invariant random variables (Hsiao, 1986). Instead, generalized method of moments (GMM) estimators have been widely used to analyze the FE dynamic panel models (e.g., Arellano and Bond, 1991; Arellano and Bover, 1995; Ahn and Schmidt, 1995, 1997; and Blundell and Bond, 1998). An important reason for the popularity of GMM is that it provides consistent estimators under quite general FE assumptions.

In the early 2000's, research interests in the ML estimation of dynamic panel data models have been revived. For example, Lancaster (2002), and Hsiao, Pesaran and Tahmiscioglu (2002, HPT) have developed alternative ML estimators for FE dynamic models.[1] Kruiniger (2002) provides the general conditions under which the ML estimators of Lancaster and HPT are consistent. These studies consider the models in which white noise error terms are homoskedastic over time. Extending these studies, Alvarez and Arellano (2004) examine the RE and the two FE ML estimators when white-noise error terms are heteroskedastic over time.

One reason for the recent revival of the ML approach may be that the panel data GMM estimators often have poor finite-sample properties (e.g., Bond and Windmeijer, 2002). Dynamic panel data models imply a large number of moment conditions. The GMM estimators

[1]Hahn and Kuersteiner (2002) also consider the ML estimator for the FE dynamic model with large N and T. They find the ML estimator, which is the within estimator, is consistent, but it is asymptotically biased. Hahn and Kuersteiner provide a biased-corrected ML estimator.

imposing all of the available moment conditions appear to generate biased statistical inferences, especially when T is large. Thus, an important research agenda would be to develop the alternative GMM estimators that use a smaller number of moment conditions, but without substantial loss of asymptotic efficiency. Ahn and Schmidt (1995) show that when data are normally distributed, an efficient GMM estimator is asymptotically identical to the ML estimator constructed on the same first and second order moment conditions. Given that ML estimators are the GMM estimators based on exact identifying moment conditions, the ML-based approach may be a viable alternative to the popular GMM approach.

This paper considers the asymptotic and finite-sample properties of the RE ML estimator and the two FE ML estimators of Lancaster (2002) and HPT when white noise error terms are homoskedastic over time. In addition, we investigate the asymptotic properties of the estimators when data contain unit roots. Our results of efficiency comparisons are not new. For the same dynamic models as we consider, Kruiniger (2002) has studied the relative asymptotic efficiency of the three estimators, but in a less intuitive manner. Alvarez and Arellano (2004) compare the efficiency comparison of the three estimators for general models. Their results are more general than those in Kruiniger (2002) and ours, because their models are the autoregressive models of higher order with heteroskedastic noise errors. In particular, Alvarez and Arellano propose an alternative parameterization for the RE ML estimator that leads to an easy comparison between the RE and other estimators. In this paper, we use the same parameterization that we have developed independently. More intuitive efficiency comparisons are made because we focus on the autoregressive models of order one with homoskedastic errors. Both the approaches of Kruiniger and Alvarez and Arellano are limited to the cases in which data are stationary.

A novel finding in this paper is that when cross-sectional data follow unit root processes, the information matrices of the RE and the two FE ML estimators become singular. However, these estimators still can identify the parameters to be estimated. This is a case similar to the case of Sargan (1983) in which the first order condition for identification fails, but the parameters are identified. Rotnitzky, Cox, Bottai and Robins (2000, RCBR) analyze the general asymptotic properties of ML estimators for such cases. Following their approach, we derive the asymptotic distribution and convergence rate of the RE and HPT ML estimators. Their asymptotic distributions are not normal and their convergence rates are $N^{1/4}$, not $N^{1/2}$. In addition, usual Wald-type tests for unit roots generate biased inferences. In contrast, likelihood ratio (LR) test statistics for unit root follow mixed χ^2 distributions that can be easily simulated by Monte Carlo experiments. We find that the LR tests with the p-values from the mixed distributions are much better sized than other Wald-type tests.

We also consider a dynamic panel model with heterogeneous trends. For the model with large N and T, Moon and Phillips (2004) propose a GMM estimator that is obtained treating individual trends as incidental parameters (Neyman and Scott, 1948). The convergence rate of the estimator is $N^{1/6}$ when data follow unit roots with heterogeneous drifts. We consider the ML estimation of the model under the alternative assumption that the individual trends are random. For fixed T, we find that the convergence rate of the alternative ML estimator is $N^{1/4}$.

This paper is organized as follows. In Section 2, we briefly compare the RE, HPT and Lancaster estimators and examine their relative efficiency. Section 3 investigates the asymptotic properties of these estimators when data follow unit-root processes. Section 4 reports our Monte Carlo experiment results. Some concluding remarks follow in Section 5.

2. ML Estimation

The foundation of this paper is the simple dynamic model

$$y_{it} = \delta y_{i,t-1} + (\alpha_i + \varepsilon_{it}). \tag{1}$$

Here i = 1, 2, ..., N denotes cross-section unit (individual) and t = 1, 2, ..., T denotes time. Our parameter of interest is δ. We initially assume that δ is within a unit circle, *i.e.*, $|\delta| < 1$. We will relax this assumption later. The composite error $(\alpha_i + \varepsilon_{it})$ contains a time invariant individual effect α_i and random noise ε_{it}. The initial observed value of y for individual i is y_{i0}. We assume that the random error vector $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})'$ is uncorrelated with y_{i0} and α_i, and that the ε_i, y_{i0}, and α_i are cross-sectionally independent. We assume that the error terms ε_{it} are serially uncorrelated. For the maximum likelihood estimation of model (1), we need to make distributional assumptions about the y_{i0} and α_i. We assume that all of the ε_i, y_{i0}, and α_i are normally distributed:

$$\begin{pmatrix} y_{i0} \\ \alpha_i \\ \varepsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0_{T\times 1} \end{pmatrix}, \begin{pmatrix} \sigma_{y_o}^2 & \sigma_{y_o,\alpha} & 0_{1\times T} \\ \sigma_{y_o,\alpha} & \sigma_\alpha^2 & 0_{1\times T} \\ 0_{T\times 1} & 0_{T\times 1} & \nu I_T \end{pmatrix} \right), \tag{2}$$

where $\nu = \text{var}(\varepsilon_{it})$. Here, we assume that the y_{i0} and α_i have zero means. This assumption is just for convenience. We can allow nonzero means without altering our main results. Since we make an explicit normality assumption about the individual effects α_i, we shall refer to (2) as the random-effect (RE) assumption.

In this paper, we do not consider the more general models that contain exogenous regressors. However, our results can be easily extended to such models. The ML estimation of model (1) has been considered by Anderson and Hsiao (1981), Bhargava and Sargan (1983), and Hsiao (1986). We can easily derive the log-likelihood function, viewing (1) as a recursive simultaneous equations model (treating y_{i0}, y_{i1}, ... , y_{iT} as endogenous variables).

The log-likelihood function for model (1) depends on the five parameters, δ, $\sigma_{y_o,\alpha}$, $\sigma_{y_o}^2$, σ_α^2 and ν. However, we find a convenient reparameterization by which the parameter $\sigma_{y_o}^2$ can be orthogonalized to other parameters. Using this method, the parameter of our interest δ can be estimated independently from $\sigma_{y_o}^2$. In addition, this reparameterization will facilitate our comparisons of the RE ML estimator with the two alternative fixed-effects (FE) ML estimators developed by Hsiao, Pesaran and Tahmiscioglu (2002, HPT), and Lancaster (2002).

We define

$$p_i \equiv (\delta - 1)y_{i0} + \alpha_i \,; E(p_i \mid y_{i0}) = \psi y_{i0} \,; \;\; u_i = \left(p_i - \psi y_{i0}\right) + \varepsilon_{i1} \,,$$

where $\psi = (\delta - 1) + \sigma_{y_o,\alpha} / \sigma_{y_o}^2$.[2] With this notation, model (1) can be written in the following alternative form:

$$\begin{pmatrix} \Delta y_{i1} \\ \Delta y_i \end{pmatrix} = \begin{pmatrix} y_{i0} \\ 0_{(T-1)\times 1} \end{pmatrix} \psi + \begin{pmatrix} 0 \\ \Delta y_{i,-1} \end{pmatrix} \delta + \begin{pmatrix} u_i \\ \Delta \varepsilon_i \end{pmatrix}, \tag{3}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$, and

$$\Delta y_i = (\Delta y_{i2}, ..., \Delta y_{iT})' \,; \;\; \Delta y_{i,-1} = (\Delta y_{i1}, ..., \Delta y_{i,T-1})' \,; \;\; \Delta \varepsilon_i = (\Delta \varepsilon_{i2}, ..., \Delta \varepsilon_{iT})' \,.$$

Under the RE assumption, the variance matrix of the error vector $(u_i, \Delta\varepsilon_i')'$ is given:

$$Var\begin{pmatrix} u_i \\ \Delta \varepsilon_i \end{pmatrix} \equiv \nu \Omega_T(\omega) \equiv \nu \begin{pmatrix} \omega + 1 & -c_{T-1}' \\ -c_{T-1} & B_{T-1} \end{pmatrix}, \tag{4}$$

[2]If the exogenous regressors, say x_{i1}, ... , x_{iT}, were included in the model, we may construct the likelihood function assuming $E(\alpha_i \mid y_{i0}, x_{i1}, ..., x_{iT})$ is linear in the conditional variables (Wooldridge, 2000).

where $\omega = \text{var}(p_i - \psi y_{i0})/\nu = (\sigma_\alpha^2 - \sigma_{y_o,\alpha}^2/\sigma_{y_o}^2)/\nu$, c_{T-1} is a (T-1)×1 vector whose first entry equals one while other entries equal zero, and $B_{T-1} = [B_{T,jh}]$ is a (T-1)×(T-1) matrix such that $B_{T-1,jj} = 2$, $B_{T-1,j,j+1} = B_{T-1,j-1,j} = -1$, and all other entries equal zero. We can show that

$$[\Omega_T(\omega)]^{-1} = \begin{pmatrix} 0 & 0_{1\times(T-1)} \\ 0_{(T-1)\times 1} & B_{T-1}^{-1} \end{pmatrix} + \frac{T}{\xi_T}\begin{pmatrix} 1 \\ k_{T-1} \end{pmatrix}\begin{pmatrix} 1 & k'_{T-1} \end{pmatrix}, \tag{5}$$

where $k_{T-1} = ((T-1)/T, (T-2)/T, ..., 1/T)'$, and $\xi_T = \det[\Omega_T(\omega)] = T\omega + 1$.[3]

In the reparameterized model (3), the parameter vector to be estimated jointly is given by $\theta = (\delta, \nu, \omega, \psi)'$. The initial observation y_{i0} is uncorrelated with the error vector $(u_i, \Delta\varepsilon_i')'$. So, we can construct a likelihood function treating y_{i0} as predetermined. This is so because under our normality assumption,

$$f(r_i, y_{i0} \mid \theta, \sigma_{y_0}^2) = f(r_i \mid y_{i0}, \theta) f(y_{i0} \mid \sigma_{y_0}^2), \tag{6}$$

where $r_i = (\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT})'$ and "f" denotes a normal probability density function. Under the RE assumption, and using (4) and (5), we can easily derive the log-density function of r_i conditional on y_{i0}.[4,5]

$$\begin{aligned} \ell_{RE,i}(\theta) &\equiv \ln f(r_i \mid \theta) \\ &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\nu) - \frac{1}{2}\ln(\xi_T) - \frac{1}{2\nu}\left(\Delta y_i - \delta\Delta y_{i,-1}\right)' B_{T-1}^{-1}\left(\Delta y_i - \delta\Delta y_{i,-1}\right) \\ &\quad - \frac{T}{2\nu\xi_T}\left((\Delta y_{i1} - \psi y_{i0}) + k_{T-1}{}'(\Delta y_i - \delta\Delta y_{i,-1})\right)^2. \end{aligned}$$

Thus, maximizing the likelihood function $\Sigma_{i=1}^N \ell_{RE,i}(\theta)$, we can obtain the conditional RE ML estimator of θ, which is equivalent to the unconditional RE ML estimator based on the unconditional density $f(r_i, y_{i0} \mid \theta, \sigma_{y_0}^2)$. We obtain the equality (6) from the assumption that the initial values y_{i0} are normally distributed. However, it may be worth noting that the normality of y_{i0} is not required. The equality (6) holds as long as the ε_i and α_i are normal conditional

[3]See HPT for the derivation of $\det[\Omega_T(\omega)]$. The parameter ω in their paper is equivalent to $(\omega+1)$ in this paper.

[4]For the cases in which the y_{i0} and α_i do have non-zero expectations, the term $(\Delta y_{i1} - \psi y_{i0})$ will be replaced by $(\Delta y_{i1} - \psi\Delta y_{i0} - a)$, where a is a constant.

[5]We can estimate ξ_T instead of ω. In simulations, we found that the ML algorithms converge faster when ξ_T is estimated. Accordingly, we have estimated ξ_T for our simulations.

on y_{i0} and the conditional mean of α_i is linear in y_{i0}

There is an important case in which the equivalence of the unconditional and conditional ML estimators of θ breaks down. Suppose that the y_{it} follow a stationary process that has considered by Arellano and Bover (1995) and Ahn and Schmidt (1995, 1997):

$$y_{i0} = \eta_i + q_{i0}; \quad y_{it} = \eta_i + q_{it}; \quad q_{it} = \delta q_{i,t-1} + \varepsilon_{it}, \quad t = 1,...,T, \tag{7}$$

where $\varepsilon_i \sim N(0_{T\times 1}, \nu I_T)$, η_i and q_{i0} are also normally distributed with zero means, and all of ε_i, η_i and q_{i0} are mutually independent. This process implies that $y_{it} = \delta y_{i,t-1} + (1-\delta)\eta_i + \varepsilon_{it}$, where $(1-\delta)\eta_i$ equals α_i in our notation. Let $E(\eta_i \mid y_{i0}) = \tau y_{i0}$. Then, the data generation process (7) appears to be the same as the process given by (1) and (2), if we set:

$$\alpha_i = (1-\delta)\eta_i; \quad \psi = (1-\delta)\tau .$$

However, there exists an important difference between the two data generating processes. In the model given by (1) and (2), we do not assume that the initial values y_{i0} are distributed around the effect α_i. Thus, we do not impose any restriction on $\sigma_{y_o,\alpha} = \text{cov}(y_{i0}, \alpha_i)$. Without any restriction on $\sigma_{y_o,\alpha}$, the parameter vector θ is orthogonal to the variance ($\sigma^2_{y_0}$) of the initial values y_{i0}. Thus, the estimation of θ based on the conditional log-density (6) is equivalent to the joint estimation of θ and $\sigma^2_{y_0}$ based on the unconditional log-density of $(y_{i0}, \Delta y_{i1}, ..., \Delta y_{iT})'$. However, this is not the case for (7). The process (7) implies that the y_{i0} are distributed around $\eta_i = \alpha_i / (1-\delta)$, and that $\sigma_{y_o,\alpha} = (1-\delta)\sigma_{y_o,\eta} = (1-\delta)\sigma^2_\eta$. But this restriction implies that $\psi = (1-\delta)\sigma^2_\eta / \sigma^2_{y_0}$, and

$$\omega = (1-\delta)^2(\sigma^2_\eta \sigma^2_{y_o} - \sigma^2_\eta)/(\sigma^2_{y_0}\nu) = (1-\delta-\psi)\psi\sigma^2_{y_o} / \nu . \tag{8}$$

Thus, under (7), ω is no longer a free parameter: It depends on a new parameter vector $\overline{\theta} = (\delta, \nu, \sigma^2_{y_o}, \psi)'$. Observe that the marginal density function of y_{i0}, as well as the conditional density of $(\Delta y_{i1}, ..., \Delta y_{iT})'$, depends on $\overline{\theta}$. Thus, the estimator of $\overline{\theta}$ based on the conditional log-density (7) is not equivalent to the estimator of θ based on the unconditional log-density of $(y_{i0}, \Delta y_{i1}, ..., \Delta y_{iT})'$. The ML estimator of $\overline{\theta}$ based on the unconditional density will be more efficient. Apparently, the unconditional density has more information on δ if data are generated by (7). Furthermore, as we show below, when data follow unit root processes, the conditional ML estimator of δ has a non-standard asymptotic distribution, while the unconditional ML estimator computed with the restriction (8) remains asymptotically normal. Of course, however, the unconditional ML estimator becomes inconsistent if the restriction (8) does not hold. Thus, in this paper, we focus on the ML estimation of the model (1).

The function (6) provides a foundation by which the RE ML estimator of δ can be compared to the HPT and Lancaster estimators. HPT propose to estimate δ based on the following differenced model:[6]

$$\begin{pmatrix} \Delta y_{i1} \\ \Delta y_i \end{pmatrix} = \begin{pmatrix} 0 \\ \Delta y_{i,-1} \end{pmatrix} \delta + \begin{pmatrix} u_{HPT,i} \\ \Delta \varepsilon_i \end{pmatrix},$$

where,

$$Var\begin{pmatrix} u_{HPT,i} \\ \Delta \varepsilon_i \end{pmatrix} \equiv \nu \Omega_{HPT,T}(\omega_{HPT}) = \nu \begin{pmatrix} \omega_{HPT}+1 & c'_{T-1} \\ c_{T-1} & B_{T-1} \end{pmatrix}.$$

If we assume the normality of the error vector $(u_{HPT,i}, \Delta\varepsilon'_i)'$, the HPT model leads to the following log-density function:

$$\begin{aligned} \ell_{HPT,i}(\theta_{HPT}) \equiv & -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\nu) - \frac{1}{2}\ln(\xi_{HPT,T}) - \frac{1}{2\nu}(\Delta y_i - \Delta y_{i,-1}\delta)' B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta) \\ & - \frac{1}{2\nu\xi_{HPT,T}}\left(\Delta y_{i1} + k'_{T-1}(\Delta y_i - \delta \Delta y_{i,-1})\right)^2, \end{aligned} \quad (9)$$

where $\theta_{HPT} = (\delta, \nu, \omega_{HPT})'$ and $\xi_{HPT,T} = \det(\Omega_{HPT,T}) = T\omega_{HTP} + 1$. Observe that this log-density function and the corresponding log-likelihood function depend on only three parameters, while the RE log-density function (6) depends upon four parameters.

While the function (9) is almost identical to (6), a critical difference between the two functions is that the former does not depend on the initial value y_{i0}. That is, (9) is the unconditional log-density of Δy_{i1}, ..., Δy_{iT}. Note that under the RE assumption,

$$u_{HPT,i} = \psi y_{i0} + u_i, \quad \omega_{HPT} = \mathrm{var}(p_i)/\nu = \left(\psi^2 \sigma_{y_0}^2\right)/\nu + \omega. \quad (10)$$

The HPT ML method treats the y_{i0} as unobservables. This treatment does not lead to inconsistent estimators. But it will lead to inefficiency under the RE assumption. The HPT

[6] HPT in fact include an intercept term a for the Δy_{i1} equation. But the interceptor term equals zero under our zero-mean assumptions.

estimation does not exploit the information about δ that is contained in the level data y_{i0}. It only exploits the information contained in differenced data. In contrast, the RE estimator utilizes the information about δ contained in level data (through the non-zero correlation with Δy_{i1} and y_{i0}). Of course, both the estimators become asymptotically equivalent if $\psi = 0$.

The RE and HPT estimators are motivated from the RE assumption (2). However, the consistency of the RE and HPT ML estimator does not require that all variables should be normal. So long as $E(\varepsilon_i \mid y_{i0}, \alpha_i) = 0_{T\times 1}$ and $E(\varepsilon_i \varepsilon_i' \mid y_{i0}, \alpha_i) = \sigma_\varepsilon^2 I_T$, both the estimators are consistent. This can be shown easily because the score vectors of the RE and HPT log-likelihood functions have zero expectations under the two conditions.

We now turn to the Lancaster estimator which has a Bayesian flavor. We derived model (3) by first-differencing model (1). Instead, if we difference out y_{i0} from y_{it}, model (1) reduces to

$$\Delta_0 y_i = \Delta_0 y_{i,-1}\delta + 1_T p_i + \varepsilon_i , \tag{11}$$

where $p_i = (\delta - 1) y_{i0} + \alpha_i$, $\mathbf{1}_T$ is a $T\times 1$ vector of ones, $\Delta_0 y_i = (y_{i1} - y_{i0}, y_{i2} - y_{i0}, ..., y_{iT} - y_{i0})'$ and $\Delta_0 y_{i,-1} = (0, y_{i1} - y_{i0}, y_{i2} - y_{i0}, ..., y_{i,T-1} - y_{i0})'$. Lancaster treats the p_i as the unobservable effects instead of the α_i. This treatment is similar to that of HPT in that both do not exploit the information contained in the level of y_{i0}.

The density of $\Delta_0 y_i$ conditional on δ, ν and p_i equals

$$\begin{aligned} f(\Delta_0 y_i \mid \delta, \nu, p_i) &= \frac{1}{(2\pi)^{T/2}\nu^{T/2}} \exp\left(-\frac{1}{2\nu}(\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T)'(\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T) \right) \\ &= \frac{1}{(2\pi)^{T/2}\nu^{T/2}} \exp\left(\begin{array}{l} -\frac{1}{2\nu}(\Delta y_i - \Delta y_{i,-1}\delta)' B_{T-1}^{-1} (\Delta y_i - \Delta y_{i,-1}\delta) \\ -\frac{T}{2\nu}(\Delta y_{i1} + k_{T-1}{}'(\Delta y_i - \Delta y_{i,-1}\delta) - p_i)^2 \end{array} \right), \end{aligned} \tag{12}$$

where the second equality is shown in the appendix. Observe that $y_{it} - y_{i0} = \Sigma_{s=1}^{t} \Delta y_{is}$. Thus, the second equality of (12) implies that the density $f(\Delta_0 y_i \mid \delta, \nu, p_i)$ can be also viewed as a density of $r_i = (\Delta y_{i1}, ..., \Delta y_{iT})'$ conditional on δ, ν and p_i.

