PANEL LM UNIT ROOT TESTS WITH LEVEL SHIFTS

Kyung So Im, Junsoo Lee and Margie Tieslau[¤]

December 26, 2002

Abstract

This paper proposes a new panel LM unit root test with structural shifts in level. We show that the asymptotic distribution of the proposed panel test with level shifts is the same as that of the test without level shifts. This invariance property permits us to use the same critical values of the panel LM test with no shift, regardless of locations of structural shifts. As such, our proposed panel LM test is robust to the presence of heterogeneous stuructural shifts.

JEL Classification: C12, C15, C22

Key Words: Heterogeneous Dynamic Panels, Panel Unit Root Test, LM Test, Structural Change, Level Shift.

[¤]Associate Professor at the University of Central Florida; Associate Professor at the University of Central Florida, and Associate Professor at the Univerisy of North Texas; Corresponding author: Junsoo Lee, Department of Economics, University of Central Florida, Orlando, FL 32816, e-mail: Junsoo.Lee@Bus.ucf.edu, phone: 407-823-2070, fax: 407-823-3269.

1. Introduction

Recent developments in unit root tests for panel data have attracted much attention in both theoretical and applied works. For example, the tests proposed by Levin, Lin and Chu (2002, hereafter LLC), Im, Pesaran and Shin (2003, hereafter IPS), as well as other extensions, have been widely applied in the literature; see Baltagi and Kao (2000) for a survey. Applications using these tests have focused on various key economic issues, such as purchasing power parity and income convergence, with the hope that the increased power of these tests would provide more reliable evidence on these important theories.

Another significant line of research that has developed in the last decade in the unit root literature is testing for a unit root in the presence of structural changes. Since the pioneering work of Perron (1989), the importance of allowing for a structural change in testing for a unit root has been widely documented in the literature. Failure to allow for an existing structural change can lead to biased estimates of the parameters in the unit root testing regression and a significant loss of power. Potentially, this problem also exists when testing for a unit root in a panel framework. Consider, for example, the case of the IPS test. The IPS test statistic is obtained as a linear combination of the univariate unit root test statistics not allowing for structural changes. As such, the IPS test is expected to suffer from a similar loss of power and produce biased estimates if a structural shift does in fact occur. This problem occurs when the panel test statistics are obtained from the univariate unit root test statistics ignoring existing structural shifts. The same problem would occur in other panel unit root tests of any type.

Consequently, one may be tempted to allow for structural changes in the panel setting as the way in the univariate case. However, allowing for structural changes in the panel setting is not an easy task. To the best of our knowledge, no significant development has been made for a panel unit root test that can allow for structural changes while providing relevant asymptotic results. One possible reason for this lies in the difficulty of constructing a valid panel test that is free of the nuisance parameters indicating the locations of the structural changes. One may possibly consider modifying the IPS or LLC test for which a dummy variable is included in each ADF regression to control for the effect of structural changes. However, this approach poses a problem. It is well known that the limiting distribution of Perron's ADF type t-statistic depends on the location of structural changes that occur in the data. Thus, in order to make use of an IPS-type unit root test that allows for structural changes, one would need to compute the expected values and variances of the ADF t-statistics at all different possible shift locations for each cross-section unit in the sample. This task is extremely cumbersome and, practically, not feasible.

In this paper we develop new panel unit root tests based on the Lagrangian Multiplier (LM) principle. First, we consider the test without structural changes and then extend the test by allowing for structural shifts in level. When dealing with structural shifts, our panel LM test has an important feature that cannot be found in the ADF-type extension tests. In particular, the asymptotic distribution of the panel LM test does not depend on the nuisance parameters that indicate the positions of the structural shifts. A univariate LM unit root test was initially proposed by Schmidt and Phillips (1992). Amsler and Lee (1995) showed that the extended version of the univariate LM unit root test allowing for a shift in level has the same asymptotic distribution of the Schmidt and Phillips test. We show that this invariance property of the LM univariate unit root test carries over to our new panel LM test.

This is a salient feature of the panel LM test and it has an important practical implication. As a result of the invariance property, the panel LM test with level shifts can make use of the same critical values that apply to the panel LM test without any shifts. In this case, the researcher is spared the task of having to simulate different critical values for all possible shift locations. That is, one may use the same expected values and variances that are computed for the baseline case where no structural change is present in the data—even when a structural shift is allowed at any location. Also, our new panel LM test is quite flexible. It can be applied when more than one structural shift occurs at different time periods for each cross-section unit as well as when different numbers of structural shifts (including no shift) occur for each cross-section unit. Thus, this interesting feature of the panel LM test permits us to construct a valid panel unit root test which is also practically very useful. This feature is not available in the IPS test or in other extensions.

Further, we investigate how closely and under what conditions this invariance result is satisfied in the panel framework. We find that the panel LM test remains valid as long as N/T does not diverge as $N, T \to \infty$, and the second moment of the individual LM statistic exists for all $T \geq T_0$, for some finite T_0. For both of our panel LM tests, with or without structural shifts, we can construct a convenient standard normal test through appropriate standardization since the sum of individual statistics approaches a normal random variable as N increases. The panel LM statistic follows a standard normal distribution under the null hypothesis as $N \to \infty$ (for fixed T), as long as the second moment of the individual LM statistic exists. We provide the asymptotic results with the case of one structural shift but the results can be applied to the case with more than one structural shift. We also examine the finite sample performance of the LM test via Monte Carlo simulations. All empirical sizes of the LM tests that properly control for structural shifts are reasonably close to the nominal size in most of the cases we examine.