Using the method of Cox and Reid (1987), Lancaster reparameterizes the effects p_i

defining $p_i = \tilde{p}_i \exp(-b(\delta))$, where

$$b(\delta) = T^{-1}\Sigma_{t=1}^{T-1}[(T-t)/t]\delta^t . \tag{13}$$

This reparameterization is chosen so that the new fixed effects $\tilde{p}_i$ are *information orthogonal* to $(\delta, \nu)'$, in the sense that

$$E\left(\frac{\partial^2 \ln(f(\Delta_0 y_i \mid \delta, \nu, p_i))}{\partial(\delta,\nu)'\partial\tilde{p}_i}\right) = 0_{2\times1} .$$

With this reparameterization and assuming uninformative uniform priors to the $\tilde{p}_i$, we can integrate them out from $f(\Delta_0 y_i \mid \delta, \nu, p_i)$. If we do so, we can show that the conditional density of the differenced y_{it}'s, $(\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT})'$, conditional on $(\delta, \nu)'$, is given

$$f(\Delta_0 y_i \mid \delta, \nu) \propto \frac{1}{\nu^{(T-1)/2}} \exp(b(\delta)) \exp\left(-\frac{1}{2\nu}(\Delta y_i - \Delta y_{i,-1}\delta)' B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta)\right),$$

which leads to the log-density function

$$\ell_{Lan,i}(\delta,\nu) = b(\delta) - \frac{(T-1)}{2}\ln(\nu) - \frac{1}{2\nu}(\Delta y_i - \Delta y_{i,-1}\delta)' B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta) . \tag{14}$$

Observe that Lancaster's ML depends on only two parameters. It does not depend on the variance $\nu\omega$ of the composite error term u_i in model (3). Notice that the error term u_i contains the projection error component of the effect α_i (the error in the population regression of α_i on y_{i0}). Thus, the fact that (14) does not depend on ω seems to imply that the Lancaster estimator is indeed a fixed-effect treatment. However, it is not without costs. That is, while the Lancaster estimator does not require any distributional assumption about the effect α_i, it loses the usual ML properties as we see below.[7]

---

[7] An intriguing question would be whether the Lancaster and/or HPT estimators are consistent under broader circumstances than the RE estimator is. Consider the two conditions: (i) $p\lim_{N\to\infty} N^{-1}\Sigma_i p_i^2 < \infty$ is finite; and (ii) $p\lim_{N\to\infty} N^{-1}\Sigma_i (y_{io}, \alpha_i)'(y_{i0}, \alpha_i)$ is finite. Obviously, condition (ii) is stronger than (i). Kruiniger (2002) finds that a necessary condition for consistency of the Lancaster estimator is (i). It can be shown that the same condition is required for the consistency of the HPT estimator. As long as the condition holds and the errors ε_{it} are i.i.d. and uncorrelated with p_i, both the Lancaster and HPT estimators are consistent, even if data are not normal. In contrast, it can be shown that the consistency of the RE estimator requires the stronger restriction (ii). Thus, it is true that that the Lancaster and/or the HPT estimator are consistent under more general conditions. However, there are few realistic cases in which condition (i) holds while (ii) is violated. Condition (ii) would be an acceptable assumption for most of the panel studies (at least the studies with short panels). If (ii) holds, all of the three estimators are consistent, and thus the distinction between RE and FE becomes unimportant.

An interesting property of the Lancaster estimator, $(\hat{\delta}_{Lan}, \hat{\nu}_{Lan})'$, is that the asymptotic covariance matrix of the Lancaster estimator is not of the inverted Hessian form. Define

$$H_{Lan,i}(\delta,\nu) = \frac{\partial^2 \ell_{Lan,i}}{\partial(\delta,\nu)'\partial(\delta,\nu)}; B_{Lan,i}(\delta,\nu) = \left(\frac{\partial \ell_{Lan,i}}{\partial(\delta,\nu)'}\right)\left(\frac{\partial \ell_{Lan,i}}{\partial(\delta,\nu)}\right)'.$$

Then, it can be shown[8]:

$$\sqrt{N}\begin{pmatrix} \hat{\delta}_{Lan} - \delta_o \\ \hat{\nu}_{Lan} - \nu_o \end{pmatrix} \rightarrow_d N\left(0_{2\times 1}, \left(-E(H_{Lan,i}(\delta_o,\nu_o))\right)^{-1} E(B_{Lan,i}(\delta_o,\nu_o))\left(-E(H_{Lan,i}(\delta_o,\nu_o))\right)^{-1}\right),$$

where "$\rightarrow_d$" means "converges in distribution," and $(\delta_o,\nu_o)'$ denotes the true value of $(\delta,\nu)'$. The reason why the asymptotic covariance matrix of the Lancaster estimator is not of the inverted Hessian form is that $\int f_{Lan}(\Delta_0 y_i \mid \delta,\nu) d(\Delta y_{i1},...\Delta y_{iT}) \neq 1$. That is, $f_{Lan}(\Delta_0 y_i \mid \delta,\nu)$ is not a proper density. Thus, the Lancaster estimator is not really a ML estimator.[9]

Kruiniger (2002) shows that the Lancaster estimator of δ is inefficient compared to the HPT estimator of δ when $T > 2$, while the asymptotic variance matrices of the two estimators are the same when $T = 2$.[10] We can obtain the same result in a more intuitive way.

Proposition 1: Suppose that the prior density of p_i, $f(p_i \mid \omega_{HPT},\nu)$, is given $N(0,\nu\omega_{HPT})$. Then,

$$\int_{-\infty}^{\infty} f(\Delta_0 y_i \mid \delta,\nu,p_i) f(p_i \mid \omega_{HPT},\nu) dp_i = f(r_i \mid \delta,\omega_{HPT},\nu).$$

All of the proofs are in the appendix. While a rigorous proof of Proposition 1 is given in the appendix, the result of the proposition is intuitive given the law of iterative expectations. The proposition implies that if $f(\Delta_0 y_i \mid \delta,\nu,p_i)$ were integrated with the normal prior $f(p_i \mid \omega_{HPT},\nu)$, the Lancaster estimator becomes equivalent to the HPT estimator. This explains why the HPT estimator should be more efficient than the Lancaster estimator under our RE assumption or the assumptions justifying the HPT ML method. Under the RE assumption, the prior $f(p_i \mid \omega_{HPT},\nu)$ is informative.

[8] See Kruiniger (2002).
[9] Lancaster also acknowledges this point.
[10] However, even when $T = 2$, $f_{Lan}(\Delta_0 y_i \mid \delta,\nu)$ is not a proper density function.

3. Maximum Likelihood When Data Are Random Walks

In this section, we investigate the asymptotic distribution of the three estimators discussed above when data contain unit roots with and without drifts. A number of studies have proposed different test methods for unit roots. Examples are, among many, Levin, Lin and Chu (2004), Im, Pesaran and Shin (2003), and, Moon and Phillips (2004), and Moon, Perron and Phillips (2005). These studies consider the cases of large *N* and *T*. Alternative unit-root tests for the data with fixed *T* have been studied by Breitung and Meyer (1994), and Harris and Tzavalis (1999). By investigating the asymptotic distributions of the ML estimators, we derive the distributions of the Likelihood-Ratio (LR) statistics for testing unit roots in data with fixed *T*.

3.1. Random Walks without Drifts

In this subsection, we consider the asymptotic distribution of the random effects ML estimator when data follow unit root processes. We begin by considering the cases in which the y_{it} are random walks without drifts: $y_{it} = y_{i,t-1} + \varepsilon_{it}$, for t = 1, 2, ... , T. We can easily see that this unit root process is equivalent to the following three parametric restrictions on model (3):

$$H_o^{UND} : \theta_o = (\delta_o, \nu_o, \omega_o, \psi_o)' = (1, \nu_*, 0, 0)' \equiv \theta_* \qquad (15)$$

where the superscripted "UND" refers to "unit-root with no drift", the subscripted "o" indicates the true value of the corresponding parameter, and ν_* is unrestricted.

For notational convenience, we use $\ell_{RE,i,\theta}$ and $\ell_{RE,i,\theta\theta}$ to denote the score vector and the Hessian matrix of the log-density $\ell_{RE,i}$, respectively. We will use the same rule to denote the derivatives of the log-density function with respect to individual parameters: for example, $\ell_{RE,i,\delta} = \partial \ell_{RE,i} / \partial \delta$ and $\ell_{RE,i,\delta\delta} = \partial^2 \ell_{RE,i} / \partial \delta^2$.

In usual maximum likelihood theory, the information matrices of log-density (in our case, $(E(-\ell_{RE,i,\theta\theta}(\theta_o)))$ are assumed to be nonsingular. However, when data are generated with the restrictions (15), this is no longer the case. We state this finding formally.

Proposition 2: Under H_o^{UND}, both of $E(-\ell_{RE,i,\theta\theta}(\theta_*))$ and $E(\ell_{RE,i,\theta}(\theta_*)\ell_{RE,i,\theta}(\theta_*)')$ are singular.

The proposition implies that the information matrix of the RE log-density function (6) is singular. This proposition also applies to the HPT and Lancaster estimation. It means that the usual asymptotic theory of ML estimation does not apply to the RE, HPT and Lancaster ML estimation when data follow random walks without drift. Singular information matrices often imply model non-identification. However, as Sargan (1983) finds from the analysis of the models linear in variables and nonlinear in parameters, a singular information matrix (first-order condition for lack of identification) is not a necessary condition of non-identification. Similarly to his cases, Proposition 2 does not imply that the RE model (3) is not identifiable when $\theta_o = \theta_*$. If the RE model were not identifiable, it should be the case that θ_* is not a unique maximizer of $E(\ell_{RE,i}(\theta))$ when $\theta_o = \theta_*$. But this is not the case. To see why, consider a simple case of T = 2 at $\theta_o = \theta_*$. If we concentrate $E(\ell_{RE,i}(\theta))$ by maximizing it with respect to the nuisance parameter vector $(\nu,\omega,\psi)'$ given δ, we can obtain:

$$E\left(\ell^c_{RE,i}(\delta)\right) = -\frac{1}{2}\ln\left(\frac{1}{2}+\frac{\delta^2}{2}\right) - \frac{1}{2}\ln\left(\frac{5}{2}-2\delta+\frac{1}{2}\delta^2\right) + g(\nu_*,T),$$

where the superscript "c" means "concentrated", and $g(\nu_*,T)$ is some function of the variance parameter ν_* and T. Then, it can be show that at $\delta = 1$,

$$E(\ell^c_{RE,i,\delta}(\delta)) = E(\ell^c_{RE,i,\delta\delta}(\delta)) = E(\ell^c_{RE,i,\delta\delta\delta}(\delta)) = 0; E(\ell^c_{RE,i,\delta\delta\delta\delta}(\delta)) = -3\,.$$

Thus, although the expected value of the second derivative of $\ell^c_{RE,i}(\delta)$ equals zero, $\delta = 1$ is still a local maximum point. Moreover, it can be shown that

$$E\left(\ell^c_{RE,i,\delta}(\delta)\right) = -\frac{2(\delta-1)^3}{\left[(\delta-2)^2+1\right](\delta^2+1)}.$$

Thus, $E\left(\ell^c_{RE,i}(\delta)\right)$ is always increasing when $\delta < 1$ and decreasing when $\delta > 1$. This indicates that $\delta = 1$ is the global maximum point of $E\left(\ell^c_{RE,i}(\delta)\right)$ when $\theta_o = \theta_*$. This shows that θ_* is the global maximum point of $E(\ell_{RE,i}(\theta))$ when $\theta_o = \theta_*$. We can generalize this result to the cases with general T. Stated formally:

Proposition 3: Under H_o^{UND}, θ_* is a unique global maximizer of $E(\ell_{RE,i}(\theta))$. Thus, the RE

ML estimator, $\hat{\theta}_{RE} = (\hat{\delta}_{RE}, \hat{\nu}_{RE}, \hat{\omega}_{RE}, \hat{\psi}_{RE})'$, is a consistent estimator under H_o^{UND}.

This proposition applies to the HPT estimator. However, somewhat surprisingly, it does not apply to the Lancaster ML estimation. When data are random walks, the point $\delta = 1$ is an inflexion point of the expectation of Lancaster log-likelihood function.

Proposition 4: Under H_o^{UND}, θ_* does not maximize $E\left(\ell_{Lan,i}(\delta,\nu)\right)$. In fact, θ_* is an inflexion point.

This proposition is shown by investigating the concentrated expected log-likelihood function $E\left(\ell^c_{Lan,i}(\delta)\right)$ constructed similarly to $E\left(\ell^c_{RE,i}(\delta)\right)$. For example, when $T = 2$, we obtain:

$$E\left(\frac{\partial \ell^c_{Lan,i}(\delta)}{\partial \delta}\right) = \frac{(\delta-1)^2}{2(\delta^2+1)} \geq 0 .$$

Thus, $E\left(\ell^c_{Lan,i}(\delta)\right)$ is always increasing except when $\delta = 1$.

Proposition 4 implies that the estimator of δ which directly maximizes the log-likelihood function $\Sigma_{i=1}^N \ell_{Lan,i}(\delta,\nu)$ may not be consistent when data contain a unit root. In fact, in unreported simulations with data containing unit roots, we often failed to locate the maximizing values of $(\delta,\nu)'$, although, when located, they were close to one. In contrast, Given that $\delta = 1$ is the unique root of $E\left(\partial \ell^c_{Lan,.i}(\delta)/\partial\delta\right) = 0$ when $\theta_o = \theta_*$, the GMM estimator using the score vector of $\Sigma_{i=1}^N \ell_{Lan,i}(\delta,\nu)$ as moment functions would be consistent. However, the GMM estimator would not be normal even asymptotically. It can be shown that the expectation of the Hessian matrix of $\Sigma_{i=1}^N \ell_{Lan,i}(\delta,\nu)$ is singular under H_o^{UND}. In this paper we will not further explore the asymptotic distribution of the Lancaster estimator, although it would be an interesting research agenda. We focus on the asymptotic distributions of the RE and HPT estimators. Given that the Lancaster estimator does not use a proper density, it could be viewed as a GMM estimator or a pseudo ML estimator in the sense of Gouriéroux, Monfort and Trognon (1984). The approach of Rotnitzky, Cox, Bottai and Robins (2000), which we use below to derive the asymptotic distributions of the RE and HPT ML estimator, is for the ML estimators. It would be a valuable research agenda to investigate how their approach can be generalized to GMM or pseudo ML estimators.

Proposition 4 is a somewhat counterintuitive result in that the expected value of the HPT log-likelihood function has a global maximum at $\delta = 1$ when $\theta_o = \theta_*$. Recall that when $\delta_o < 1$ and $T = 2$, the Lancaster and HPT estimators are asymptotically equivalent. Proposition 4 implies that this equality breaks down when data are random walks ($\theta_o = \theta_*$).

Sargan (1983) shows, for the models linear in variables but nonlinear in parameters, that when information matrices are singular, ML estimators are not asymptotically normal, although they may be consistent. Similar to his case, Proposition 3 warrants that the RE and HPT estimators are consistent, but their asymptotic distributions are not normal when $\theta_o = \theta_*$ as we show below. While the asymptotic distributions of the two RE and HPT estimators are similar to those of the ML estimators Sargan considered, his results do not directly apply to the dynamic panel data model (3). The model Sargan (1983) has analyzed subsumes linear simultaneous equation models without variance-covariance restrictions on error terms. The dynamic panel model (3) can be viewed as a linear simultaneous equations model if we treat Δy_{i1}, ... , Δy_{iT} as endogenous regressors. However, the model (3) is not a special case of Sargan's model because it imposes variance-covariance restrictions on the error terms.

Fortunately, Rotnitzky, Cox, Bottai and Robins (2000, RCBR) derive the asymptotic distributions of the ML estimators of more general models when their information matrices are singular. We can derive the asymptotic distributions of the RE and HPT estimators using their method. Here, we will focus on the RE estimator only. All of the results we obtain below also apply to the HPT estimator.

RCBR study the cases in which the derivatives of log-density functions are linearly dependent (so that Hessian matrices become singular). For such cases, they derive the asymptotic distributions of ML estimators and likelihood-ratio (LR) test statistics. To use their method, we need to show that the derivatives of $\ell_{RE,i}$ are linearly dependent at $\theta = \theta_*$. Stated formally:

Proposition 5: $\ell_{RE,i,\delta}(\theta_*) - \ell_{RE,i,\omega}(\theta_*) + \nu_* \ell_{RE,i,\nu}(\theta_*) = 0$.

To get tractable asymptotic results, we now need to reparameterize the model.

Following RCBR, we define the following new parameter vector:

$$\theta_r \equiv \begin{pmatrix} \delta_r \\ \nu_r \\ \omega_r \\ \psi_r \end{pmatrix} = \begin{pmatrix} \delta \\ \nu \\ \omega \\ \psi \end{pmatrix} + \begin{pmatrix} 0 \\ \left[E\left(\ell_{RE,i,\varphi}(\theta_*)\ell_{RE,i,\varphi}(\theta_*)'\right)\right]^{-1} E\left(\ell_{RE,i,\varphi}(\theta_*)\ell_{RE,i,\delta}(\theta_*)\right) \end{pmatrix}(\delta-1)$$

$$= \begin{pmatrix} \delta \\ \nu \\ \omega \\ \psi \end{pmatrix} + \begin{pmatrix} 0 \\ -\nu_* \\ 1 \\ 0 \end{pmatrix}(\delta-1) = \begin{pmatrix} \delta \\ \nu-\nu_*(\delta-1) \\ \omega+(\delta-1) \\ \psi \end{pmatrix},$$

where $\varphi = (\nu, \omega, \psi)'$, and E(•) is computed assuming H_o^{UND}. This parameterization is chosen to secure that when $\theta = \theta_*$, $\theta_r = \theta_*$. The reparameterization requires the knowledge of the true variance of ε_{it}, ν_*. However, this problem can be resolved by replacing ν_* by ν_r.

The above reparameterization means that we retain δ and ψ in $\ell_{RE,i}(\theta)$, but treat ω and ν as the functions of δ and the new parameters ω_r and ν_r:

$$\omega = \omega_r - (\delta-1); \nu = \nu_r + \nu_r(\delta-1) = \nu_r\delta. \tag{16}$$

Let $\theta_r = (\delta, \varphi_r')'$, $\varphi_r' = (\nu_r, \omega_r, \psi)$ and $\tilde{\ell}_{RE,i}(\theta_r) = \ell_{RE,i}(\delta, \nu(\delta,\nu_r), \omega(\delta,\omega_r), \psi)$. Under this reparameterization, $\theta_{r,o} = \theta_*$ whenever $\theta_o = \theta_*$. In addition, the reparameterization is designed to have

$$\tilde{\ell}_{RE,i,\delta}(\theta_*) = \ell_{RE,i,\delta}(\theta_*) - \ell_{RE,i,\omega}(\theta_*) + \nu_*\ell_{RE,i,\nu}(\theta_*) = 0.$$

This reparameterization is necessary because the RCBR approach is basically for the cases where a first derivative of a log-likelihood function equals (not asymptotically, but exactly) zero. The asymptotic distributions of the ML estimators of ω and ν are different from those of the ML estimators of ω_r and ν_r, although the former can be derived from the latter (see RCBR). However, the distribution of the ML estimator of δ can be directly obtained from that of the ML estimator from the reparameterized model.

Define:

$$s_i = \left(s_{i,1} \quad s_{i2}\right)'; \; s_{i,1} = \tilde{\ell}_{RE,i,\delta\delta}(\theta_*)/2; \; s_{i,2} = \tilde{\ell}_{RE,i,\varphi_r}(\theta_*);$$

$$\Upsilon = E(s_i s_i') = [\Upsilon_{ij}], \quad i,j = 1,2\,; \;\; \Upsilon^{-1} = [\Upsilon^{ij}]\,.$$

To use the RCBR approach, we need to check (i) whether or not $\tilde{\ell}_{RE,i,\delta\delta}(\theta_*)$ equals zero or is linearly related to $\tilde{\ell}_{RE,i,\delta}(\theta_*)$ and $s_{i,2}$, and (ii) whether or not $\tilde{\ell}_{RE,i,\delta\delta\delta}(\theta_*)$ is a linear combination of s_i. We find that only $\tilde{\ell}_{RE,i,\delta}(\theta_*)$ equals zero, not $\tilde{\ell}_{RE,i,\delta\delta}(\theta_*)$, and that $\tilde{\ell}_{RE,i,\delta\delta}(\theta_*)$ is not a linear combination of $\tilde{\ell}_{RE,i,\delta}(\theta_*)$ and $s_{i,2}$. We also find that $\tilde{\ell}_{RE,i,\delta\delta\delta}(\theta_*)$ is not a linear combination of $s_i(\theta_*)$. Based on this finding, we obtain the following result:

Proposition 6: Let $Z = (Z_1, Z_2)'$ denote a mean-zero normal vector with $Var(Z) = \Upsilon^{-1}$, where Z_1 is a scalar and Z_2 is a 3×1 vector. Let $\hat{\theta}_r = (\hat{\delta}, \hat{\varphi}_r')'$ be the ML estimator that maximizes $\Sigma_{i=1}^N \tilde{\ell}_{RE,i}(\theta)$. Assume that H_o^{UND} holds; that is, $(\delta_o, \omega_o, \psi_o)' = (1,1,0)'$. Then,

$$\begin{pmatrix} N^{1/4}(\hat{\delta}-1) \\ N^{1/2}(\hat{\varphi}_r - \varphi_*) \end{pmatrix} \to_d \begin{pmatrix} (-1)^B Z_1^{1/2} \\ Z_2 \end{pmatrix} 1(Z_1 > 0) + \begin{pmatrix} 0 \\ Z_2 - \Upsilon^{21}(\Upsilon^{11})^{-1} Z_1 \end{pmatrix} 1(Z_1 < 0), \quad (17)$$

where $\varphi_* = (v_*, 1, 0)'$, "$\to_d$" means "converges in distribution," and *B* is a Bernoulli random variable with success probability equal to half and independent of *Z*. The Likelihood Ratio (LR) statistic for H_o^{UND} is a mixture of $\chi^2(2)$ and $\chi^2(3)$ with the mixing probability equal to one half. The LR test for the hypothesis that $\delta_o = 1$ is a mixture of $\chi^2(1)$ and zero with the mixing probability equal to one half.[11]

Several comments follow from Proposition 6. First, the ML estimator $\hat{\theta}_r$ is not normal even asymptotically. Since $\hat{\delta}_{RE} = \hat{\delta}$, the above theorem indicates that the convergence rate of $\hat{\delta}_{RE}$ is $N^{1/4}$. The result also implies that the ML-based *t*-test for the hypothesis of $\delta_o = 1$ would not be properly sized.[12] Second, the probability that the ML estimator $\hat{\delta}_{RE}$ is equal to the true value

[11] If the means of the y_{i0} and α_i are non-zero, we need to include an intercept term, say *a*, in the log-likelihood function replacing $(\Delta y_{i1} - \psi y_{i0})$ by $(\Delta y_{i1} - \psi y_{i0} - a)$. Then, the unit-root hypothesis (15) implies *a = 0* in addition to the restrictions in θ_*. For this case, the LR statistic for testing all of the restrictions implied by (15) is a mixture of $\chi^2(3)$ and $\chi^2(4)$.

[12] We may consider an alternative Wald-type test statistic, $N(\hat{\delta}_{RE} - 1)^4 / \Upsilon^{11}$. Proposition 6 implies that under H_o^{UND},

of δ converges to half as $N \to \infty$. This implies that the asymptotic distribution of the ML estimator $\hat{\delta}_{RE}$ will be spiked at $\delta = 1$ when $\theta_o = \theta_*$. Third, the two LR test statistics from the reparameterized log-likelihood $\tilde{\ell}_{RE,i}(\theta_r)$ are the same as the LR test statistics directly obtained from the original log-likelihood $\ell_{RE,i}(\theta)$. Thus, using the LR statistics for testing H_o^{UND}, we do not need the parameterization (16). Fourth, we can think of the LR statistic for testing the joint hypothesis that $\delta_o = 1$ and $\omega_o = 0$. It can be shown that when $\theta_o = \theta_*$, this test statistic follows a mixture of $\chi^2(1)$ and $\chi^2(2)$.[13,14]

Finally, we can obtain the similar results for the HPT estimator. The unit-root hypothesis (15) implies that $\delta_o = 1$ and $\omega_{HPT,o} = 0$. The HPT-based LR test for this joint hypothesis follows a mixture of $\chi^2(2)$ and $\chi^2(1)$.

3.2. Random Walks with Homogenous Drifts

We now consider the ML estimation of the dynamic panel data model with homogeneous trends, and the LR test for the hypothesis of random walk with homogeneous drifts. Specifically, we consider the following model:

$$y_{it} = \delta y_{i,t-1} + \beta t + (\alpha_i + \varepsilon_{it}) . \tag{18}$$

We assume that $(y_{i0}, \alpha_i, \varepsilon_i')'$ satisfies the RE assumption (2), but we now allow the initial values y_{i0} and the effect α_i to have nonzero means. Assume that $E(\alpha_i \mid y_{i0}) = \gamma_1 + \gamma_2 y_{i0}$. Then, similarly to what we have done from (1) to (3), we can transform (18) into

---

this statistic follows a mixture of $\chi^2(1)$ and zero.

[13] These LR tests are two-tail tests. When T is fixed, the RE likelihood function is well defined in the neighborhood of $\delta = 1$. Thus the RE ML estimator is not subject to the boundary problem raised by Andrews (1999), and the hypothesis of $\delta_o = 1$ can be tested against the two-tail alternative hypothesis of $\delta_0 \neq 1$. This justifies use of the LR tests. However, some one-tail alternatives of the LR tests would be more powerful since δ_o is unlikely to be greater than one, although we do not investigate them here.

[14] Arellano and Bover propose to use for GMM the moment conditions, $E[\Delta y_{it}(y_{is} - \delta y_{i,s-1})] = 0$, $t < s$. These moment conditions are valid under (7), but not under (1) and (2). Observe that these moment conditions can identify the true value of δ even if data follow unit root processes without drifts. In contrast, the moment conditions by Arellano and Bond (1991), $E(y_{it}(\Delta y_{is} - \delta \Delta y_{i,s-1})) = 0$, $t < s$, are motivated by the model given by (1) and (2). As Blundell and Bond (1998) find, these moment conditions are unable to identify δ if data follow random walk processes because the level instruments y_{it} are uncorrelated with differenced regressors $\Delta y_{i,s-1}$.

$$\begin{pmatrix}\Delta y_{i1}\\ \Delta y_i\end{pmatrix}=\begin{pmatrix}y_{i0} & 1\\ 0_{(T-1)\times 1} & 0_{(T-1)\times 1}\end{pmatrix}\begin{pmatrix}\psi\\ a\end{pmatrix}+\begin{pmatrix}0 & 0\\ \Delta y_{i,-1} & 1_{T-1}\end{pmatrix}\begin{pmatrix}\delta\\ \beta\end{pmatrix}+\begin{pmatrix}u_i\\ \Delta\varepsilon_i\end{pmatrix}, \tag{19}$$

where $a=\gamma_1+\beta$, $\psi=(\delta-1)+\gamma_2$ and $u_i=(\alpha_i-E(\alpha_i\mid y_{i0}))+\varepsilon_{i1}$. The individual log-density function for this model (conditional on y_{i0}) is given:

$$\begin{aligned}\ell^D_{RE,i}(\phi^D)=&-\frac{T}{2}\ln(2\pi)-\frac{T}{2}\ln(\nu)-\frac{1}{2}\ln(\xi_T)\\&-\frac{1}{2\nu}\left(\Delta y_i-\delta\Delta y_{i,-1}-1_{T-1}\beta\right)' B^{-1}_{T-1}\left(\Delta y_i-\delta\Delta y_{i,-1}-1_{T-1}\beta\right)\\&-\frac{T}{2\nu\xi_T}\left((\Delta y_{i1}-\psi y_{i0}-a)+k_{T-1}{}'(\Delta y_i-\delta\Delta y_{i,-1}-1_{T-1}\beta)\right)^2,\end{aligned} \tag{20}$$

where $\phi^D=(\delta,\nu,\omega,\psi,a,\beta)'$, and $\xi_T=T\omega+1$, as before.

Suppose that data are generated by the following trend-stationary process:

$$y_{it}=\eta_i+gt+q_{it};\;\; q_{it}=\delta q_{i,t-1}+\varepsilon_{it},\;\; t=0,...,T, \tag{21}$$

where $\varepsilon_i\sim N(0_{T\times 1},\nu I_T)$, η_i and q_{i0} are also normally distributed, and all of ε_i, η_i and q_{i0} are mutually independent. Assume that $E(\eta_i\mid y_{i0})=\tau_1+\tau_2 y_{i0}$. This data generation process is the same as the process (18), if we set:

$$\alpha_i=(1-\delta)\eta_i;\;\; \psi=(1-\delta)\tau_2;\;\; a=(1-\delta)\tau_1+g;\;\; \beta=(1-\delta)g.$$

However, there exists an important difference between (18) and (21). In the former, we do not assume that the initial values y_{i0} are distributed around the effect α_i. Thus, we do not impose any restriction on $\sigma_{y_o,\alpha}=\mathrm{cov}(y_{i0},\alpha_i)$. Without any restriction on $\sigma_{y_o,\alpha}$, the parameter vector ϕ^{D} is orthogonal to the variance ($\sigma^2_{y_0}$) of the initial values y_{i0}. Thus, the estimation of ϕ^D based on the conditional log-density (20) is equivalent to the joint estimation of ϕ^D and $\sigma^2_{y_0}$ based on the unconditional log-density of $(y_{i0},\Delta y_{i1},...,\Delta y_{iT})'$. However, this is not the case for (21). The process (21) implies that the y_{i0} are distributed around $\eta_i=\alpha_i/(1-\delta)$, and that $\sigma_{y_o,\alpha}=(1-\delta)\sigma_{y_o,\eta}=(1-\delta)\sigma^2_\eta$. But this restriction implies that $\psi=(1-\delta)\sigma^2_\eta/\sigma^2_{y_0}$, and

$$\omega=(1-\delta)^2(\sigma^2_\eta\sigma^2_{y_o}-\sigma^2_\eta)/(\sigma^2_{y_0}\nu)=(1-\delta-\psi)\psi\sigma^2_{y_o}/\nu. \tag{22}$$

Observe that ω does not contain any free parameter given ψ, δ, ν and $\sigma^2_{y_0}$. This means that any knowledge of $\sigma^2_{y_0}$ would help to obtain a more efficient estimator of ϕ^D. This implies that the estimation of ϕ^D based on the conditional log-density (20) is not equivalent to the joint estimation of ϕ^D and $\sigma^2_{y_0}$ based on the unconditional log-density of $(y_{i0}, \Delta y_{i1}, ..., \Delta y_{iT})'$. The ML estimators based on the unconditional density will be more efficient. Of course, these unconditional ML estimators will be inconsistent if condition (22) is violated.

We now turn to the cases where data are random walks with homogeneous drifts; that is, $\delta = 1$ in (21). We can show that the unconditional ML estimators computed with the restriction (22) are consistent and asymptotically normal.[15] Thus, the usual ML theory applies. However, the conditional ML estimators without the restriction have a different story. The hypothesis of random walk with a homogeneous drift implies the following restrictions on (20):

$$H_o^D : \phi_o^D = \phi_*^D \equiv (1, \nu_*, 0, 0, a_*, 0)' , \tag{23}$$

where "D" refers to "(homogeneous) drifts", and ν_* and a_* are unrestricted. When this hypothesis holds, we obtain essentially the same results as Proposition 5. Stated formally:

Proposition 7: $\ell^D_{RE,i,\delta}(\phi_*^D) - a_*\ell^D_{RE,i,\beta}(\phi_*^D) - \ell^D_{RE,i,\omega}(\phi_*^D) + \nu_*\ell^D_{RE,i,\nu}(\phi_*^D) = 0$.

Proposition 7 implies that the Hessian matrix (not just the information matrix) of the RE ML estimator of ϕ^D is singular under H_o^D. In order to derive the asymptotic distribution of the ML estimator, we can use the following reparameterization: Define $\phi_r^D = (\delta, \nu_r, \omega_r, \psi, a, \beta_r)'$, where

$$\nu = \nu(\nu_r, \delta) = \nu_r\delta \,; \quad \omega(\omega_r, \delta) = \omega_r - (\delta - 1) \,; \quad \beta = \beta(\beta_r, \delta, a) = \beta_r - a(\delta - 1) \,. \tag{24}$$

Then, $\phi^D_{r,0} = \phi^D_*$ whenever $\phi^D_o = \phi^D_*$. Let

$$\tilde{\ell}^D_{RE,i}(\phi_r^D) = \ell^D_{RE,i}(\delta, \nu(\nu_r, \delta), \omega(\omega_r, \delta), \psi, a, \beta(\beta_r, \delta, a)) ,$$

so that

$$\tilde{\ell}^D_{RE,i,\delta}(\phi_*^D) = \ell^D_{RE,i,\delta}(\phi_*^D) - a_*\ell^D_{RE,i,\beta}(\phi_*^D) - \ell^D_{RE,i,\omega}(\phi_*^D) + \nu_*\ell^D_{RE,i,\nu}(\phi_*^D) = 0 .$$

With this parameterization, we redefine:

$$\varphi_r = (\nu_r, \omega_r, \psi, a, \beta_r)' ;$$

---

[15] The restricted ML estimators are also consistent and asymptotically normal for the models without drifts.

$$s_i = (s_{i,1}, s'_{i,2})';\;\; s_{i,1} = \tilde{\ell}^D_{RE,i,\delta\delta}(\phi^D_*)/2;\;\; s_{i,2} = \tilde{\ell}^D_{RE,i,\varphi_r}(\phi^D_*);\;\; \Upsilon = Var(s_i \mid \phi^D_{r,o} = \phi^D_*).$$

Similarly to Proposition 6, let $Z = (Z_1, Z_2)'$ denote a 6×1 mean-zero random vector with $Var(Z) = (\Upsilon)^{-1}$, where Z_1 is a scalar. Then, the ML estimator of ϕ^D_r, $\hat{\phi}^D_r = (\hat{\delta}, \hat{\varphi}_r{}')'$, has the same asymptotic distribution as (17). Accordingly, the LR statistic for testing H^D_o is a mixture of $\chi^2(3)$ and $\chi^2(4)$, $B\chi^2(4) + (1-B)\chi^2(3)$, where B is a Bernoulli random variable with success probability equal to one half. In addition, under H^{UD}_o, the LR statistic for testing $\delta_o = 1$ is $B\chi^2(1)$. Similar to the cases of unit-roots without drifts, these LR statistics can be computed using the original log-likelihood $\Sigma^N_{i=1}\ell^D_{RE,i}$. The reparameterization (24) is not required.

3.3. Random Walks with Heterogeneous Drifts

In this section, we examine the ML estimation of a dynamic panel data model with heterogeneous trends, and the corresponding LR test for the hypothesis of random walk with heterogeneous drifts. Under this model the data follow the following process:

$$y_{it} = \delta y_{i,t-1} + \beta_i t + (\alpha_i + \varepsilon_{it}), \tag{25}$$

As before, the error terms ε_{it} are assumed to be normal and i.i.d. over both time and individuals. The error terms are not correlated with any of α_i, β_i and y_{i0}. The identification of the model requires that $T \geq 3$. Assume that $(\alpha_i + \beta_i)$ and β_i are normal conditional on the initial value y_{i0} with conditional means linear in y_{i0}, and the conditional variance matrix, νW, where $W = [\omega_{jh}]$ (*j, h* = 1, 2) is a 2×2 matrix.

Moon and Phillips (2004) have proposed a GMM estimator for model (25). In their study, T is large and the unobservables α_i and β_i are nuisance parameters. An interesting property of the GMM estimator is that its convergence rate is $N^{1/6}$. In this subsection, we consider an alternative random-trends treatment.

Define $\Delta^2 y_{it} = \Delta y_{it} - \Delta y_{i,t-1}$, $\Delta^2 y_i = (\Delta^2 y_{i3}, ... \Delta^2 y_{iT})'$, and $\Delta^2 y_{i,-1} = (\Delta^2 y_{i2}, ... \Delta^2 y_{i,T-1})'$. We also define $u_{i1} = (\alpha_i + \beta_i - E(\alpha_i + \beta_i \mid y_{i0})) + \varepsilon_{i1}$, and $u_{2i} = (\beta_i - E(\beta_i \mid y_{i0})) + \Delta\varepsilon_{i2}$. Using this notation, we can transform the model (25) into:

$$\begin{pmatrix} \Delta y_{i1} \\ \Delta y_{i2} \\ \Delta^2 y_i \end{pmatrix} = \begin{pmatrix} a_1 + \psi_1 y_{i0} \\ a_2 + \psi_2 y_{i0} + \delta \Delta y_{i1} \\ \delta \Delta^2 y_{i,-1} \end{pmatrix} + \begin{pmatrix} u_{i1} \\ u_{i2} \\ \Delta^2 \varepsilon_i \end{pmatrix}. \tag{26}$$

Define:

$$L_T^{HD\prime} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & 0 & \dots & 0 \\ 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}_{T\times T} \quad ; \quad J_T = \begin{pmatrix} I_2 \\ 0_{(T-2)\times 2} \end{pmatrix},$$

such that $L_T^{HD\prime}(\Delta y_{i1},...,\Delta y_{iT})' = (\Delta y_{i1}, \Delta y_{i2}, \Delta^2 y_i')'$. Then, the variance matrix of the error vector in (26), $(u_{i1}, u_{i2}, \Delta^2 \varepsilon_i')'$, is of the form:

$$\nu \Omega_T^{HD}(W) = \nu \left(L_T^{HD\prime} L_T^{HD} + J_T W J_T' \right).$$

With this variance matrix, the log-density of $(\Delta y_{i1}, \Delta y_{i2}, \Delta^2 y_i')'$ conditional on y_{i0} is given:

$$\ell_{RE,i}^{HD}(\phi^{HD}) = -\frac{1}{2}\ln(2\pi) - \frac{T}{2}\ln(\nu) - \frac{1}{2}\ln\{\det[\Omega_T^{HD}(W)]\}$$
$$-\frac{1}{2\nu}\begin{pmatrix} \Delta y_{i1} - a_1 - \psi_1 y_{io} \\ \Delta y_{i2} - a_2 - \psi_2 y_{i0} - \delta \Delta y_{i1} \\ (\Delta^2 y_i - \delta \Delta^2 y_{i,-1}) \end{pmatrix}' [\Omega_T^{HD}(W)]^{-1} \begin{pmatrix} \Delta y_{i1} - a_1 - \psi_1 y_{io} \\ \Delta y_{i2} - a_2 - \psi_2 y_{i0} - \delta \Delta y_{i1} \\ (\Delta^2 y_i - \delta \Delta^2 y_{i,-1}) \end{pmatrix}, \tag{27}$$

where $\phi^{HD} = (\delta, \nu, \omega_{11}, \omega_{12}, \omega_{22}, \psi_1, \psi_2, a_1, a_2)'$. An alternative representation of the log likelihood function, which is computationally more convenient, is given in the appendix.

Suppose now that the data follow the unit-root process with heterogeneous drifts: $\Delta y_{it} = \beta_i + \varepsilon_{it}$. This process implies the following five restrictions on ϕ^{HD}:

$$H_o^{HD} : \phi_o^{HD} = (1, \nu_*, \omega_{11,*}, 0, 0, \psi_{1,*}, 0, a_{1,*}, 0)' \equiv \phi_*^{HD}, \tag{28}$$

where "HD" refers to "heterogeneous drifts,", and ν_*, $\omega_{11,*}$, $\psi_{1,*}$, and $a_{1,*}$ are unrestricted ($\delta_o = 1$, $\omega_{12,o} = \omega_{22,o} = 0$, and $\psi_{2,o} = a_{2,o} = 0$). Similar to the cases of unit-roots without drifts and with homogeneous drifts, we can show that under H_o^{HD}, the information matrix of the log-density (24) is singular. Stated formally:

Proposition 8: At $\phi^{HD} = \phi_*^{HD}$,

$$\ell_{RE,i,\delta}^{HD} - a_1 \ell_{RE,i,a_2}^{HD} - \psi_1 \ell_{RE,i,\psi_2}^{HD} + \nu \ell_{RE,i,\nu}^{HD} - (\omega_{11}+1)\ell_{RE,i,\omega_{11}}^{HD} - \omega_{11}\ell_{RE,i,\omega_{12}}^{HD} = 0.$$

Proposition 8 implies that similarly to Proposition 6, we can derive the asymptotic distribution of the ML estimator of ϕ^{HD}. The LR statistic for testing H_o^{HD} is a mixture of $\chi^2(5)$ and $\chi^2(4)$, $B\chi^2(5)+(1-B)\chi^2(4)$. The ML estimator of δ is $N^{1/2}$-consistent. This convergence rate is contrasted to $N^{1/6}$ of the GMM estimator of Moon and Phillips (2004). We are unable to identify the source of their different convergence rates. Indeed, their convergence rates are obtained under different settings: large N and fixed T for the ML estimator; and both large N and T for the GMM estimator. The two estimators could be better compared by investigating the asymptotic distribution of the ML estimator as T tends to infinity. We will leave this comparison to a future study.

4. Monte Carlo Experiments

In this section we consider the finite-sample properties of the RE and HPT ML estimators. We here consider the RE and HPT ML estimators only.[16] Alvarez and Arellano (2004) compare the efficiency gains of the RE estimator over the HPT and Lancaster estimators through calculations of population asymptotic variances. Thus, our results partly supplement theirs.

We also investigate the size and power properties of the *t*-test and the Likelihood Ratio (LR) test based on the RE and HPT estimators. Harris and Tzavalis (1999) proposed one-tailed *t*-tests for the hypothesis of unit root. The *t*-statistics for their tests, which are obtained by correcting the biases in within-type estimators, are asymptotically standard normal under the null hypothesis of unit root. For comparison, we also report the test results from their method. For our experiments, we consider three cases: the ML estimation without trend, with homogeneous trends, and with heterogeneous trends.

4.1. ML Estimation without Trend

[16]The Lancaster estimator seems to be quite sensitive to the choice of the starting parameter values used for algorithms. Some limited simulation results on the Lancaster estimator are available upon request from the authors.

In this subsection, we consider the finite-sample properties of the RE and HPT estimators computed with the assumption of no trend. The foundation of our Monte Carlo experiments here is the stationary data generating process (11), where $\eta_i \sim N(0,\sigma_\eta^2)$, $q_{i0} \sim N(0,\sigma_{q0}^2)$ and $\varepsilon_i \sim N(0,\nu)$. We set $\sigma_\eta^2 = 2$, $\sigma_{qo}^2 = 2$, $\nu = 1$, and $T = 5$. We try four different values, 0.5, 0.8, 0.9 and 1, for δ_o. Observe that when $\delta_o = 1$, the y_{it} are random walks without drifts. For each trial, we use 10,000 iterations. We consider two cases, $N = 500$ and 100, to examine both the large-sample and small-sample properties of the estimators.

Table 1 reports the bias and MSE (mean square error) of the RE ML estimator of δ . We also report the finite-sample size and power properties of the *t*-tests based on the RE ML estimation. For all true values of δ (δ_o), the biases of the ML estimator are small, even when $N =$ 100. MSE generally increases with δ_o when $\delta_o \le 0.9$, but it slightly decreases as δ_o increases from 0.9 to one. When δ_o is small, the *t*-test is properly sized. However, as δ_o increases, the test tends to over-reject correct hypotheses.

The power of the *t*-test to correctly reject the unit-root hypothesis increases with *N*. When $N = 100$, the power of the *t*-test is generally low especially when $0.8 \le \delta_o < 1$. When $\delta_o = 1$, the *t*-test rejects the correct unit-root hypothesis too much. This trend is not dependent on the sample size. This result is consistent with Proposition 6. The HPT estimator has similar properties as the RE estimator.

Figures 1-2 show the finite-sample distributions of the sampling errors of the RE estimator (that is, $(\hat{\delta}_{RE} - \delta_o)$). Figure 1 is for the case with $N = 500$ and Figure 2 is for the case with $N = 100$. In either figure, the distribution of the RE estimator becomes wider as δ_o increases. Up to $\delta_o = 0.9$, the estimator is roughly normally distributed. Then, when $\delta_o = 1$, the estimator is no longer normally distributed. Its distribution has a hump near $\delta = 1$ as Proposition 6 suggests.

Figures 3 compares the sampling error distributions of the RE and HPT ML estimators when $N = 100$. Somewhat surprisingly, the two estimators are similarly distributed. At our choice of parameter values, the efficiency gains of the RE estimator over the HPT estimator are not substantial. However, we find from unreported experiments that the efficiency gain of the RE estimator increases with the conditional variance of y_{i0} given the individual effects η_i (σ_{q0}^2 in our set-up).

We now examine the finite-sample size and power properties of the LR tests for all of the parametric restrictions implied by the hypothesis of unit root. In Table 2, LR-RE denotes the RE-based LR test for three parametric restrictions $(\delta_o, \omega_o, \psi_o)' = (1,0,0)'$. LR-HPT represents the HPT-based LR test for two restrictions $(\delta_o, \omega_{HPT,o}) = (1,0)'$, while the *t*-test of Harris and Tzavalis for the hypothesis of unit root without drifts is represented by T-HT. The asymptotic distribution of the LR-RE statistic is asymptotically a mixture of $\chi^2(3)$ and $\chi^2(2)$, while that of the LR-HPT statistic is a mixture of $\chi^2(2)$ and $\chi^2(1)$. For sensitivity analysis, we report test results obtained with five different combinations of the values of σ_η^2, σ_{qo}^2, and ν.

The LR-HPT test is generally better sized than the LR-RE and T-HT test, but by a small margin. For example, for our base choice of $\sigma_\eta^2 = 2$, $\sigma_{qo}^2 = 2$, and $\nu = 1$, when *N* = 500 and the normal size is 5%, the rejection rates of LR-RE, LR-HPT and T-HT are 5.40%, 5.03%, and 5.38%, respectively. Even when *N* = 100, both the LR-HPT and LR-HPT tests perform well. LR-RE tends to over-reject the unit root hypothesis, but the size of this distortion is not substantial, clearly smaller than that of the *t*-test reported in Table 1.

Table 2 also reports the finite-sample power property of the LR tests to reject the unit-root hypothesis. Not surprisingly, the power of each test increases with *N*. Overall, LR-RE has the highest power, with a few exceptions in the cases with *N* = 100 and the 1% significance level. The powers of LR-HPT and T-HT are generally compatible. The power of the LR-RE test depends on the sample size (*N*), the true value of δ (δ_o), and the relative sizes of the conditional variance of the initial values of y (σ_{q0}^2) and the variance of the random noises (ν). Their powers are positively related to the sample size and the conditional variance of the initial values of y (conditional on the unobservable effects η_i), while it is negatively related to the variance of random noises and the true value of δ. To be more specific, consider the case of *N* = 500. When δ_o = 0.9, the power of the LR-RE test is 100% or very close to it for all of the cases reported in Table 2. Even the true value of δ is closer to one ($\delta_o = 0.95$), the power remains reasonably high, when the conditional variance is relatively larger than the variance of the random noises. If we compare our base case of $(\sigma_\eta^2, \sigma_{qo}^2, \nu)$ = (2,2,1) with the cases of $(\sigma_{\eta}^2, \sigma_{qo}^2, \nu)$ = (2,1,1) and $(\sigma_{\eta}^2, \sigma_{qo}^2, \nu)$ = (2,2,4), we can see that the power of the LR-RE test is negatively related to the ratio of ν and σ_{qo}^2. The power is somewhat low when the ratio is large, especially when 1% of significance level is used for the test. But, the power of the LR test using 10% of significance level is greater than 89% for all of the cases reported in Table 2.

We now consider the power of the LR-RE test when N = 100. Clearly, the power is lower when N = 100 than when N = 500, especially when the ratio of ν and σ_{qo}^2 is large and the true of δ is close to one. But, when the ratio of ν and σ_{qo}^2 is small, the power of the test is reasonably high, even if the true value of δ is very close to one (see the case of $(\sigma_{\eta}^2, \sigma_{qo}^2, \nu)$ = (2,4,1)). For dynamic models, we can expect that the ratio of ν and σ_{qo}^2 should be low because the initial values of y is the weighted sum of the past random noises. The ratio would be even lower when the true value of δ is close to one. Thus, we can guess that for quite general cases, the LR test would have high power even in small samples.

Table 3 reports the finite-sample performances of the LR tests for the single parametric restriction $\delta_o = 1$ based on the RE and HPT estimators. Comparing Tables 2 and 3, we can see that the LR tests for the multiple restrictions implied by the hypothesis of unit root (*e.g.*, $(\delta_o, \omega_o, \psi_o) = (1,0,0)$ for the cases of the RE estimation) are generally better sized than the LR tests for the single restriction of $\delta_o = 1$. The former tests also have much higher power.

The general findings from our first Monte Carlo experiments can be summarized as follows. First, when the true value of δ is near one, the RE and HPT estimators do not follow usual normal distributions even if the sample size is large. This finding is consistent with our result in Proposition 6. Second, when the data follow the random walks without drifts, the *t*-tests based on the ML estimators reject correct unit-root hypothesis too often. In contrast, the LR tests based on the RE or HPT estimators perform well. Even when N is small, they are sized properly and have good power to reject the unit root hypothesis. Third, we find that in terms of the power of the LR test, use of the RE estimator instead of the HPT estimator is desirable especially when the initial values have a large variance. Finally, the LR test for the single restriction of $\delta_o = 1$ has low power. Use of the LR test for the multiple restrictions implied by the unit-root hypothesis is recommended.

4.2. ML Estimation with Homogenous Trends

We now consider the LR test for unit root based on the RE estimation with homogeneous trends. For simulations, data are generated following (21). The parametric restrictions we test are given $H_o : (\delta_o, \omega_o, \psi_o, \beta_o)' = (1,0,0,0)'$. Table 4 reports the rejection rates of the LR test determined by using the mixed distribution of $\chi^2(3)$ and $\chi^2(4)$. Again as with Tables 2 and 3, we report the results we obtain with different combinations of the values of σ_η^2, σ_{qo}^2, and ν. For all simulations, the trend parameter *g* in (21) is fixed at 0.5. The choice of different values of *g* alters simulation outcomes only immaterially. Six different combinations of *N* and *T* are used. For each of *N* = 100 and 200, we try three different values for the number of time series observations: *T* = 5, 10, and 25. We do so because the size of *T* would be an important factor determining the finite-sample properties of the estimators for the models with time trends.

For the cases with *T* = 5, the results reported in Table 4 are very comparable to those in Table 2. The introduction of homogeneous drift does not lead to a large change in the size or power of the LR test. The power of the LR tests becomes higher as either *N* or *T* increases. For our base choice of $\sigma_\eta^2 = 2$, $\sigma_{qo}^2 = 2$, $\nu = 1$, and *T* = 5, when *N* = 500 and the normal size is 5%, the rejection rate of the LR test based on the RE estimator is 4.77%. Thus, the LR test is again reasonably well sized. In addition, the test has a power to reject unit root even when δ_o is close to one. As with the no-drift simulations when *N* = 100, the LR test performs well. It tends to slightly over-reject the unit root hypothesis but retains a good power property even when *N* is small, especially when the variance of the initial values y_{i0} or the number of time series observations are large.

4.3. Random Walks with Heterogeneous Drifts

In this subsection, we examine the finite-sample properties of the RE ML estimator computed under the assumption of heterogeneous trends. The data generating process we use is given:

$$y_{it} = \eta_i + g_i t + q_{it}; \quad q_{it} = \delta q_{i,t-1} + \varepsilon_{it}, \quad t = 0, ..., T,$$

where $\varepsilon_i \sim N(0_{T\times 1}, \nu I_T)$, η_i and q_{i0} are also normally distributed with zero means, and all of ε_i, η_i, q_{i0} and g_i are mutually independent. The mean of the normally distributed trends g_i is fixed at 0.5 while their variance is set at three different values, 0, 0.5, and 1. The data are generated with $\sigma_\eta^2 = 2$, $\sigma_{qo}^2 = 2$, $\nu = 1$, our default case from the previous two experiments. Note that when $\delta_o = 1$ and $\sigma_g^2 > 0$, the data are random walks with heterogeneous drifts. When $\delta_o = 1$ and $\sigma_g^2 = 0$, the data are random walks with homogeneous drifts. As with the pervious experiments we use 10,000 iterations and six different combinations of *N* and *T* are tried.

To save space, we only consider the finite-sample properties of the LR test based on the RE estimation (LR-RE). For each simulation, the five parametric restrictions imposed on (27), $(\delta_o, \omega_{12,o}, \omega_{22,o}, \psi_{2,o}, a_{2,o})' = (1,0,0,0,0)'$, all of which are implied by the hypothesis of unit root with heterogeneous drifts, are jointly tested. Thus, under the null hypothesis of unit roots with heterogeneous trends, the LR statistic has the asymptotic distribution mixing $\chi^2(5)$ and $\chi^2(4)$. For comparison, we also consider another one-tailed *t*-test developed by Harris and Tzavalis (T-HT). The *t*-test is designed for testing the hypothesis of unit root allowing for heterogeneous drifts. The simulation results are reported in Table 5.

Since we use $(\sigma_\eta^2, \sigma_{qo}^2, \nu)$ = (2,2,1) to generate the results in Table 4, the data generating processes are the same for the case of $\sigma_g^2 = 0$ in Table 5 and the case of $(\sigma_\eta^2, \sigma_{qo}^2, \nu)$ = (2,2,1) in Table 4. Comparing these two cases, we can see that when *T* is small, the power property of the LR test based on the RE ML estimation with heterogeneous trends is quite different from that of the LR test based on the estimation with homogeneous trends. For example, for the cases with *T* = 5 and $\delta_o = 0.9$, the rejection rates of the LR test from the estimation with homogeneous time trends are all 100% at the 1%, 5%, 10% of normal sizes (Table 4). In contrast, the same rates of the LR test from the estimation allowing heterogeneous trends are merely 2.35%, 9.98% and 17.28, respectively (Table 5). For the estimation allowing heterogeneous trends, the power of the LR test improves only mildly when *N* increases from 100 to 500. However, the power increases in a large scale as *T* increases. We observe the similar results from the cases in which heterogeneous trends exist in data ($\sigma_g^2 > 0$).

Tables 5 also report the finite-sample performances of the *t*-test by Harris and Tzavalis (T-HT). Similarly to the LR test, the *t*-test has low power to reject the hypothesis of unit roots.

Overall, the t-test show the performances compatible to those of the LR tests.

5. Concluding Remark

This paper has considered the asymptotic and finite-sample properties of a random effects ML estimator and the two fixed-effects ML estimators of Lancaster (2002) and Hsiao, Pesaran and Tahmiscioglu (2002, HPT). When data are stationary, the random effects ML estimator is asymptotically more efficient than the other two fixed effects ML estimators when both the individual effects and the initial observations are normal. We also consider the asymptotic distributions of the random effects and HPT ML estimators when data contain unit roots. The two estimators, as well as the Lancaster estimator, are non-normal. This distortion results because the information matrices of the estimators are singular when data contain unit root. Thus, the *t*-tests for unit root are inappropriate. Consistent with the prediction of Rotnitzky, Cox, Bottai and Robins (2000), the Likelihood-Ratio tests for unit root with the *p*-values from the mixed chi-square distributions perform much better than the *t*-tests. They also have good power properties even if the number of observations is small.

Our results depend on the assumption of homoskedastic random errors (ε_{it}). Alvarez and Arellano (2004) have shown that when the errors are heteroskedastic or autocorrelated over time, the ML estimator computed incorporating such heteroskedasticity and autocorrelations is generally consistent and asymptotically normal even if data may follow random walks. However, Thomas (2005) found that the finite-sample distribution of the ML estimator considerably deviates from the normal distribution, if data follow unit root processes and the error variances change only slowly over time. In addition, there are some special cases in which the information matrix of the ML estimator becomes singular. An example is the case in which only the variance of the random error at the last time period is different from the variances at other time periods. Another example is the case in which the random errors follow a MA(1) process with the MA coefficient equal to one (Thomas, 2005). We are currently working on such special cases.

Our results also convey a message to the studies of Monte Carlo experiments on dynamic panel data models. The homoskedasticity assumption is not an innocuous simplifying assumption for the models. The asymptotic and finite-sample properties of the estimators developed for the models could crucially depend on whether or not the homoskedasticity assumption holds.

Appendix

Proof of Equation (12): Define $P_T = T^{-1}1_T 1_T'$, $Q_T = I_T - P_T$; and

$$D_{T-1}' = \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ : & : & : & & : & : \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}_{(T-1)\times T},$$

where 1_T is a (T×1) vector of ones. Observe that $D_{T-1}'(y_{i1},\ldots,y_{iT})' = \Delta y_i$. Notice that $D_{T-1}' 1_T = 0_{(T-1)\times 1}$, Thus, by Rao (1973, p. 77), we have $Q_T = D_{T-1}B_{T-1}^{-1}D_{T-1}'$. Also, observe that $\overline{\Delta_0 y_i} \equiv T^{-1}\Sigma_{t=1}^T (y_{it} - y_{i0}) = \Delta y_{i1} + k_{T-1}'\Delta y_i$. Similary, $\overline{\Delta_0 y_{i,-1}} = k_{T-1}'\Delta y_{i,-1}$. Using these results, we obtain

$$\begin{aligned}
&(\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T)'(\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T) \\
&\quad = (\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T)' Q_T (\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T) \\
&\quad\quad + (\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T)' P_T (\Delta_0 y_i - \Delta_0 y_{i,-1}\delta - p_i 1_T) \\
&\quad = (y_i - y_{i,-1}\delta)' Q_T (y_i - y_{i,-1}\delta) + T(\overline{\Delta_0 y_i} - \overline{\Delta_0 y_{i,-1}}\delta - p_i)^2 \\
&\quad = (\Delta y_i - \Delta y_{i,-1}\delta)' B_{T-1}^{-1} (\Delta y_i - \Delta y_{i,-1}\delta) + T(\Delta y_{i1} + k_{T-1}'(\Delta y_i - \Delta y_{i,-1}\delta) - p_i)^2.
\end{aligned}$$

Proof of Proposition 1: A straight algebra shows

$$\begin{aligned}
&f(\Delta_0 y_i \mid \delta, \nu, p_i) f(p_i \mid \omega_{HPT}, \nu) \\
&\quad = \frac{1}{(2\pi)^{T/2}\nu^{T/2}} \exp\left[-\frac{1}{2}\left(\Delta y_i - \Delta y_{i,-1}\delta\right)' B_{T-1}^{-1} \left(\Delta y_i - \Delta y_{i,-1}\delta\right) - \frac{T}{2\nu}(p_i - g_i)^2 \right] \\
&\quad\quad \times \frac{1}{(2\pi)^{1/2}\nu^{1/2}(\omega_{HPT})^{1/2}} \exp\left(-\frac{1}{2\nu\omega_{HPT}} p_i^2 \right) \\
&\quad = \frac{1}{(2\pi)^{T/2}\nu^{T/2}} \exp\left[-\frac{1}{2}\left(\Delta y_i - \Delta y_{i,-1}\delta\right)' B_{T-1}^{-1} \left(\Delta y_i - \Delta y_{i,-1}\delta\right) \right] \\
&\quad\quad \times \frac{1}{(2\pi)^{1/2}\nu^{1/2}(\omega_{HPT})^{1/2}} \exp\left[-\frac{T}{2\nu} p_i^2 - \frac{1}{2\nu\omega_{HPT}} p_i^2 + \frac{T}{\nu} p_i g_i - \frac{T}{2\nu} g_i^2 \right].
\end{aligned}$$

where $g_i = \Delta y_{i1} + k_{T-1}'(\Delta y_i - \Delta y_{i,-1}\delta)$. We can also show:

$$\frac{1}{(2\pi)^{1/2}v^{1/2}(\omega_{HPT})^{1/2}}\exp\left[-\frac{T}{2v}p_i^2-\frac{1}{2v\omega_{HPT}}p_i^2+\frac{T}{v}p_ig_i-\frac{T}{2v}g_i^{\ 2}\right]$$

$$=(\xi_{HPT,T})^{-1/2}\frac{(\xi_{HPT,T})^{1/2}}{(2\pi)^{1/2}v^{1/2}(\omega_{HPT})^{1/2}}\exp\left[-\frac{\xi_{HPT,T}}{2v\omega_{HPT}}\left(p_i-\frac{T\omega_{HPT}}{\xi_{HPT,T}}g_i\right)^2\right]$$

$$\times\exp\left[-\frac{T}{2v\xi_{HPT,T}}g_i^{\ 2}\right].$$

Thus,

$$\int_{-\infty}^{\infty}f(\Delta_0 y_i\,|\,\delta,v,p_i)f(p_i\,|\,\omega_{HPT},v)dp_i$$

$$=\frac{1}{(2\pi)^{T/2}v^{T/2}(\xi_{HPT,T})^{1/2}}\exp\left[-\frac{1}{2v}(\Delta y_i-\Delta y_{i,-1}\delta)'B_{T-1}^{-1}(\Delta y_i-\Delta y_{i,t-1}\delta)-\frac{T}{2v\xi_{HPT,T}}g_i^2\right]$$

$$\times\int_{-\infty}^{\infty}\frac{(\xi_{HPT,T})^{1/2}}{(2\pi)^{1/2}v^{1/2}(\omega_{HPT})^{1/2}}\exp\left[-\frac{\xi_{HPT,T}}{2v\omega_{HPT}}\left(p_i-\frac{T\omega_{HPT}}{\xi_{HPT,T}}g_i^2\right)^2\right]dr_i$$

$$=\frac{1}{(2\pi)^{T/2}v^{T/2}(\xi_{HPT,T})^{1/2}}\exp\left[-\frac{1}{2v}(\Delta y_i-\Delta y_{i,-1}\delta)'B_{T-1}^{-1}(\Delta y_i-\Delta y_{i,t-1}\delta)-\frac{T}{2v\xi_{HPT,T}}g_i^2\right]$$

$$=f(r_i\,|\,\delta,\omega_{HTP},v).$$

The following lemmas are useful to prove the propositions in Section 3. The first two lemmas provide alternative forms of $\Omega_T(\omega)$ and $[\Omega_T(\omega)]^{-1}$.

Lemma A.1: Let $\Omega_s(\omega)$ be the *s*×*s* (*s* = 2, ..., *T*) matrix of the form of $\Omega_T(\omega)$ in Section 2. Let:

$$L_s'=\begin{pmatrix}1&0&0&\cdots&0&0\\-1&1&0&\cdots&0&0\\0&-1&1&\cdots&0&0\\\vdots&\vdots&\vdots&&\vdots&\vdots\\0&0&0&\cdots&1&0\\0&0&0&\cdots&-1&1\end{pmatrix}_{s\times s};\quad c_s=\begin{pmatrix}1\\0\\0\\0\\\vdots\\0\end{pmatrix}_{s\times s}.$$

Then, $\Omega_s(\omega)=L_s'L_s+\omega c_sc_s'$.

Lemma A.2: $[\Omega_T(\omega)]^{-1} = (L_T' L_T)^{-1} - \frac{T\omega}{T\omega+1}\bar{k}_T \bar{k}_T{}'$, where $\bar{k}_T{}' = (1, k_{T-1}')'$.

Lemma A.3: $B_{T-1}^{-1} = (\tilde{D}_{T-1}' \tilde{D}_{T-1})^{-1} - T m_{T-1} m_{T-1}'$, where $m_{T-1} = T^{-1}(1,2,...,T-1)$ and $\tilde{D}_{T-1}$ is the square matrix of the first (*T*-1) columns of D_{T-1}'. In addition, $trace\left(B_{T-1}^{-1}\right) = (T-1)(T+1)/6$; and $trace\left(k_{T-1}k_{T-1}'\right) = (T-1)(2T-1)/(6T)$.

Lemma A.4: The first-order and second-order derivatives of $\ell_{RE,i}(\theta)$ are given:

$$\ell_{RE,i,\delta} = \frac{1}{\nu}\Delta y_{i,-1}' B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta) + \frac{T}{\nu \xi_T}\Delta y_{i,-1}' k_{T-1} d_i(\psi,\delta);$$

$$\ell_{RE,i,\omega} = -\frac{T}{2\xi_T} + \frac{T^2}{2\nu\xi_T{}^2}(d_i(\psi,\delta))^2;$$

$$\ell_{RE,i,\nu} = -\frac{T}{2\nu} + \frac{1}{2\nu^2}\left(\Delta y_i - \Delta y_{i,-1}\delta\right)' B_{T-1}^{-1}\left(\Delta y_i - \Delta y_{i,-1}\delta\right) + \frac{T}{2\nu^2\xi_T}\left(d_i(\psi,\delta)\right)^2;$$

$$\ell_{RE,i,\psi} = \frac{T}{\nu\xi_T} y_{i0} d_i(\psi,\delta);$$

$$\ell_{RE,i,\delta\delta} = -\frac{1}{\nu}\Delta y_{i,-1}' B_{T-1}^{-1}\Delta y_{i,-1} - \frac{T}{\nu\xi_T}\Delta y_{i,-1}' k_{T-1} k_{T-1}' \Delta y_{i,-1};$$

$$\ell_{RE,i,\delta\nu} = -\frac{1}{\nu^2}\Delta y_{i,-1}' B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta) - \frac{T}{\nu^2\xi_T}\Delta y_{i,-1}' k_{T-1} d_i(\psi,\delta);$$

$$\ell_{RE,i,\delta\omega} = -\frac{T^2}{\nu\xi_T{}^2}\Delta y_{i,-1}' k_{T-1} d_i(\psi,\delta);\quad \ell_{RE,i,\delta\psi} = -\frac{T}{\nu\xi_T}\Delta y_{i,-1}' k_{T-1} y_{i0};$$

$$\ell_{RE,i,\nu\nu} = \frac{T}{2\nu^2} - \frac{1}{\nu^3}(\Delta y_i - \Delta y_{i,-1}\delta)' B_{T-1}^{-1}(\Delta y_i - \Delta y_{i,-1}\delta) - \frac{T}{\nu^3\xi_T}\left(d_i(\psi,\delta)\right)^2;$$

$$\ell_{RE,i,\nu\omega} = -\frac{T^2}{2\nu^2\xi_T{}^2}\left(d_i(\psi,\delta)\right)^2;\quad \ell_{RE,i,\nu\psi} = -\frac{T}{\nu^2\xi_T} d_i(\psi,\delta) y_{i0};$$

$$\ell_{RE,i,\omega\omega} = \frac{T^2}{2\xi_T{}^2} - \frac{T^3}{\nu\xi_T{}^3}\left(d_i(\psi,\delta)\right)^2;\quad \ell_{RE,i,\lambda\psi} = -\frac{T^2}{\nu\xi_T{}^2} d_i(\psi,\delta) y_{i0};\quad \ell_{RE,i,\psi\psi} = -\frac{T}{\nu\xi_T}(y_{i0})^2,$$

where $d_i(\psi,\delta) = \Delta y_{i1} - \psi y_{i0} + k_{T-1}'(\Delta y_i - \delta\Delta y_{i,-1})$.

Lemma A.5: Under the RE assumption,

$$E\left(\Delta y_{i,-1}' B_{T-1}^{-1} \Delta y_{i,-1}\right) = \psi^2 \sigma_{y_0}^2 c_{T-1}' H_{T-1}(\delta_o) B_{T-1}^{-1} H_{T-1}(\delta_o)' c_{T-1} + \nu \times trace\left(B_{T-1}^{-1} H_{T-1}(\delta_o)' \Omega_{T-1,o} H_{T-1}(\delta_o)\right); \tag{A.1}$$

$$E\left(\Delta y_{i,-1}' k_{T-1} k_{T-1}' \Delta y_{i,-1}\right) = \psi^2 \sigma_{y_0}^2 c_{T-1}' H_{T-1}(\delta_o) k_{T-1} k_{T-1}' H_{T-1}(\delta_o)' c_{T-1} + \nu_o \times trace\left(k_{T-1}' H_{T-1}(\delta_o)' \Omega_{T-1,o} H_{T-1}(\delta_o) k_{T-1}\right); \tag{A.2}$$

$$E\left(\Delta y_{i,-1}' B_{T-1}^{-1} \Delta \varepsilon_i\right) = -\nu_o b(\delta_o)'; \tag{A.3}$$

$$E\left(\Delta y_{i,-1}' k_{T-1}\left(\Delta y_{i1} - \psi y_{i0} + k_{T-1}'(\Delta y_i - \Delta y_{i,-1}\delta_o)\right)\right) = \frac{\nu_o \xi_{T,o}}{T} b(\delta_o)'; \tag{A.4}$$

$$E\left[(\Delta y_i - \Delta y_{i,-1}\delta_o)' B_{T-1}^{-1} (\Delta y_i - \Delta y_{i,-1}\delta_o)\right] = (T-1)\nu_o; \tag{A.5}$$

$$E[(\Delta y_{i1} - \psi_o y_{i0} + k_{T-1}'(\Delta y_i - \Delta y_{i,-1}\delta_o))^2] = \frac{\xi_{T,o}\nu_o}{T}, \tag{A.6}$$

where $\xi_{T,0} = T\omega_o + 1$, $\Omega_{T-1,o} = \Omega_{T-1}(\omega_o)$, $b(\delta)$ is defined in (13), $b(\delta)' = db/d\delta$, and

$$H_{T-1}(\delta)' = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \delta & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \delta^{T-3} & \delta^{T-4} & \delta^{T-5} & \dots & 0 \\ \delta^{T-2} & \delta^{T-3} & \delta^{T-4} & \dots & 1 \end{pmatrix}_{(T-1)\times(T-1)}.$$

Proof: Define $h_{i,T-1} = (u_i, \Delta\varepsilon_{i2}, \dots, \Delta\varepsilon_{i,T-1})$, so that $Var(h_{i,T}) = \nu_o \Omega_{T-1,o}$, and

$$\Delta y_{i,-1} = H_{T-1}(\delta_o)' c_{T-1} \psi_o y_{i0} + H_{T-1}(\delta_o)' h_{i,T-1}.$$

Then,

$$\begin{aligned} &E\left(\Delta y_{i,-1}' B_{T-1}^{-1} \Delta y_{i,-1}\right) \\ &= \psi^2 \sigma_{y_0}^2 c_{T-1}' H_{T-1}(\delta_o) B_{T-1}^{-1} H_{T-1}(\delta_o)' c_{T-1} + E\left(h_{i,T-1}' H_{T-1}(\delta_o) B_{T-1}^{-1} H_{T-1}(\delta_o)' h_{i,T-1}\right) \\ &= \psi^2 \sigma_{y_0}^2 c_{T-1}' H_{T-1}(\delta_o) B_{T-1}^{-1} H_{T-1}(\delta_o)' c_{T-1} + tr\left(E\left(H_{T-1}(\delta_o) B_{T-1}^{-1} H_{T-1}(\delta_o)' h_{i,T-1} h_{i,T-1}'\right)\right) \\ &= \psi^2 \sigma_{y_0}^2 c_{T-1}' H_{T-1}(\delta_o) B_{T-1}^{-1} H_{T-1}(\delta_o)' c_{T-1} + \nu_o \times tr\left(B_{T-1}^{-1} H_{T-1}(\delta_o)' \Omega_{T-1,o} H_{T-1}(\delta_o)\right). \end{aligned}$$

(A.2)-(A.6) can be obtained by the similar method.

Lemma A.6: Under the RE assumption,

$$E\left(-\ell_{RE,i,\theta\theta}(\theta_o)\right)=\begin{pmatrix} B+A & 0 & \frac{Tb(\delta_o)'}{\xi_{T,o}} & \frac{T\psi_o\sigma_{y_0}^2}{\nu_o\xi_{T,o}}b(\delta_o)' \\ 0 & \frac{T}{2\nu_o{}^2} & \frac{T}{2\nu\xi_{T,o}} & 0 \\ \frac{Tb(\delta_o)'}{\xi_{T,o}} & \frac{T}{2\nu_o\xi_{T,o}} & \frac{T^2}{2\xi_{T,o}^2} & 0 \\ \frac{T\psi\sigma_{y_0}^2}{\nu_o\xi_{T,o}}b(\delta_o)' & 0 & 0 & \frac{T\sigma_{y_0}^2}{\nu_o\xi_{T,o}} \end{pmatrix}, \qquad (A.7)$$

where,

$$A = tr\left(B_{T-1}^{-1}H_{T-1}(\delta_o)'\Omega_{T-1,0}H_{T-1}(\delta_o)\right)+\frac{T}{\xi_{T,o}}trace\left(k_{T-1}'H_{T-1}(\delta_o)'\Omega_{T-1,o}H_{T-1}(\delta_o)k_{T-1}\right);$$

$$B=\frac{1}{\nu_o}\psi^2\sigma_{y_0}^2c_{T-1}'H_{T-1}(\delta_o)B_{T-1}^{-1}H_{T-1}(\delta_o)'c_{T-1}+\frac{T}{\xi_{T,0}\nu_0}\psi^2\sigma_0^2c_{T-1}'H_{T-1}(\delta_o)'k_{T-1}k_{T-1}'H_{T-1}(\delta_o)'c_{T-1}.$$

Proof of Proposition 2: When $\theta_o=\theta_*=(1,\nu_*,0,0)'$, Lemma A.1 and the definition of $H_{T-1}(\delta)$ in Lemma 5 imply that $H_{T-1}(1)'\Omega_{T-1}(0)H_{T-1}(1)=I_{T-1}$. This is so because $\Omega_{T-1}(0)=L_{T-1}'L_{T-1}$ and $H_{T-1}(1)=(L_{T-1}')^{-1}$. In addition, in (A.7), B = 0 because $\psi_o=0$. Using Lemma A.3 and the fact that $\xi_{T,o}=1$ when $\theta_o=\theta_*$, we can show A = T(T-1)/2 = Tb(1)′. Substituting these results, $\psi_o=0$, and $\xi_{T,o}=1$, into (A.7), we have:

$$E\left(-\ell_{RE,i,\theta\theta}(\theta_*)\right)=\begin{pmatrix} Tb(1)' & 0 & Tb(1)' & 0 \\ 0 & \frac{T}{2\nu_o{}^2} & \frac{T}{2\nu_o} & 0 \\ Tb(1)' & \frac{T}{2\nu_o} & \frac{T^2}{2} & 0 \\ 0 & 0 & 0 & \frac{T\sigma_{y_0}^2}{\nu_o} \end{pmatrix}, \qquad (A.8)$$

which has a zero determinant. The likelihood theory indicates $E\left(H_{RE,i}(\theta_*)+B_{RE,i}(\theta_*)\right)=0$, at $\theta_o=\theta_*$. Thus, $E\left(B_{RE,i}(\theta_*)\right)$ must be also singular.

Proof of Proposition 3: We first derive $E(\ell_{RE,i}(\theta))$ under H_o^{UND}. We can show that

$$E\left(\left(d_i(\psi,\delta)\right)^2\right)=\nu_*\frac{(T+1)(2T+1)}{6T}-2\nu_*\frac{(T-1)(T+1)}{3T}\delta+\sigma_{y_0}^2\psi^2+\nu_*\frac{(T-1)(2T-1)}{6T}\delta^2;$$

$$E\left(\Delta\varepsilon_i(\delta)'B_{T-1}^{-1}\Delta\varepsilon_i(\delta)\right)=\frac{\nu_*}{6}\left((T-1)(T+1)-2(T-1)(T-2)\delta+(T-1)(T+1)\delta^2\right),$$

where $d_i(\psi,\delta)=\Delta y_{i1}-\psi y_{io}+k_{T-1}'(\Delta y_i-\Delta y_{i,-1}\delta)$ and $\Delta\varepsilon_i(\delta)=\Delta y_i-\Delta y_{i,-1}\delta$. Thus, we have

$$\begin{aligned}E\left(\ell_{RE,i}(\delta,\nu,\omega,\psi)\right)=&-\frac{T}{2}\ln(\pi)-\frac{T}{2}\ln(\nu)-\frac{1}{2}\ln\xi_T\\&-\frac{1}{12}\frac{\nu_*}{\nu}\left((T-1)(T+1)-2(T-1)(T-2)\delta+(T-1)(T+1)\delta^2\right)\\&-\frac{1}{12}\frac{\nu_*}{\nu\xi_T}\begin{pmatrix}(T+1)(2T+1)-4(T-1)(T+1)\delta\\+\sigma_{y_0}^2\psi^2+(T-1)(2T-1)\delta^2\end{pmatrix},\end{aligned}$$

where $\xi_T=T\omega+1$. We now concentrate out ψ, ω, and ν from $E(\ell_{RE,i}(\delta,\nu,\omega,\psi))$. It is obvious that $\psi=0$ maximizes $E(\ell_{RE,i}(\delta,\nu,\omega,\psi))$. Thus we have the concentrated value of $E(\ell_{RE,i}(\delta,\nu,\omega,\psi))$:

$$\begin{aligned}E(\ell_{RE,i}^c(\delta,\nu,\omega))=&-\frac{T}{2}\ln(\pi)-\frac{T}{2}\ln(\nu)-\frac{1}{2}\ln\xi_T\\&-\frac{1}{12}\frac{\nu_*}{\nu}\left((T-1)(T+1)-2(T-1)(T-2)\delta+(T-1)(T+1)\delta^2\right)\\&-\frac{1}{12}\frac{\nu_*}{\nu\xi_T}E\left((T+1)(2T+1)-4(T-1)(T+1)\delta+(T-1)(2T-1)\delta^2\right).\end{aligned}$$

Now, solve the first-order condition with respect to ξ_T (instead of ω):

$$\frac{\partial E(\ell_{RE,i}^c(\delta,\nu,\omega))}{\partial\xi_T}=-\frac{1}{2\xi}+\frac{1}{12}\frac{\nu_*}{\nu\xi_T{}^2}\begin{pmatrix}(T+1)(2T+1)-4(T-1)(T+1)\delta\\+(T-1)(2T-1)\delta^2\end{pmatrix}=0.$$

Then we obtain:

$$\xi_T=\frac{1}{6}\frac{\nu_*}{\nu}\left((T+1)(2T+1)-4(T-1)(T+1)\delta+(T-1)(2T-1)\delta^2\right).$$

Thus,

$$\begin{aligned}E(\ell_{RE,i}^c(\delta,\nu))=&-\frac{T}{2}\ln(\pi)-\frac{T}{2}\ln(\nu)-\frac{1}{2}\ln\left(\frac{\nu_*}{\nu}\right)-\frac{1}{2}\ln\left(\frac{1}{6}\right)\\&-\frac{1}{2}\ln\left(\frac{(T+1)(2T+1)}{6}-2\frac{(T-1)(T+1)}{3}\delta+\frac{(T-1)(2T-1)}{6}\delta^2\right)\\&-\frac{1}{12}\frac{\nu_*}{\nu}\left((T-1)(T+1)-2(T-1)(T-2)\delta+(T-1)(T+1)\delta^2\right).\end{aligned}$$

Now, solving

$$\frac{\partial E(\ell^c_{RE,i}(\delta,\nu))}{\partial \nu} = -\frac{T-1}{2\nu} + \frac{1}{12}\frac{\nu_*}{\nu^2}\left((T-1)(T+1) - 2(T-1)(T-2)\delta + (T-1)(T+1)\delta^2\right) = 0,$$

we have:

$$\nu = \nu_* \frac{1}{T-1}\left(\frac{(T-1)(T+1)}{6} - 2\frac{(T-1)(T-2)}{6}\delta + \frac{(T-1)(T+1)}{6}\delta^2\right).$$

Substituting this solution to $E(\ell^c_{RE,i}(\delta,\nu))$ yields

$$\begin{aligned} E(\ell^c_{RE,i}(\delta)) = &-\frac{T}{2}\ln(\pi) - \frac{T}{2}\ln(\nu_*) - \frac{T-1}{2}\ln\left(\frac{1}{T-1}\right) - \frac{1}{2}\ln\left(\frac{1}{6}\right) \\ &-\frac{T-1}{2}\ln\left(\frac{(T-1)(T+1)}{6} - 2\frac{(T-1)(T-2)}{6}\delta + \frac{(T-1)(T+1)}{6}\delta^2\right) \\ &-\frac{1}{2}\ln\left(\frac{(T+1)(2T+1)}{6} - 2\frac{(T-1)(T+1)}{3}\delta + \frac{(T-1)(2T-1)}{6}\delta^2\right). \end{aligned}$$

Then, a little algebra shows:

$$\frac{\partial E(\ell^c_{RE,i}(\delta))}{\partial \delta} = -\frac{(\delta-1)^3 T(T+1)(2T-1)}{\left[(2T-1)\left\{\delta - \frac{2(T+1)}{2T-1}\right\}^2 + \frac{3}{2T-1}\right]\left[(T+1)\left\{\delta - \frac{T-2}{T+1}\right\}^2 + \frac{3(T-1)}{T+2}\right]}.$$

Since the denominator of $\partial E(\ell^c_{RE,i}(\delta))/\partial\delta$ is positive for any $T \geq 2$, it is positive for $\delta < 1$ and negative for $\delta = 1$. The derivative equals zero only at $\delta = 1$. This indicates that $\delta = 1$ is the global maximum point.

Proof of Proposition 4: It can be shown that when $\theta_o = \theta_*$,

$$\begin{aligned} E\left(\ell_{Lan,i}(\delta,\nu)\right) = b(\delta) &- \frac{T-1}{2}\ln(\nu) \\ &- \frac{\nu_*}{2\nu}\left(\frac{(T-1)(T+1)}{6} - \frac{(T-1)(T-2)}{3}\delta + \frac{(T-1)(T+1)}{6}\delta^2\right). \end{aligned}$$

Without loss of generality we set $\nu_* = 1$. Then, the first-order maximization condition with respect to ν yields

$$\nu = \frac{1}{T-1}\left(\frac{(T-1)(T+1)}{6} - \frac{(T-1)(T-2)}{3}\delta + \frac{(T-1)(T+1)}{6}\delta^2\right).$$

Substituting this into the expected value of $E(\ell_{Lan,i}(\delta,\nu))$, we can get

$$E\left(\ell^c_{Lan,i}(\delta)\right) = b(\delta) - \frac{T-1}{2}\ln\left(\frac{1}{T-1}\right) - \frac{T-1}{2}$$
$$-\frac{T-1}{2}\ln\left(\frac{(T-1)(T+1)}{6} - \frac{(T-1)(T-2)}{3}\delta + \frac{(T-1)(T+1)}{6}\delta^2\right).$$

If we differentiate the concentrated likelihood function, we yield

$$\frac{\partial E\left(\ell^c_{Lan,i}(\delta)\right)}{\partial \delta} = b(\delta)' - \frac{(T-1)\left((T+1)\delta - T(T-2)\right)}{(T+1)\delta^2 - 2T(T-2)\delta + (T+1)};$$

$$\frac{\partial^2 E\left(\ell^c_{Lan,i}(\delta)\right)}{\partial \delta^2} = b(\delta)'' - \frac{2(T-1)\left(T(\delta-1)+\delta+2\right)^2}{\left(\left(T(\delta-2)+4+\delta\right)\delta+(T+1)\right)^2}$$
$$+\frac{(T-1)(T+1)}{\left(\left(T(\delta-2)+4+\delta\right)\delta+(T+1)\right)}.$$

$$\frac{\partial^3 E\left(\ell^c_{Lan,i}(\delta)\right)}{\partial \delta^3} = b(\delta)''' + \frac{8(T-1)\left(T(\delta-1)+\delta+2\right)^3}{\left(\left(T(\delta-2)+4+\delta\right)\delta+(T+1)\right)^3}$$
$$-\frac{6(T-1)(T+1)\left(T(\delta-1)+\delta+2\right)}{\left((T+1)\delta^2 - 2(T-2)\delta + (T+1)\right)^2}.$$

It can be shown that when $\delta = 1$,

$$b(1)' = \frac{T-1}{2}; b(1)'' = \frac{(T-1)(T-2)}{6}; b(1)''' = \frac{(T-1)(T-2)(T-3)}{12}.$$

Using these results, we can easily show that at $\delta = 1$,

$$\frac{\partial E\left(\ell^c_{Lan,i}(1)\right)}{\partial \delta} = \frac{\partial^2 E\left(\ell^c_{Lan,i}(1)\right)}{\partial \delta^2} = 0,$$

but,

$$\frac{\partial^3 E\left(\ell^c_{Lan,i}(1)\right)}{\partial \delta^3} = \frac{(T-1)(T-2)(T-3)}{12} - \frac{(T-1)(T+1)}{2} \neq 0.$$

Thus, $\delta = 1$ is an inflexion point of $E\left(\ell^c_{Lan,i}(\delta)\right)$.

Proof of Proposition 5: At $\theta = \theta_*$, using the alternative representation of $[\Omega_T(\omega)]^{-1}$ given in Lemma A.2, we can have

$$\ell_{RE,i,\nu}(\theta_*) = -\frac{T}{2\nu} + \frac{1}{2\nu^2}\begin{pmatrix} \Delta y_{i1} \\ \Delta y_i - \Delta y_{i,-1} \end{pmatrix}' (L_T' L_T)^{-1} \begin{pmatrix} \Delta y_{i1} \\ \Delta y_i - \Delta y_{i,-1} \end{pmatrix}$$
$$= -\frac{T}{2\nu} + \frac{1}{2\nu^2} r_i' L_T (L_T' L_T)^{-1} L_T' r_i = -\frac{T}{2\nu} + \frac{1}{2\nu^2}\Sigma_i r_i' r_i,$$

where $r_i = (\Delta y_{i1}, ..., \Delta y_{iT})'$ and $\nu = \nu_*$ Also, we have:

$$\ell_{RE,i,\omega}(\theta_*) = -\frac{T}{2} + \frac{T^2}{2\nu}\left(\Delta y_{i1} + k_{T-1}'(\Delta y_i - \Delta y_{i,-1})\right)^2 = -\frac{T}{2} + \frac{T^2}{2\nu}\left(\bar{k}_T' L_T' r_i\right)^2$$
$$= -\frac{T}{2} + \frac{T^2}{2\nu} r_i' L_T \bar{k}_T \bar{k}_T' L_T' r_i = -\frac{T}{2} + \frac{1}{2\nu} r_i' 1_T 1_T' r_i.$$

Thus,

$$\ell_{RE,i,\omega}(\theta_*) - \nu \ell_{RE,i,\nu}(\theta_*) = -\frac{1}{2\nu} r_i'(I_T - 1_T 1_T') r_i = \frac{1}{\nu}\Sigma_{s=2}^{T}\Sigma_{t=1}^{s-1}\Delta y_{it}\Delta y_{is}.$$

Now,

$$\ell_{RE,i,\delta}(\theta_*) = \frac{1}{\nu}\begin{pmatrix} 0 \\ \Delta y_{i,-1} \end{pmatrix}' (L_T' L_T)^{-1} \begin{pmatrix} \Delta y_{i1} \\ \Delta y_i - \Delta y_{i,-1} \end{pmatrix}' = \frac{1}{\nu}\begin{pmatrix} 0 \\ \Delta y_{i1} \\ \vdots \\ \Sigma_{t=1}^{T-1}\Delta y_{it} \end{pmatrix}' \begin{pmatrix} \Delta y_{i1} \\ \Delta y_{i2} \\ \vdots \\ \Delta y_{iT} \end{pmatrix} = \frac{1}{\nu}\Sigma_{s=2}^{T}\Sigma_{t=1}^{s-1}\Delta y_{it}\Delta y_{is}.$$

Thus, $\ell_{RE,i,\delta}(\theta_*) - \ell_{RE,i,\omega}(\theta_*) + \nu_* \ell_{RE,i,\nu}(\theta_*) = 0.$

Proof of Proposition 6: We first check whether (i) whether $\tilde{\ell}_{RE,i,\delta\delta}(\theta_*)$ equals zero or is linearly related to $\tilde{\ell}_{RE,i,\delta}(\theta_*)$ and $s_{i,2}$. Note that

$$\tilde{\ell}_{RE,i,\nu_r}(\theta_*) = \ell_{RE,i,\nu}(\theta_*);\ \tilde{\ell}_{RE,i,\omega_r}(\theta_*) = \ell_{RE,i,\omega}(\theta_*);\ \tilde{\ell}_{RE,i,\psi}(\theta_*) = \ell_{RE,i,\psi}(\theta_*).$$

Since $\tilde{\ell}_{RE,i}(\delta, \nu_r, \omega_r, \psi) = \ell_{RE,i}(\delta, \nu(\nu_r, \delta), \omega(\omega_r, \delta), \psi)$, we have:

$$\tilde{\ell}_{RE,i,\delta} = \ell_{RE,i,\delta} - \ell_{RE,i,\omega} + \nu_r \ell_{RE,i,\nu};$$

$$\tilde{\ell}_{RE,i,\delta\delta} = \ell_{RE,i,\delta\delta} - \ell_{RE,i,\delta\omega} + \nu_r \ell_{RE,i,\delta\nu} - \ell_{RE,i,\omega\delta} + \ell_{RE,i,\omega\omega} - \nu_r \ell_{RE,i,\omega\nu}$$
$$+\nu_r\left(\ell_{RE,i,\nu\delta} - \ell_{RE,i,\nu\omega} + \nu_r \ell_{RE,i,\nu\nu}\right)$$
$$= \left(\ell_{RE,i,\delta\delta} - 2\ell_{RE,i,\delta\omega} + \ell_{RE,i,\omega\omega}\right) + 2\nu_r\left(\ell_{RE,i,\delta\nu} - \ell_{RE,i,\omega\nu}\right) + \nu_r^2 \ell_{RE,i,\nu\nu}.$$

Then, with a little algebra and Lemma A.3, we can show that at $\theta_r = \theta_*$,

$$\tilde{\ell}_{RE,i,\delta\delta}(\theta_*) = \left(\ell_{RE,i,\delta\delta}(\theta_*) - 2\ell_{RE,i,\delta\omega}(\theta_*) + \ell_{RE,i,\omega\omega}(\theta_*)\right)$$
$$+2\nu_*\left(\ell_{RE,i,\delta\nu}(\theta_*) - \ell_{RE,\omega\nu}(\theta_*)\right) + \nu_*^2\ell_{RE,i,\nu\nu}(\theta_*)$$
$$= \begin{pmatrix} -\frac{1}{\nu_*}\Delta y_{i,-1}' B_{T-1}^{-1}\Delta y_{i,-1} - \frac{T}{\nu_*}\Delta y_{i,-1}' k_{T-1}k_{T-1}'\Delta y_{i,-1} \\ +\frac{2T^2}{\nu_*}\Delta y_{i,-1}' k_{T-1}(\Delta y_{i1} + k_{T-1}'(\Delta y_i - \Delta y_{i,-1})) - \frac{T^3}{\nu_*}\left(\Delta y_{i1} + k_{T-1}'(\Delta y_i - \Delta y_{i,-1})\right)^2 \end{pmatrix}$$
$$+\frac{T(T+1)}{2}$$
$$= \left(-\frac{1}{\nu_*}\Delta y_{i,-1}'\left(\tilde{D}_T{}'\tilde{D}_T\right)^{-1}\Delta y_{i,-1} - \frac{2T}{\nu_*}\Delta y_{i,-1}' m_{T-1}(\Delta y_{iT}) - \frac{T}{\nu_*}(\Delta y_{iT})^2\right) + \frac{T(T+1)}{2},$$

which is neither zero, nor a linear combination of $\tilde{\ell}_{RE,i,\delta}(\theta_*)$ and $s_{i,2}$.

We now check (ii) whether or not $\tilde{\ell}_{RE,i,\delta\delta\delta}(\theta_*)$ is a linear function of s_i. We can show

$$\tilde{\ell}_{RE,i,\delta\delta\delta} = \left(\ell_{RE,i,\delta\delta\delta} - 3\ell_{RE,,\delta\delta\omega} + 3\ell_{RE,i,\omega\omega\omega} - \ell_{RE,i,\omega\omega\omega}\right)$$
$$+3\nu_r\left(\ell_{RE,,i,\delta\delta\nu} - 2\ell_{RE,i,\delta\omega\nu} + \ell_{RE,i,\omega\omega\nu}\right) + 3\nu_r{}^2\left(\ell_{RE,i,\delta\nu\nu} - \ell_{RE,i,\omega\nu\nu}\right) + \nu_r^3\left(\ell_{RE,i,\nu\nu\nu}\right).$$

But, at $\theta = \theta_*$,

$$\ell_{RE,i,\delta\delta\delta}(\theta_*) = 0\,;\quad \ell_{RE,i,\delta\delta\omega} = \frac{T^2}{\nu_*}\Delta y_{i,-1}' k_{T-1}k_{T-1}'\Delta y_{i,-1}\,;$$

$$\ell_{RE,i,\omega\omega\delta}(\theta_*) = \frac{2T^3}{\nu_*}\Delta y_{i,-1}' k_{T-1}\left(\Delta y_{i1} + k_T'(\Delta y_i - \Delta y_{i,-1})\right);$$

$$\ell_{RE,i,\omega\omega\omega}(\theta_*) = -T^3 + \frac{3T^4}{\nu_*}(\Delta y_{i1} + k_{T-1}'(\Delta y_i - \Delta y_{i,-1}))^2\,;$$

$$\ell_{RE,i,\delta\delta\nu}(\theta_*) = -\frac{1}{\nu_*}\ell_{RE,i,\delta\delta}(\theta_*)\,;\quad \ell_{RE,i,\delta\omega\nu}(\theta_*) = -\frac{1}{\nu_*}\ell_{RE,i,\delta\omega}(\theta_*)\,;$$

$$\ell_{RE,i,\omega\omega\nu}(\theta_*) = -\frac{1}{\nu_*}\ell_{RE,i,\omega\omega}(\theta_*) + \frac{T^2}{2\nu_*}\,;\quad \ell_{RE,i,\delta\nu\nu}(\theta_*) = -\frac{2}{\nu_*}\ell_{RE,i,\delta\nu}(\theta_*)\,;$$

$$\ell_{RE,i,\nu\nu\omega}(\theta_*) = -\frac{2}{\nu_*}\ell_{RE,i,\nu\omega}(\theta_*)\,;\quad \ell_{RE,i,\nu\nu\nu}(\theta_*) = -\frac{3}{\nu_*}\ell_{RE,i,\nu\nu}(\theta_*) + \frac{T}{2\nu_*{}^3}.$$

Using these results, we can show:

$$\tilde{\ell}_{RE,i,\delta\delta\delta}(\theta_*) = \frac{2T^3+3T^2+T}{2} - \frac{3T^2}{\nu_*}\Delta y_{i,-1}' m_{T-1} m_{T-1}' \Delta y_{i,-1}$$
$$-\frac{6T^2}{\nu_*}(m_{T-1}'\Delta y_{i,-1})\Delta y_{iT} - \frac{3T^2}{\nu_*}(\Delta y_{iT})^2 - 3\tilde{\ell}_{RE,i,\delta\delta}(\theta_*),$$

which is neither zero, nor a linear combination of s_i. For example, when $T = 2$:

$$\tilde{\ell}_{RE,i,\delta\delta\delta}(\theta_*) = 15 - \frac{3}{\nu_*}\left(\Delta y_{i1}\right)^2 - \frac{12}{\nu_*}\Delta y_{i1}\Delta y_{i2} - \frac{12}{\nu_*}(\Delta y_{i2})^2 - 3\tilde{\ell}_{RE,i,\delta\delta}(\theta_*);$$

$$\tilde{\ell}_{RE,i,\delta\delta}(\theta_*) = \left(-\frac{1}{\nu_*}\left(\Delta y_{i1}\right)^2 - \frac{2}{\nu_*}\Delta y_{i1}\Delta y_{i2} - \frac{2}{\nu_*}(\Delta y_{i2})^2\right) + 3;$$

$$\tilde{\ell}_{RE,i,\nu_r}(\theta_*) = \ell_{RE,i,\nu}(\theta_*) = -\frac{1}{\nu_*} + \frac{1}{2\nu_*^{\,2}}\left((\Delta y_{i1})^2 + (\Delta y_{i2})^2\right)$$

$$\tilde{\ell}_{RE,i,\omega_r}(\theta_*) = \ell_{RE,i,\omega}(\theta_*) = -1 + \frac{1}{2\nu_*}(\Delta y_{i1} + \Delta y_{i2})^2;$$

$$\tilde{\ell}_{RE,i,\psi_r}(\theta_*) = \ell_{RE,i,\psi}(\theta_*) = \frac{1}{\nu_*} y_{i0}\left(\Delta y_{i1} + \Delta y_{i2}\right).$$

All of these derivatives are linearly independent.

We now consider the asymptotic distribution of the ML estimator of θ_r. Let $L_N(\theta_r) = \Sigma_{i=1}^N \tilde{\ell}_{RE,i}(\theta_r)$, where $\theta_r = (\delta, \varphi_r')'$. For convenience, redefine φ_r as $\varphi_r = (\omega_r, \psi, \nu_r)'$ instead of $(\nu_r, \omega_r, \psi)'$. Correspondingly, we also reorder $s_{i,2}$ and Z_2. Partition Z_2 into $Z_2 = (Z_{21}, Z_{22})'$, where Z_{21} is a 2×1 vector and Z_{22} is a scalar. We also partition Υ and Υ^{-1}, accordingly:

$$\Upsilon = \begin{pmatrix} \Upsilon_{11} & \Upsilon_{21}' \\ \Upsilon_{21} & \Upsilon_{22} \end{pmatrix} = \begin{pmatrix} \Upsilon_{11} & \Upsilon_{21,1}' & \Upsilon_{22,1}' \\ \Upsilon_{21,1} & \Upsilon_{21,21} & \Upsilon_{22,21}' \\ \Upsilon_{22,1} & \Upsilon_{22,21} & \Upsilon_{22,22} \end{pmatrix}; \ \Upsilon^{-1} = \begin{pmatrix} \Upsilon^{11} & (\Upsilon^{21})' \\ \Upsilon^{21} & \Upsilon^{22} \end{pmatrix} = \begin{pmatrix} \Upsilon^{11} & \Upsilon^{21,1\prime} & \Upsilon^{22,1\prime} \\ \Upsilon^{21,1} & \Upsilon^{21,21} & (\Upsilon^{22,21})' \\ \Upsilon_{22,1} & \Upsilon_{22,21} & \Upsilon^{22,22} \end{pmatrix}.$$

Given the conditions (i)-(ii) are satisfied, Theorem 3 of RCBR implies that in the bounded neighborhood of θ_*, the log-likelihood function can be approximated by:

$$
\begin{aligned}
L_N(\theta_r) &= L_N(\theta_*) + \sqrt{N}\begin{pmatrix}(\delta-1)^2\\ \varphi_r-\varphi^*\end{pmatrix}' \Upsilon \begin{pmatrix}Z_1\\ Z_2\end{pmatrix} - \frac{1}{2}N\begin{pmatrix}(\delta-1)^2\\ \varphi_r-\varphi^*\end{pmatrix}' \Upsilon \begin{pmatrix}(\delta-1)^2\\ \varphi_r-\varphi^*\end{pmatrix} + o_p(1) \\
&= L_N(\theta_*) + \sqrt{N}\begin{pmatrix}(\delta-1)^2\\ \varphi_r-\varphi^*\end{pmatrix}' (\Upsilon^{-1})^{-1} \begin{pmatrix}Z_1\\ Z_2\end{pmatrix} - \frac{1}{2}N\begin{pmatrix}(\delta-1)^2\\ \varphi_r-\varphi^*\end{pmatrix}' (\Upsilon^{-1})^{-1} \begin{pmatrix}(\delta-1)^2\\ \varphi_r-\varphi^*\end{pmatrix} + o_p(1)
\end{aligned}
\tag{A.9}
$$

Suppose that $Z_1 > 0$. Then, it is straightforward to show that the ML estimator of θ_r equals

$$
\begin{pmatrix}\sqrt{N}(\hat{\delta}-1)^2\\ \sqrt{N}(\hat{\varphi}_r-\varphi^*)\end{pmatrix} = \begin{pmatrix}Z_1\\ Z_2\end{pmatrix} + o_p(1).
$$

That is,

$$
\begin{pmatrix}\sqrt{N^{1/2}}(\hat{\delta}-1)\\ \sqrt{N}(\hat{\varphi}_r-\varphi^*)\end{pmatrix} = \begin{pmatrix}(-1)^B\sqrt{Z_1}\\ Z_2\end{pmatrix} + o_p(1). \tag{A.10}
$$

This is so because the solution for $\sqrt{N^{1/2}}(\hat{\delta}-1)$ has two solutions and both solutions have equal chances. Now, suppose $Z_1 < 0$. Then, we do not have an interior solution for $\sqrt{N}(\hat{\delta}-1)^2 > 0$. The corner solution for $\sqrt{N}(\hat{\delta}-1)^2$ equals zero. For this case, we have:

$$
\begin{pmatrix}\sqrt{N^{1/2}}(\hat{\delta}-1)\\ \sqrt{N}(\hat{\varphi}_r-\varphi^*)\end{pmatrix} = \begin{pmatrix}0\\ Z_2-\Upsilon^{21}(\Upsilon^{11})^{-1}Z_1\end{pmatrix} + o_p(1). \tag{A.11}
$$

Thus, we obtain (17).

We now derive the asymptotic distributions of LR test statistics. Substituting (A.10) and (A.11) into (A.9), we have:

$$
2[L_N(\hat{\theta}_r) - L_N(\theta_*)] = 1(Z_1>0)\times Z'(\Upsilon^{-1})^{-1}Z + 1(Z_2<0)\times Z_2^{*\prime}\Upsilon_{22}Z_2^*, \tag{A.12}
$$

where $1(\bullet)$ is the index function, $\Upsilon_{22} = [\Upsilon^{22} - \Upsilon^{21}(\Upsilon^{11})^{-1}\Upsilon^{12}]^{-1}$ and $Z_2^* = Z_2 - \Upsilon^{21}(\Upsilon^{11})^{-1}Z_1$.

A tedious algebra shows:

$$
\begin{aligned}
Z(\Upsilon^{-1})^{-1}Z &= Z_1'(\Upsilon^{11})^{-1}Z_1 + Z_2^{*\prime}\Upsilon_{22}Z_2^* \\
&= Z_1'(\Phi_{11})^{-1}Z_1 + Z_{21}^{**\prime}(\Phi_{21,21})^{-1}Z_{21}^{**} + Z_{22}^{*}{}'(\Phi_{22,22})^{-1}Z_{22}^{**},
\end{aligned}
\tag{A.13}
$$

where, $\Phi_{11} = \Upsilon^{11}$, $\Phi_{21,21} = \Upsilon^{21,21} - \Upsilon^{21,1}(\Upsilon^{11})^{-1}\Upsilon^{21,1\prime}$, $Z_{21}^{**} = Z_{21} - \Upsilon^{21,1}(\Upsilon^{11})^{-1}Z_1$,

$$
\Phi_{22,22} = \Upsilon^{22,22} - \begin{pmatrix}\Upsilon^{22,1} & \Upsilon^{22,21}\end{pmatrix}\begin{pmatrix}\Upsilon^{11} & \Upsilon^{21,1\prime}\\ \Upsilon^{21,1} & \Upsilon^{21,21}\end{pmatrix}^{-1}\begin{pmatrix}\Upsilon^{22,1\prime}\\ \Upsilon^{22,21\prime}\end{pmatrix};
$$

$$Z_{22}^{**} = Z_{22} - \begin{pmatrix} \Upsilon^{22,1} & \Upsilon^{22,21} \end{pmatrix} \begin{pmatrix} \Upsilon^{11} & \Upsilon^{21,1\prime} \\ \Upsilon^{21,1} & \Upsilon^{21,21} \end{pmatrix}^{-1} \begin{pmatrix} Z_1 \\ Z_{21} \end{pmatrix}.$$

Notice that Z_1, Z_{21}^{**} and Z_{22}^{**} are uncorrelated, and

$$Z_1 \sim N(0, \Phi_{11}); \; Z_{21}^{**} \sim N(0, \Upsilon_{21,21}); \; Z_{22}^{**} \sim N(0, \Upsilon_{22,22}). \tag{A.14}$$

Using (A.13), we can rewrite (A.12) as:

$$\begin{aligned} 2[L_N(\hat{\theta}_r) - L_N(\theta_*)] &= 1(Z_1 > 0) \times (Z_1'(\Phi_{11})^{-1} Z_1 + Z_{21}^{**\prime}(\Phi_{21,21})^{-1} Z_{21}^{**} + Z_{22}^{**\prime}(\Phi_{22,22})^{-1} Z_{22}^{**}) \\ &+ 1(Z_1 < 0) \times (Z_{21}^{**\prime}(\Phi_{21,21})^{-1} Z_{21}^{**} + Z_{22}^{**\prime}(\Phi_{22,22})^{-1} Z_{22}^{**}). \end{aligned} \tag{A.15}$$

Let $\tilde{\theta}_r = (1,1,0,\tilde{\nu})'$ be the restricted ML estimator of θ_r which maximizes (A.9) with the restrictions, $\delta = 1$ and $\omega = \psi = 0$. It can be shown that:

$$\sqrt{N}(\tilde{\nu} - \nu_*) = Z_{22}^{**} + o_p(1).$$

Substituting this solution into (A.9), we can have

$$\begin{aligned} 2[L_N(\tilde{\theta}_r) - L_N(\theta_*)] &= 2\sqrt{N} \begin{pmatrix} 0 \\ 0_{2\times1} \\ \tilde{\nu} - \nu_* \end{pmatrix}' \Upsilon \begin{pmatrix} Z_1 \\ Z_{21} \\ Z_{22} \end{pmatrix} - N \begin{pmatrix} 0 \\ 0_{2\times1} \\ \tilde{\nu} - \nu * \end{pmatrix}' \Upsilon \begin{pmatrix} 0 \\ 0_{2\times1} \\ \tilde{\nu} - \nu * \end{pmatrix} + o_p(1) \\ &= Z_{22}^{**\prime}(\Phi_{22,22})^{-1} Z_{22}^{**} + o_p(1) \end{aligned} \tag{A.16}$$

Using (A.14)-(A.16), we can show that when $\theta_{r,o} = \theta_*$, the LR test statistic for testing the joint hypotheses of $\delta_o = 1$, $\omega_{r,o} = 0$, and $\psi_o = 0$ follows a mixed Chi-square distribution:

$$\begin{aligned} 2[L_N(\hat{\theta}_r) - L_N(\tilde{\theta}_r)] &= 1(Z_1 > 0) \times (Z_1'(\Phi_{11})^{-1} Z_1 + Z_{21}^{**\prime}(\Phi_{21,21})^{-1} Z_{21}^{**}) \\ &+ 1(Z_1 < 0) \times (Z_{21}^{**\prime}(\Phi_{21,21})^{-1} Z_{21}^{**}) + o_p(1) \\ &\to_d B \times \chi^2(3) + (1 - B) \times \chi^2(2). \end{aligned}$$

Finally, consider the LR statistic for testing the hypothesis that $\delta_o = 1$. Let $\widehat{\theta}_r = (1, \widehat{\varphi}_r')'$ be restricted ML estimator that maximize (A.9) with the restriction $\delta_o = 1$. We can show that

$$\sqrt{N}(\widehat{\varphi}_r - \varphi_*) = Z_2^* + o_p(1).$$

Substituting this solution into (A.9) yields:

$$2[L_N(\widehat{\theta}_r) - L_N(\theta_*)] = 2\sqrt{N}\begin{pmatrix} 0 \\ \widehat{\varphi}_r - \varphi_* \end{pmatrix}' \Upsilon \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} - N\begin{pmatrix} 0 \\ \widehat{\varphi}_r - \varphi_* \end{pmatrix}' \Upsilon \begin{pmatrix} 0 \\ \widehat{\varphi}_r - \varphi_* \end{pmatrix} + o_p(1) \quad \text{(A.17)}$$

$$= Z_2^{*\prime}\Upsilon_{22}Z_2^* + o_p(1) = Z_{21}^{**\prime}(\Phi_{21,21})^{-1}Z_{21}^{**} + Z_{22}^{**\prime}(\Phi_{22,22})^{-1}Z_{22}^{**} + o_p(1).$$

Then, using (A.14), (A.15) and (A.17), we have:

$$2[L_N(\hat{\theta}_r) - L_N(\widehat{\theta}_r)] = 1(Z_1 > 0) \times Z_1'(\Phi_{11})^{-1}Z_1 + o_p(1) \to_d B\chi^2(1).$$

This completes our proof.

Proof of Proposition 7: It is easy to see that $(\ell^D_{RE,i,\delta}(\phi^D_*) - a_*\ell^D_{RE,i,\beta}(\phi^D_*))$, $\ell^D_{RE,i,\nu}(\phi^D_*)$, and $\ell^D_{RE,i,\omega}(\phi^D_*)$ are the same as $\ell_{RE,i,\delta}(\theta_*)$, $\ell_{RE,i,\nu}(\theta_*)$, and $\ell_{RE,i,\omega}(\theta_*)$ with the Δy_{it} replaced by $\Delta y_{it} - a$. Thus, we can obtain the result by the same method used for Propositions 5.

Proof of Proposition 8: Define

$$L_T^{HD\prime} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & 0 & \dots & 0 \\ 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}_{T\times T}; \quad J_T = \begin{pmatrix} I_2 \\ 0_{(T-2)\times 2} \end{pmatrix}; \quad G_T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T-1 \end{pmatrix},$$

We can easily show that $G = (L_T^{HD\prime})^{-1}J_T$ and $\Omega_T^{HD}(W) = L_T^{HD\prime}L_T^{HD} + J_TWJ_T'$. Let

$$\begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix} \equiv G_T'G_T = \begin{pmatrix} T & \dfrac{T(T-1)}{2} \\ \dfrac{T(T-1)}{2} & \dfrac{T(T-1)(2T-1)}{6} \end{pmatrix}.$$

Then, we have:

$$\begin{aligned}(\Omega_T^{HD})^{-1} &= \left[L_T^{HD\prime}L_T^{HD} + J_TWJ_T'\right]^{-1} \\ &= \left(L_T^{HD\prime}L_T^{HD}\right)^{-1} - (L_T^{HD\prime}L_T^{HD})^{-1}J_TW[J_T'(L_T^{HD\prime}L_T^{HD})^{-1}J_TW + I_2]^{-1}J_T'(L_T^{HD\prime}L_T^{HD})^{-1} \\ &= (L_T^{HD})^{-1}(L_T^{HD\prime})^{-1} - (L_T^{HD})^{-1}G_TW[G_T'G_TW + I_2]^{-1}G_T'(L_T^{HD\prime})^{-1}.\end{aligned}$$

Thus, the log-likelihood function can be written:

$$\ell_{RE,i}^{HD} = -\frac{1}{2}\ln(2\pi) - \frac{T}{2}\ln(\nu) - \frac{1}{2}\ln[\det(L_T^{HD\prime}L_T^{HD} + J_T W J_T')]$$

$$-\frac{1}{2\nu}\begin{pmatrix} \Delta y_{i1} - a_1 - \psi_1 y_{i0} \\ \Delta y_{i2} - a_2 - \psi_2 y_{i0} - \Delta y_{i1}\delta \\ \Delta^2 y_i - \Delta^2 y_{i,-1}\delta \end{pmatrix}' (L_T^{HD\prime}L_T^{HD} + J_T W J_T')^{-1} \begin{pmatrix} \Delta y_{i1} - a_1 - \psi_1 y_{i0} \\ \Delta y_{i2} - a_2 - \psi_2 y_{i0} - \Delta y_{i1}\delta \\ \Delta^2 y_i - \Delta^2 y_{i,-1}\delta \end{pmatrix}$$

Define $d_{it} = \Delta y_{it} - a_{1,*} - \psi_{1,*} y_{i0}$. By a tedious but straightforward algebra, we can show at $\phi^{HD} = \phi_*^{HD}$:

$$\ell_{RE,i,\delta}^{HD} = \frac{1}{\nu}\Sigma_{t=1}^{T-1}\Sigma_{s=1}^{t}\Delta y_{is} d_{i,t+1} - \frac{\omega_{11}}{\nu(g_{11}\omega_{11}+1)}(\Sigma_{t=1}^{T-1}(T-t)\Delta y_{it})\Sigma_{t=1}^{T} d_{it};$$

$$\ell_{RE,i,a_2}^{HD} = \frac{1}{\nu}\Sigma_{t=2}^{T}(t-1)d_{it} - \frac{w_{11}}{\nu(g_{11}\omega_{11}+1)}\frac{T(T-1)}{2}\Sigma_{t=1}^{T} d_{it};$$

$$\ell_{RE,i,\psi_2}^{HD} = \frac{1}{\nu}\Sigma_{t=2}^{T}(t-1)d_{it}y_{i0} - \frac{w_{11}}{\nu(g_{11}\omega_{11}+1)}\frac{T(T-1)}{2}\Sigma_{t=1}^{T} d_{it}y_{i0},$$

where $\nu = \nu_*$ and $\omega_{11} = \omega_{11,*}$. Thus, we have:

$$\ell_{RE,i,\delta}^{HD} - a_1\ell_{RE,i,a_2}^{HD} - \psi_1\ell_{RE,i,\psi_2}^{HD} = \frac{1}{\nu}\Sigma_{t=1}^{T-1}\Sigma_{s=1}^{t} d_{i,s}d_{i,t+1} - \frac{\omega_{11}}{\nu(g_{11}\omega_{11}+1)}(\Sigma_{t=1}^{T-1}(T-t)d_{it})(\Sigma_{t=1}^{T} d_{it}).$$

We also can show:

$$\ell_{RE,i,\nu}^{HD} = -\frac{T}{2\nu} + \frac{1}{2\nu^2}\Sigma_{t=1}^{T} d_{it}^{\ 2} - \frac{\omega_{11}}{2\nu^2(g_{11}\omega_{11}+1)}(\Sigma_{t=1}^{T} d_{it})^2;$$

$$\ell_{RE,i,\omega_{11}}^{HD} = -\frac{1}{2}\frac{g_{11}}{g_{11}\omega_{11}+1} - \frac{g_{11}\omega_{11}}{2\nu(g_{11}\omega_{11}+1)^2}(\Sigma_{t=1}^{T} d_{it})^2 + \frac{1}{2\nu(g_{11}\omega_{11}+1)}(\Sigma_{t=1}^{T} d_{it})^2;$$

$$\ell_{RE,i,\omega_{12}}^{HD} = -\frac{g_{12}}{(g_{11}\omega_{11}+1)} - \frac{g_{12}\omega_{11}}{\nu(g_{11}\omega_{11}+1)^2}(\Sigma_{t=1}^{T} d_{it})^2 + \frac{1}{\nu(g_{11}\omega_{11}+1)}(\Sigma_{t=1}^{T} d_{it})(\Sigma_{t=1}^{T-1} t d_{i,t+1}).$$

But,

$$\ell^{HD}_{RE,i,\delta} + \nu \ell^{HD}_{RE,i,\nu} - (g_{11}\omega_{11}+1)\ell^{HD}_{RE,i,\omega_{11}}$$

$$= \frac{1}{2\nu}\begin{bmatrix} \Sigma_{t=1}^{T} d_{it}^2 + 2\Sigma_{t=1}^{T}\Sigma_{s=1}^{t-1} d_{is} d_{it} \\ -(\Sigma_{t=1}^{T} d_{it})^2 \end{bmatrix} - \frac{\omega_{11}\left(\Sigma_{t=1}^{T-1}(T-t)d_{it}\right)\left(\Sigma_{t=1}^{T} d_{it}\right)}{\nu(g_{11}\omega_{11}+1)} + \frac{(g_{11}-1)\omega_{11}\left(\Sigma_{t=1}^{T} d_{it}\right)^2}{2\nu(g_{11}\omega_{11}+1)}$$

$$= -\frac{\omega_{11}\left(\Sigma_{t=1}^{T-1}(T-t)d_{it}\right)\left(\Sigma_{t=1}^{T} d_{it}\right)}{\nu(g_{11}\omega_{11}+1)} + \frac{(g_{11}-1)\omega_{11}\left(\Sigma_{t=1}^{T} d_{it}\right)^2}{2\nu(g_{11}\omega_{11}+1)}$$

$$= -\frac{\omega_{11}\left((T-1)\Sigma_{t=1}^{T} d_{it} - \Sigma_{t=1}^{T}(t-1)d_{it}\right)\left(\Sigma_{t=1}^{T} d_{it}\right)}{\nu(g_{11}\omega_{11}+1)} + \frac{(g_{11}-1)\omega_{11}\left(\Sigma_{t=1}^{T} d_{it}\right)^2}{2\nu(g_{11}\omega_{11}+1)}$$

$$= -\frac{(T-1)\omega_{11}\left(\Sigma_{t=1}^{T} d_{it}\right)^2}{\nu(g_{11}\omega_{11}+1)} + \frac{\omega_{11}\left(\Sigma_{t=1}^{T}(t-1)d_{it}\right)\left(\Sigma_{t=1}^{T} d_{it}\right)}{\nu(g_{11}\omega_{11}+1)} + \frac{(g_{11}-1)\omega_{11}\left(\Sigma_{t=1}^{T} d_{it}\right)^2}{2\nu(g_{11}\omega_{11}+1)}$$

$$= -\frac{(T-1)\omega_{11}}{2\nu(T\omega_{11}+1)}\left(\Sigma_{t=1}^{T} d_{it}\right)^2 + \frac{\omega_{11}}{\nu(T\omega_{11}+1)}\left(\Sigma_{t=1}^{T}(t-1)d_{it}\right)\left(\Sigma_{t=1}^{T} d_{it}\right),$$

where the last equality results from $g_{11} = T$. Now observe that:

$$(T-1)\ell^{HD}_{RE,i,\omega_{11}} - \ell^{HD}_{RE,i,\omega_{12}} = \frac{(T-1)}{2\nu(T\omega_{11}+1)}\left(\Sigma_{t=1}^{T} d_{it}\right)^2 - \frac{1}{\nu(T\omega_{11}+1)}\left(\Sigma_{t=1}^{T} d_{it}\right)\left(\Sigma_{t=1}^{T}(t-1)d_{it}\right).$$

Thus, we have:

$$\ell^{HD}_{RE,i,\delta} + \nu\ell^{HD}_{RE,i,\nu} - (T\omega_{11}+1)\ell^{HD}_{RE,i,\omega_{11}} + \omega_{11}[(T-1)\ell^{HD}_{RE,i,\omega_{11}} - \ell^{HD}_{RE,i,\omega_{12}}]$$
$$= \ell^{HD}_{RE,i,\delta} + \nu\ell^{HD}_{RE,i,\nu} - (\omega_{11}+1)\ell^{HD}_{RE,i,\omega_{11}} - \omega_{11}\ell^{HD}_{RE,i,\omega_{12}} = 0.$$

An alternative representation of the log likelihood function (27) is given:

$$\ell^{HD}_{RE,i} = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln[\det(B^{HD}_{T-2})] - \frac{T}{2}\ln(\nu) - \frac{1}{2}\ln[\det(\Xi)]$$
$$-\frac{1}{2\nu}(\Delta^2 y_i - \Delta^2 y_{i,-1}\delta)'(B^{HD}_{T-2})^{-1}(\Delta^2 y_i - \Delta^2 y_{i,-1}\delta)$$
$$+\frac{1}{2\nu}\left(\begin{pmatrix} \Delta y_{i1} - a_1 - \psi_1 y_{i0} \\ \Delta y_{i2} - a_2 - \psi_2 y_{i0} - \delta\Delta y_{i1} \end{pmatrix} + K^{HD\prime}_{T-2}(\Delta^2 y_i - \Delta^2 y_{i,-1}\delta)\right)'$$
$$\times\Xi^{-1}\left(\begin{pmatrix} \Delta y_{i1} - a_1 - \psi_1 y_{i0} \\ \Delta y_{i2} - a_2 - \psi_2 y_{i0} - \delta\Delta y_{i1} \end{pmatrix} + K^{HD\prime}_{T-2}(\Delta^2 y_i - \Delta^2 y_{i,-1}\delta)\right),$$

where,

$$C_{T-2}^{HD} = \begin{pmatrix} -1 & 3 \\ 0 & -1 \\ 0 & 0 \\ : & : \\ 0 & 0 \end{pmatrix}_{(T-2)x2} \quad ; \quad B_{T-2}^{HD} = \begin{pmatrix} 6 & -4 & 1 & 0 & ... & 0 \\ -4 & 6 & -4 & 1 & ... & 0 \\ 1 & -4 & 6 & -4 & ... & 0 \\ : & : & : & : & & : \\ 0 & 0 & 0 & 0 & ... & -4 \\ 0 & 0 & 0 & & ... & 6 \end{pmatrix}_{(T-2)x(T-2)} \quad ; \quad K_{T-2}^{HD} = (B_{T-2}^{HD})^{-1}(C_{T-2}^{HD});$$

$$\Xi \equiv \begin{pmatrix} \xi_{11,T} & \xi_{12,T} \\ \xi_{12,T} & \xi_{22,T} \end{pmatrix} = \begin{pmatrix} \omega_{11}+1 & \omega_{12}-1 \\ \omega_{12}-1 & \omega_{22}+2 \end{pmatrix} - (C_{T-2}^{HD})'(B_{T-2}^{HD})^{-1}(C_{T-2}^{HD}).$$

Our unreported simulations show that it is much easer to minimize the above log likelihood function with respect to $(\delta, \nu, \xi_{T,11}, \xi_{T,12}, \xi_{T,22}, \psi_1, \psi_2, a_1, a_2)'$ than to minimize the equivalent log likelihood function (27) with respect to $\phi^{HD} = (\delta, \nu, \omega_{11}, \omega_{12}, \omega_{13}, \psi_1, \psi_2, a_1, a_2)'$.

REFERENCES

Ahn, S.C., and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, Journal of Econometrics, 68, 5-27.

Ahn, S.C. and P. Schmidt, 1997, Efficient estimation of dynamic panel data models: alternative assumptions and simplified assumptions, Journal of Econometrics 76, 309-321.

Alvarez, A. and M. Arellano, 2004, Robust likelihood estimation of dynamic panel data models, mimeo, CEMFI, Spain.

Anderson, T.W. and C. Hsiao, 1981, Estimation of dynamic models with error components, Journal of the American Statistical Association 76, 598-606.

Andrews, D., 1999, Estimation when a parameter is on a boundary, Econometrica 67, 1341-1384.

Arellano, M. and S. Bond, 1991, Tests of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of Economic Studies 58, 277-297.

Arellano, M. and O. Bover, 1995, Another look at the instrumental variables estimation of error-component models, Journal of Econometrics 68, 29-51.

Bhargava, A. and J.D. Sargan, 1983, Estimating dynamic random effects models from panel data covering short time periods, Econometrica 51, 1635-1660.

Blundell, R. and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, Journal of Econometrics 87, 115-143.

Bond, S. and F. Windmeijer, 2002, Finite sample inference for GMM estimators in linear panel data models, IFS, mimeo.

Breitung, J. and W. Meyer, 1994, Testing for unit roots using panel data; are wages on different bargaining levels cointegrated, Applied Economics 26, 353-361.

Cox, D.R. and N. Reid, 1987, Parameter orthogonally and approximate conditional inference (with discussion), Journal of the Royal Statistical Society, B, 49 (1), 1-39.

Gouriéroux, C., A. Monfort and A. Trognon, 1984, Pseudo maximum likelihood methods: Theory, Econometrica 52, 681-700.

Hahn, J. and G. Kuersteiner, 2002, Asymptotically unbiased inference for a dynamic panel data model with fixed effects when both n and T are large, Econometrica 70, 1639-1657.

Harris, R. and E. Tzavalis, 1999, Inference for unit roots in dynamic panels where the time dimension is fixed, Journal of Econometrics 91, 2001-226.

Hsiao, C., 1986, Analysis of Panel Data, Cambridge, University Press, Cambridge, UK.

Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu, 2002, Maximum likelihood estimation of fixed effects dynamic panel data models covering short time period, Journal of Econometrics 109, 107-150.

Im, K., H. Pesaran and Y. Shin, 2003, Testing for unit roots in heterogeneous panels, Journal of Econometrics 115, 53-74.

Kruiniger, H., 2002, On the estimation of panel regression models with fixed effects, Queen Mary, University of London, mimeo.

Lancaster, T., 2002, Orthogonal parameters and panel data, Review of Economic Studies 69, 647-666.

Levin, A., F. Lin and C. Chu, 2002, Unit root tests in panel data: asymptotic and finite-sample properties, Journal of Econometrics 108, 1-24.

Moon, H. and P.C.B. Phillips, 2004, GMM estimation of autoregressive roots near unity with panel data, Econometrica 72, 467-522.

Moon, H., P. Phillips and B. Perron, 2005, Incidental trends and the power of panel unit root tests, mimeo, University of Southern California.

Neyman, J. and E. Scott, 1948, Consistent estimates based on partially consistent observations, Econometrica 16, 1-32.

Nickell, S., 1981, Biases in dynamic models with fixed effects, Econometrica 49, 1399-1416.

Rao, C.R., 1973, Linear statistical inference and its applications, the 2nd eds., John Wiley & Sons Inc, New York, USA.

Rotnitzky, A., D.R. Cox, M. Bottai and J. Robins, 2000, Likelihood-based inference with singular information matrix, Bernoulli 6 (2), 243-284.

Sagan, J.D., 1983, Identification and lack of identification, Econometrica 51, 1605-1634.

Thomas, G., 2005, Maximum likelihood based estimation of dynamic panel data Models, Ph.D. dissertation, Arizona State University.

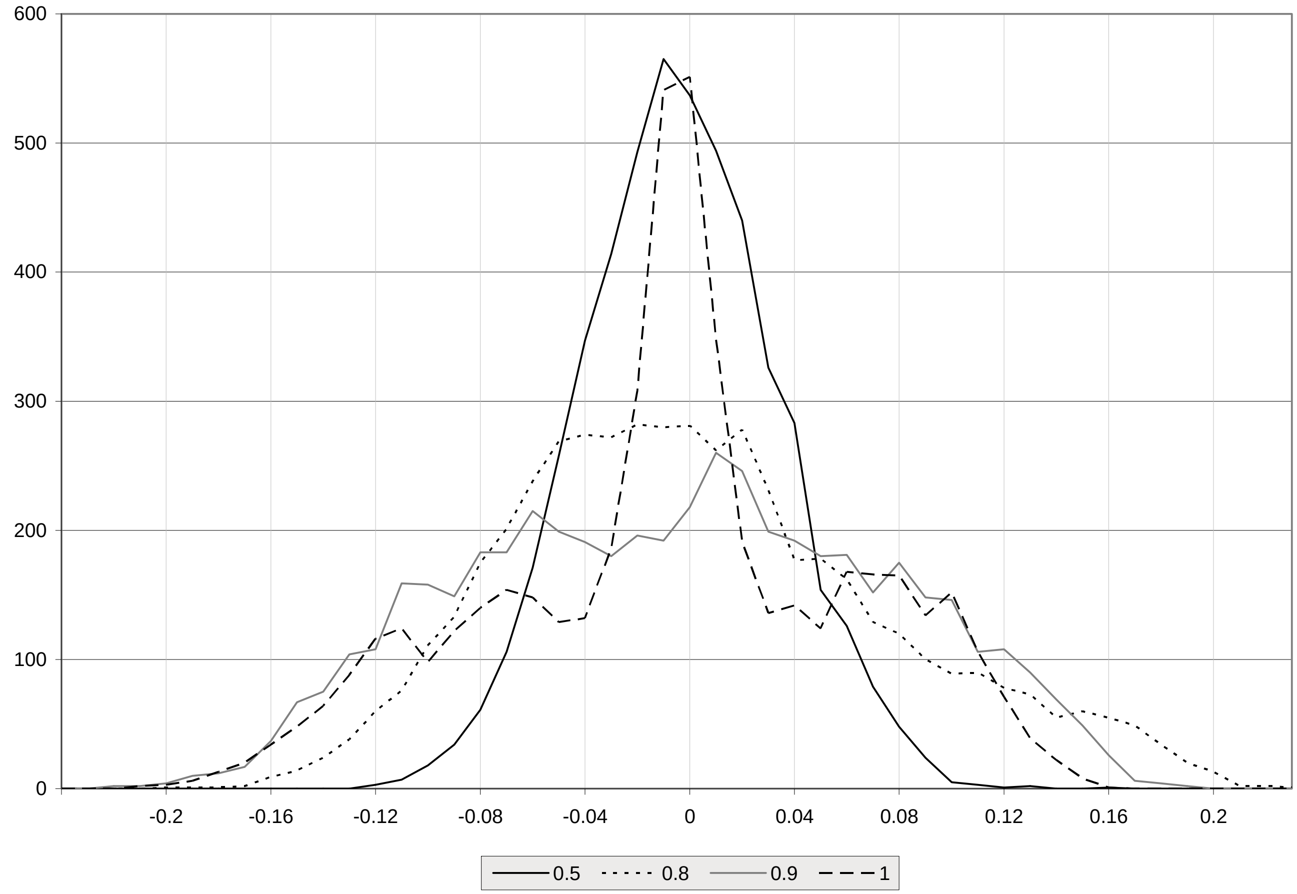
Wooldridge, J.M., 2000, A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables, Economics Letters, 68, 245-250.

Table 1
RE and HPT ML Estimation without Time Trends:
Biases in the Estimators and Size and Power Properties of the *T*-Tests

		Mean	Bias	MSE	Rejections of H_o: $\delta = \delta_o$			Rejections of H_o: $\delta = 1$		
					1%	5%	10%	1%	5%	10%
RE	N = 500									
	$\delta_o = 0.5$	0.500	0.000	0.001	1.06	5.01	9.58	100.0	100.0	100.0
	$\delta_o = 0.8$	0.809	0.009	0.005	6.33	11.45	15.88	66.15	79.34	85.03
	$\delta_o = 0.9$	0.901	0.001	0.006	9.60	17.24	23.26	20.53	29.68	35.90
	$\delta_o = 1$	0.998	-0.002	0.005	9.49	16.38	21.22	9.49	16.38	21.22
	N = 100									
	$\delta_o = 0.5$	0.503	0.003	0.006	1.31	4.34	8.49	99.30	99.58	99.62
	$\delta_o = 0.8$	0.810	0.010	0.015	9.03	15.51	20.81	26.89	38.89	46.18
	$\delta_o = 0.9$	0.891	-0.009	0.015	9.63	16.67	22.24	13.27	21.64	26.79
	$\delta_o = 1$	0.996	-0.004	0.011	9.20	15.96	21.41	9.20	15.96	21.41
HPT	N = 500									
	$\delta_o = 0.5$	0.501	0.001	0.001	1.04	4.85	9.47	100.0	100.0	100.0
	$\delta_o = 0.8$	0.8183	0.018	0.007	9.04	14.67	18.78	54.98	64.99	69.25
	$\delta_o = 0.9$	0.897	-0.003	0.006	7.26	14.66	20.43	20.78	29.57	34.68
	$\delta_o = 1$	0.998	-0.002	0.004	8.80	15.05	19.76	8.80	15.05	19.76
	N = 100									
	$\delta_o = 0.5$	0.510	0.010	0.011	2.81	5.42	8.92	97.57	97.98	98.23
	$\delta_o = 0.8$	0.811	0.011	0.016	6.87	13.57	18.39	26.33	37.23	43.11
	$\delta_o = 0.9$	0.882	-0.018	0.013	6.89	13.36	18.27	13.38	20.96	26.06
	$\delta_o = 1$	0.997	-0.003	0.010	8.02	14.33	19.08	8.02	14.33	19.08

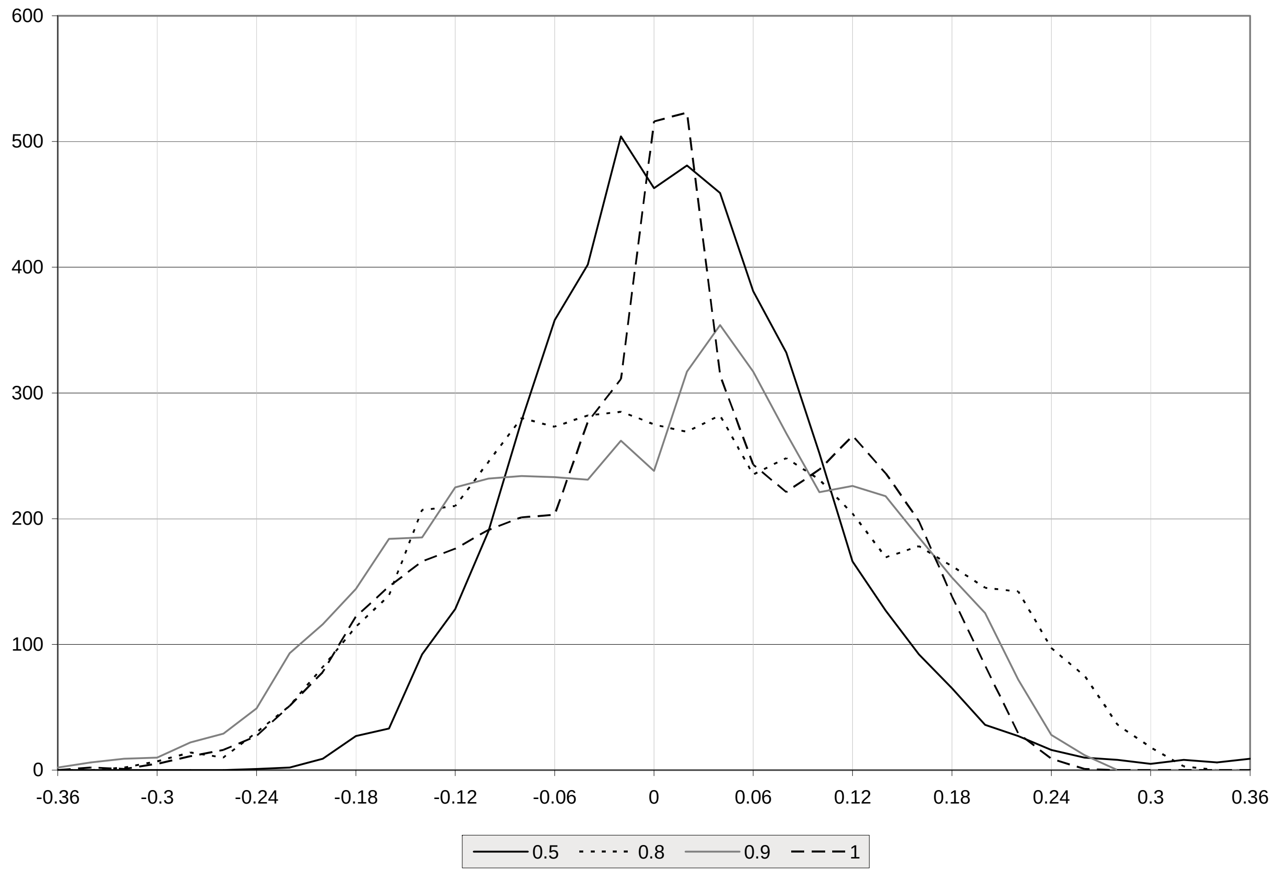
Notes: Table 1 reports the means, biases and mean square errors of the RE and HPT estimators from 10,000 simulations with the number of cross-section observations equal to *N* and the true value of δ equal to δ_o along with the rejection rates of the *t*-test with significance levels of 1%, 5% and 10%. The variables, η_i, q_{i0}, and ε_{it}, are generated from the normal distributions with zero means and the variances, variances, $\sigma_\eta^2 = 2$, $\sigma_{q0}^2 = 0$, and $\nu = 1$, respectively. The number of time series observation is fixed at $T = 5$. The data are generated without trends and with zero means. The ML estimators are estimated without trends and intercept terms.

Figure 1
The Sample Error Distributions of the RE ML Estimator without Trend When N = 500:
The true values of δ_o vary from 0.5 to 1.



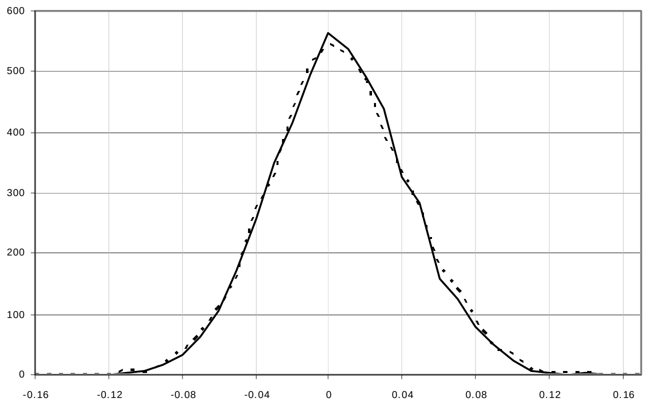
Notes: Figure 1 shows the finite-sample distribution of the sampling errors of the RE estimator for values of δ_o varying from 0.5 to 1. The variables, η_i, q_{i0}, and ε_{it}, are generated from the normal distributions with zero means and the variances, variances, $\sigma_\eta^2 = 2$, $\sigma_{q0}^2 = 0$, and $\nu = 1$, respectively. N = 500 and T = 5 are used. The data are generated without trends and with zero means. The ML estimators are estimated without trends and intercept terms

Figure 2
The Sampling Error Distribution of the RE estimator without Trend When N = 100:
The true values of δ_o vary from 0.5 to 1.

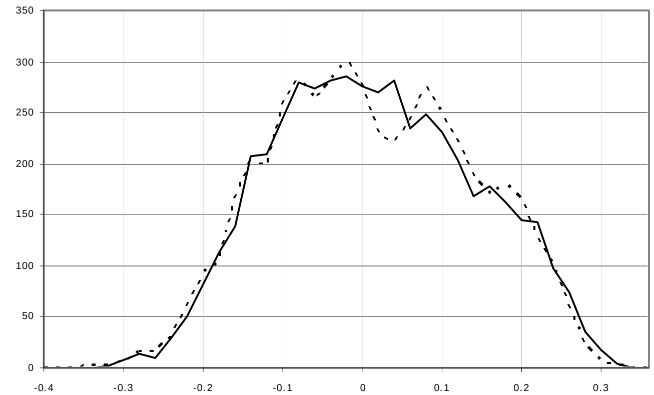


Notes: Figure 2 shows the finite-sample distribution of the sampling errors of the RE estimator for values of δ_o varying from 0.5 to 1. The variables, η_i, q_{i0}, and ε_{it}, are generated from the normal distributions with zero means and the variances, $\sigma_\eta^2 = 2$, $\sigma_{q0}^2 = 0$, and $\nu = 1$, respectively. N = 100 and T = 5 are used. The data are generated without trends and with zero means. The ML estimators are estimated without trends and intercept terms.

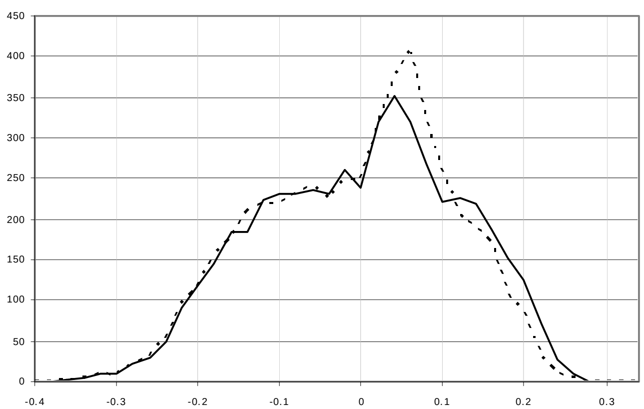
Figure 3
The Sampling Error Distributions of the RE and HPT estimators when N = 100



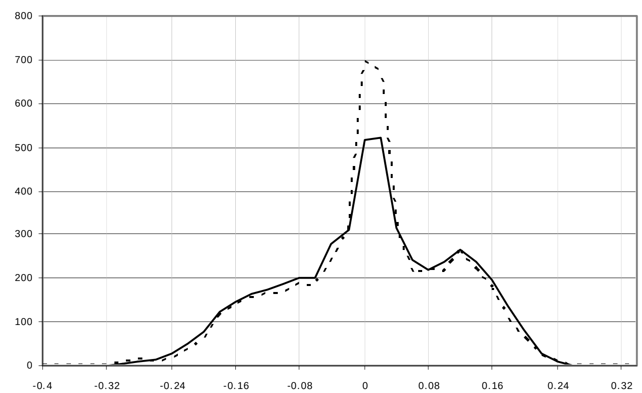
Panel A: $\delta_o = 0.5$



Panel B: $\delta_o = 0.8$



Panel C: $\delta_o = 0.9$



Panel D: $\delta_o = 1.0$

Notes: Figure 3 shows the finite-sample distributions of the sampling errors of the RE and HPT estimator for various values of δ_0. The variables, η_i, q_{i0}, and ε_{it}, are generated from the normal distributions with zero means and the variances, variances, $\sigma_\eta^2 = 2$, $\sigma_{q0}^2 = 0$, and $\nu = 1$, respectively. N = 100 and T = 5 are used. The data are generated without trends and with zero means. The ML estimators are estimated without trends and intercept terms.

Table 2
LR Tests Based on RE and HML ML Estimation with Trends:
Testing All of the Parametric Restrictions Implied by the Hypothesis of Unit Root

		$\delta_o = 1$			$\delta_o = 0.95$			$\delta_o = 0.9$		
N	Test	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 2, \nu = 1$										
N=500	LR-RE	1.06	5.40	10.48	84.54	94.75	97.41	100.0	100.0	100.0
	LR-HPT	0.95	5.03	10.07	61.75	82.57	89.79	99.55	99.94	99.98
	T-HT	(1.13)	(5.38)	(10.33)	(57.13)	(80.34)	(88.88)	(97.61)	(99.64)	(99.93)
N=100	LR-RE	1.24	6.09	11.44	12.69	32.25	44.59	60.92	82.42	94.48
	LR-HPT	1.02	5.21	10.53	8.80	24.05	35.31	34.03	59.25	71.49
	T-HT	(1.31)	(5.46)	(10.77)	(12.20)	(30.53)	(44.38)	(35.17)	(61.89)	(75.09)
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 4, \nu = 1$										
N=500	LR-RE	1.11	5.34	10.67	97.62	99.66	99.85	100.0	100.0	100.0
	LR-HPT	0.95	5.03	10.07	48.86	73.04	82.25	91.50	97.86	99.08
	T-HT	(1.13)	(5.38)	(10.33)	(46.60)	(72.67)	(83.09)	(86.54)	(96.24)	(98.58)
N=100	LR-RE	1.36	6.05	11.35	23.21	47.22	59.76	85.05	95.65	98.11
	LR-HPT	1.02	5.21	10.53	6.96	20.23	30.58	18.15	39.73	52.78
	T-HT	(1.31)	(5.46)	(10.77)	(10.03)	(26.44)	(39.67)	(22.41)	(46.39)	(60.75)
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 1, \nu = 1$										
N=500	LR-RE	1.12	5.49	10.55	72.17	88.77	93.57	99.98	100.0	100.0
	LR-HPT	0.95	5.03	10.07	68.24	86.54	92.45	99.94	100.0	100.0
	T-HT	(1.13)	(5.38)	(10.33)	(62.29)	(84.09)	(91.12)	(99.40)	(99.95)	(99.98)
N=100	LR-RE	1.20	5.87	11.47	9.52	26.53	38.23	48.78	73.23	82.84
	LR-HPT	1.02	5.21	10.53	9.88	26.36	37.72	44.76	69.23	80.26
	T-HT	(1.31)	(5.46)	(10.77)	(13.31)	(32.67)	(46.59)	(43.12)	(69.65)	(81.20)
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 2, \nu = 2$										
N=500	LR-RE	1.06	5.40	10.48	76.89	91.07	95.22	100.0	100.0	100.0
	LR-HPT	0.95	5.03	10.07	68.24	86.54	92.45	99.94	100.0	100.0
	T-HT	(1.13)	(5.38)	(10.33)	(62.29)	(84.09)	(91.12)	(99.40)	(99.95)	(99.98)
N=100	LR-RE	1.24	6.09	11.44	10.35	28.58	40.43	54.35	77.59	86.28
	LR-HPT	1.02	5.21	10.53	9.88	26.36	37.72	44.76	69.23	80.26
	T-HT	(1.31)	(5.46)	(10.77)	(13.31)	(32.67)	(46.59)	(43.12)	(69.65)	(81.20)
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 2, \nu = 4$										
N=500	LR-RE	1.06	5.40	10.48	72.35	88.65	93.72	99.99	100.0	100.0
	LR-HPT	0.95	5.03	10.07	71.65	88.25	93.75	99.97	100.0	100.0
	T-HT	(1.13)	(5.38)	(10.33)	(64.67)	(85.47)	(92.25)	(99.72)	(99.97)	(99.99)
N=100	LR-RE	1.24	6.09	11.44	9.45	26.39	38.22	51.09	74.69	84.29
	LR-HPT	1.02	5.21	10.53	10.68	27.53	39.20	50.58	74.31	84.17
	T-HT	(1.31)	(5.46)	(10.77)	(13.88)	(33.66)	(47.73)	(47.38)	(73.50)	(84.04)

Notes: Table 2 reports the empirical sizes and power (%) of the LR tests for all of the parametric restrictions implied by the hypothesis of unit root. LR-RE denotes the LR test based on the RE ML estimator with the *p*-values from a mixture of $\chi^2(3)$ and $\chi^2(2)$ over 10,000 simulations, while LR-HPT represents the LR test based on and a mixture of $\chi^2(1)$ and $\chi^2(0)$. For comparison, the size and power of a one-tailed and bias-corrected *t*-test (BT-HT) of Harris and Tzavalis (1999) are reported in parentheses. The *t*-statistic, which is obtained by correcting the bias in the usual within estimator, is asymptotically standard normal under the null hypothesis of unit root without drifts, The variables, η_i, q_{i0}, and ε_{it}, are generated from the normal distributions with zero means and, the variances, σ_η^2, σ_{q0}^2, and ν, respectively. The data are generated without trends and with zero means. The ML estimators are estimated without trends and intercept terms.

Table 3

LR Tests Based on RE and HPT ML Estimation without Trend: Testing the Single Parametric Restriction $\delta_o = 1$ Implied by the Hypothesis of Unit Root

		$\delta_o = 1$			$\delta_o = 0.95$			$\delta_o = 0.9$		
N	Test	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 2, \nu = 1$										
N = 500	LR-RE	1.04	5.75	11.56	1.09	7.42	16.61	4.80	23.32	40.98
	LR-HPT	0.95	5.04	10.26	0.88	5.34	11.85	1.28	9.61	23.12
N = 100	LR-RE	1.46	6.15	12.76	1.36	6.76	14.68	1.68	10.09	21.62
	LR-HPT	1.13	5.21	10.60	1.17	5.37	11.91	1.03	6.24	15.06
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 4, \nu = 1$										
N = 500	LR-RE	1.04	5.74	11.57	1.50	9.84	20.65	10.69	33.53	50.76
	LR_HPT	0.95	5.04	10.26	0.87	5.43	11.38	1.47	7.79	17.31
N = 100	LR-RE	1.45	6.01	12.61	1.56	7.66	16.61	2.55	13.60	26.92
	LR-HPT	1.13	5.21	10.60	1.12	5.42	11.83	1.12	6.41	13.89
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 1, \nu = 1$										
N = 500	LR-RE	1.05	5.66	11.58	1.02	6.28	14.41	2.67	17.96	36.45
	LR_HPT	0.95	5.04	10.26	0.90	5.40	12.10	1.40	12.39	29.07
N = 100	LR-RE	1.41	6.08	12.75	1.30	6.31	13.88	1.29	8.65	19.63
	LR-HPT	1.13	5.21	10.60	1.14	5.40	12.09	1.09	6.67	16.10
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 2, \nu = 2$										
N = 500	LR-RE	1.04	5.74	11.55	1.02	6.72	15.05	3.35	20.47	39.11
	LR-HPT	0.95	5.04	10.25	0.90	5.40	12.08	1.40	12.39	29.07
N = 100	LR-RE	1.46	6.15	12.76	1.34	6.41	14.17	1.48	9.15	20.41
	LR-HPT	1.13	5.21	10.60	1.14	5.40	12.09	1.09	6.67	16.10
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 2, \nu = 4$										
N = 500	LR-RE	1.04	5.74	11.55	1.00	6.28	14.34	2.73	18.98	37.98
	LR-HPT	0.95	5.04	10.25	0.86	5.37	12.30	1.67	14.55	32.13
N = 100	LR-RE	1.46	6.15	12.76	1.35	6.30	13.89	1.33	8.87	19.89
	LR-HPT	1.13	5.21	10.60	1.17	5.46	12.12	1.13	7.06	16.16

Notes: Table 3 reports the size and power (%) of the LR tests based on the RE and HPT ML estimation with their *p*-values from a mixture of $\chi^2(1)$ and $\chi^2(0)$ over 10,000 simulations. The variables, η_i, q_{i0}, and ε_{it}, are generated from the normal distributions with zero means and the variances, σ_η^2, σ_{q0}^2, and ν, respectively. The data are generated without trends and with zero means. The ML estimators are estimated without trends and intercept terms.

Table 4

LR Test Based on the RE ML Estimation with Homogenous Trends: Testing the Composite Hypothesis of Unit Root, H_o: $\delta_o = 1$, and $\omega_o = \psi_o = \beta_o = 0$

		$\delta_o = 1$			$\delta_o = 0.95$			$\delta_o = 0.9$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 2, \nu = 1$										
N=500	$T = 5$	1.23	4.77	9.57	79.89	92.57	96.04	100.0	100.0	100.0
	$T = 10$	1.13	4.78	9.31	100.0	100.0	100.0	100.0	100.0	100.0
	$T = 25$	0.86	4.57	9.18	100.0	100.0	100.0	100.0	100.0	100.0
N=100	$T = 5$	1.15	5.62	11.78	11.48	29.20	41.79	56.80	79.64	87.78
	$T = 10$	1.32	6.12	11.78	53.40	77.35	86.56	99.50	99.94	99.97
	$T = 25$	1.32	6.55	12.62	99.97	100.0	100.0	100.0	100.0	100.0
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 4, \nu = 1$										
N=500	$T = 5$	1.05	5.49	10.19	96.49	99.16	99.73	100.0	100.0	100.0
	$T = 10$	0.93	4.94	9.67	100.0	100.0	100.0	100.0	100.0	100.0
	$T = 25$	1.03	4.73	9.43	100.0	100.0	100.0	100.0	100.0	100.0
N=100	$T = 5$	1.39	6.47	11.60	21.20	43.83	56.88	81.26	93.65	97.07
	$T = 10$	1.28	5.83	11.4	67.85	87.78	93.22	99.97	99.99	100.0
	$T = 25$	1.20	5.95	11.25	100.0	100.0	100.0	100.0	100.0	100.0
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 1, \nu = 1$										
N=500	$T = 5$	1.14	5.14	9.54	66.17	85.09	91.33	99.96	100.0	100.0
	$T = 10$	0.94	4.49	8.54	99.90	100.0	100.0	100.0	100.0	100.0
	$T = 25$	1.07	4.80	9.42	100.0	100.0	100.0	100.0	100.0	100.0
N=100	$T = 5$	1.37	6.14	12.41	9.51	25.44	37.51	44.69	69.12	79.78
	$T = 10$	1.37	6.33	12.18	47.87	73.35	83.53	99.19	99.85	99.97
	$T = 25$	1.24	6.43	12.1	99.99	100.0	100.0	100.0	100.0	100.0
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 2, \nu = 2$										
N=500	$T = 5$	1.19	5.39	10.25	70.82	88.08	93.65	100.0	100.0	100.0
	$T = 10$	1.08	5.06	10.01	100.0	100.0	100.0	100.0	100.0	100.0
	$T = 25$	0.93	4.69	9.55	100.0	100.0	100.0	100.0	100.0	100.0
N=100	$T = 5$	1.34	5.99	11.66	10.14	25.98	37.52	49.39	74.41	84.23
	$T = 10$	1.32	6.05	11.83	51.24	76.04	85.17	99.49	99.93	99.98
	$T = 25$	1.25	6.08	11.90	100.0	100.0	100.0	100.0	100.0	100.0
$\sigma_\eta^2 = 2, \sigma_{q0}^2 = 2, \nu = 4$										
N=500	$T = 5$	1.02	5.07	10.01	66.18	84.89	91.43	100.0	100.0	100.0
	$T = 10$	1.19	5.77	11.16	99.96	99.99	99.99	100.0	100.0	100.0
	$T = 25$	0.94	5.02	9.97	100.0	100.0	100.0	100.0	100.0	100.0
N=100	$T = 5$	1.09	5.70	11.04	9.23	24.14	36.07	47.17	71.92	82.57
	$T = 10$	1.15	5.65	11.16	48.48	73.80	83.84	99.62	99.96	100.0
	$T = 25$	1.25	5.68	11.16	100.0	100.0	100.0	100.0	100.0	100.0

Notes: Table 4 reports the size and power of the LR test based on the RE ML estimation with homogeneous trends with the *p*-values from a mixture of $\chi^2(4)$ and $\chi^2(3)$ over 10,000 simulations. The variables, η_i, q_{i0}, and ε_{it}, are generated from the normal distributions with zero means and the variances, σ_η^2, σ_{q0}^2, and ν, respectively. The trend parameter g is fixed at 0.5. The ML estimators are computed allowing a common time trend and a nonzero intercept.

Table 5
LR Test Based on RE ML with Heterogeneous Trends:
Testing the Composite Hypothesis of Unit Root, H$_o$: $\delta_o - 1 = \omega_{12,o} = \omega_{22,o} = \psi_{2,o} = a_{2,o} = 0$

	Test	$\delta_o = 1$ 1%	5%	10%	$\delta_o = 0.95$ 1%	5%	10%	$\delta_o = 0.9$ 1%	5%	10%
$\sigma_g^2 = 0$										
N=500										
$T = 5$	LR-HT	1.14	5.63	10.76	1.28	6.28	12.01	2.35	9.98	17.28
	T-HT	(1.06)	(5.13)	(10.74)	(1.35)	(5.95)	(11.17)	(1.82)	(8.34)	(15.52)
$T = 10$	LR-HT	1.43	5.79	10.99	2.48	9.64	17.41	28.40	52.92	65.95
	T-HT	(1.11)	(5.19)	(10.06)	(2.05)	(8.74)	(15.79)	(9.37)	(27.12)	(40.26)
$T = 25$	LR-HT	1.10	5.46	10.72	64.84	84.46	90.91	100.0	100.0	100.0
	T-HT	(1.08)	(5.26)	(10.22)	(22.23)	(47.13)	(60.61)	(99.50)	(99.90)	(99.96)
N = 100										
$T = 5$	LR-HT	1.41	6.38	11.74	1.30	6.26	12.22	1.63	7.8	14.13
	T-HT	(1.01)	(5.30)	(10.30)	(1.05)	(5.26)	(10.65)	(1.47)	(6.44)	(12.45)
$T = 10$	LR-HT	1.42	6.54	12.59	1.73	7.38	13.39	4.00	14.13	23.49
	T-HT	(1.16)	(5.33)	(11.10)	(1.85)	(7.58)	(13.95)	(3.64)	(12.04)	(21.16)
$T = 25$	LR-HT	1.45	6.52	12.60	8.79	24.21	36.33	82.85	94.68	97.41
	T-HT	(1.40)	(5.69)	(10.90)	(6.04)	(17.76)	(2918)	(44.13)	(70.93)	(81.64)
$\sigma_g^2 = 0.5$										
N=500										
$T = 5$	LR-HT	1.18	5.69	11.32	1.36	6.42	11.86	2.42	8.73	15.97
	T-HT	(1.20)	(5.42)	(10.51)	(1.13)	(5.66)	(11.34)	(1.82)	(8.85)	(15.92)
$T = 10$	LR-HT	1.36	5.81	11.15	2.08	8.02	14.87	12.92	31.70	44.60
	T-HT	(0.92)	(5.07)	(9.89)	(2.25)	(9.39)	(17.02)	(9.49)	(26.42)	(40.22)
$T = 25$	LR-HT	1.38	6.07	11.43	25.25	48.37	60.87	100.0	100.0	100.0
	T-HT	(1.25)	(5.51)	(10.59)	(22.64)	(47.65)	(61.75)	(99.49)	(99.98)	(99.99)
N = 100										
$T = 5$	LR-HT	1.37	6.31	12.87	1.88	7.67	14.02	1.72	7.81	14.28
	T-HT	(1.01)	(5.01)	(10.17)	(1.22)	(5.61)	(10.89)	(1.41)	(6.85)	(13.13)
$T = 10$	LR-HT	1.46	6.78	12.85	1.63	7.42	13.91	2.70	10.89	19.42
	T-HT	(1.52)	(6.17)	(11.47)	(1.85)	(7.58)	(13.95)	(3.64)	(12.04)	(21.16)
$T = 25$	LR-HT	1.44	6.79	12.62	4.24	14.06	23.22	48.93	73.59	83.67
	T-HT	(1.40)	(5.69)	(10.90)	(6.04)	(17.81)	(29.04)	(45.30)	(70.48)	(81.75)
$\sigma_g^2 = 1$										
N=500										
$T = 5$	LR-HT	1.30	6.17	12.02	1.34	6.46	12.04	2.04	8.29	14.76
	T-HT	(1.07)	(5.11)	(10.45)	(1.22)	(5.97)	(11.43)	(1.80)	(8.01)	(15.28)
$T = 10$	LR-HT	1.42	6.15	11.94	1.89	7.71	14.76	13.08	30.41	42.68
	T-HT	(1.28)	(5.62)	(11.02)	(2.38)	(9.32)	(16.48)	(10.44)	(27.58)	(40.25)
$T = 25$	LR-HT	1.23	5.62	11.78	23.64	47.43	60.02	99.99	100.0	100.0
	T-HT	(1.02)	(5.38)	(10.62)	(22.05)	(46.12)	(60.40)	(99.38)	(99.94)	(99.99)
N = 100										
$T = 5$	LR-HT	1.53	6.97	13.48	1.40	6.67	12.99	1.46	7.39	14.32
	T-HT	(1.20)	(5.17)	(10.33)	(0.87)	(5.54)	(10.74)	(1.33)	(6.54)	(12.40)
$T = 10$	LR-HT	1.45	6.90	12.97	1.63	7.40	13.96	2.98	11.38	19.84
	T-HT	(1.16)	(5.33)	(11.10)	(1.85)	(7.58)	(13.95)	(3.38)	(12.51)	(21.59)
$T = 25$	LR-HT	1.45	6.82	12.67	4.19	13.99	23.09	48.49	73.12	83.42
	T-HT	(1.40)	(5.69)	(10.90)	(6.04)	(17.81)	(29.04)	(45.30)	(70.48)	(81.75)

Notes: Table 5 reports the size and power of the LR test based on the RE ML estimation with the *p*-values from a mixture of $\chi^2(4)$ and $\chi^2(5)$ over 10,000 simulations. The estimators are computed allowing heterogeneous time trends and nonzero intercepts. For comparison, the size and power of a one-tailed and bias-corrected *t*-test (BT-HT) of Harris and Tzavalis (1999) are reported in parentheses. The *t*-statistic, which is obtained by correcting the bias in a within-type estimator, is asymptotically standard normal under the null hypothesis of unit root with heterogeneous drifts, For each simulation, the variables, η_i, q_{i0}, and ε_{it} are generated from the normal distributions with zero means and the variances, $\sigma_\eta^2 = 2$, $\sigma_{q0}^2 = 0$, and $\nu = 1$, respectively. The trend parameters g_i are drawn from the normal distribution $N(0.5, \sigma_g^2)$.