The rest of this paper is organized as follows. In the next section we present the model and derive the panel LM statistic with no structural change. In Section 3, we extend the test to allow for the case where a structural shift is present in each cross-section unit. Section 4 reports the simulation results that examine the finite sample performance of the LM test. In Section 5, we apply our panel LM test with structural shifts to investigate the empirical validity of one of the most significant theories in international macroeconomics: purchasing power parity. Section 6 concludes.

2. The Panel LM Test without Shifts

In this section, we develop a new panel LM statistic without any structural shifts. This test is a panel version of the univariate LM unit root test of Schmidt and Phillips (1992), just as the IPS test is a panel version of the (augmented) Dickey-Fuller test. To begin with, we assume that the series y_{it} is generated according to:

$$y_{it} = z_{it} + x_{it}; \quad z_{it} = \gamma_{1i} + \gamma_{2i} t; \qquad x_{it} = \phi_i x_{i,t-1} + \varepsilon_{it} \tag{2.1}$$

where t indexes time series units such that $t = 1, 2, ..., T$, and i indexes cross section units such that $i = 1, 2, ..., N$. This data generating process (DGP) takes an unobserved representation form. We are interested in testing the null hypothesis of unit roots, implying $\phi_i = 1$ for all i. Rearranging (2.1) yields the following:

$$\Delta y_{it} = \beta_i y_{i,t-1} - \beta_i \gamma_{1i} + [1 - (\beta_i + 1)(t-1)]\gamma_{2i} + \varepsilon_{it}; \; t = 1, 2, ..., T; \; i = 1, 2, ..., N; \tag{2.2}$$

where $\beta_i = -(1 - \phi_i)$. In this case, the null hypothesis can be expressed as:

$$H_0 : \beta_i = 0 \quad \text{for all } i \tag{2.3}$$

against the alternative hypothesis:

$$H_1 : \beta_i < 0; \text{ for at least one } i \tag{2.4}$$

implying that at least one time series is trend-stationary.

Here, we derive the LM statistic in the basic case when the errors ε_{it} in (2.2) are serially uncorrelated. We assume:

Assumption 2.1. ε_{it}, $i = 1, ..., N$; $t = 1, ..., T$, are independent normal variables with mean zero and variance σ_i^2.

This leads to the following pooled log-likelihood function:

$$\ln L = \sum_{i=1}^{N} \left(-\frac{T}{2} \ln 2\pi\sigma_i^2 - \frac{1}{2\sigma_i^2} SSE_i \right); \tag{2.5}$$

where

$$SSE_i = \sum_{t=1}^{T} \{\Delta y_{it} - \beta_i y_{i,t-1} + \beta_i \gamma_{1i} - [1 - (\beta_i + 1)(t-1)]\gamma_{2i}\}^2 : \tag{2.6}$$

Let LM_{iT} be the LM statistic for the i-th time series. It is then straightforward to see that the LM statistic drived from the pooled likelihood function (2.5) becomes

$$LM_{NT} = \sum_{i=1}^{N} LM_{iT}. \tag{2.7}$$

Let

$$\tilde{S}_{i,t-1} = y_{i,t-1} - \tilde{\gamma}_{2i}(t-1), \tag{2.8}$$

where $\tilde{\gamma}_{2i} = \frac{1}{T}\sum_{t=1}^{T} \Delta y_{it} = (y_{iT} - y_{i0})/T$ is the restricted maximum likelihood estimator of γ_{2i}, obtained from the restricted regression:

$$\Delta y_{it} = \gamma_{2i} + \varepsilon_{it}. \tag{2.9}$$

Following Schmidt and Phillips (1992), we obtain the LM statistic for the i-th time series in the regression:

$$\Delta y_{it} = \text{intercept} + \beta_i \tilde{S}_{i,t-1} + \text{error}, \tag{2.10}$$

which can be expressed as:

$$LM_{iT} = \frac{\sqrt{T-2}\left(\tilde{S}'_{i,-1} M_{(i_T)} \Delta Y_i\right)}{\sqrt{\left(\tilde{S}'_{-1} M_{(i_T)} \tilde{S}_{-1}\right)\left(\Delta Y_i' M_{(i_T,\tilde{S}_{-1})} \Delta Y_i\right)}}, \tag{2.11}$$

where $\Delta Y_i = (\Delta y_{i1}, \Delta y_{i2}, \ldots, \Delta y_{iT})'$, $\tilde{S}_{i,-1} = \left(\tilde{S}_{i0}, \tilde{S}_{i1}, \tilde{S}_{i2}, \ldots, \tilde{S}_{i,T-1}\right)'$, $M_{(\cdot)}$ is a projection matrix onto the null space of $(\cdot)$, and i_T is a $T \times 1$ vector of ones.

The distribution of LM_{NT} in (2.7) depends on N and T, but it is free of other parameters under the null model. Therefore, LM_{NT} itself can be used in practice as a test statistic. However, we can construct more convenient test statistics. We denote the average of the individual LM statistics LM_{iT} as

$$\overline{LM}_{NT} = \frac{1}{N}\sum_{i=1}^{N} LM_{iT}. \tag{2.12}$$

In addition, we denote the expected value and variance of LM_{iT} under the null hypothesis as $E(L_T)$ and $V(L_T)$. More formally, these terms are defined as follows:

Definition 1. $E(L_T)$ and $V(L_T)$ are the mean and variance of the LM statistic obtained as the t-statistic testing $\beta = 0$ in the regression (2.10), where y_t follows:

$$\Delta y_t = \varepsilon_t, \text{ with } \varepsilon_t \sim \text{iid } N\left(0, \sigma^2\right) \text{ for } t = 1, 2, \ldots, T, \text{ and } y_0 = 0. \tag{2.13}$$

Then, under Assumption 2.1 and the null hypothesis, we have:

$$\Gamma_{LM} = \frac{\sqrt{N}\left[\overline{LM}_{NT} - E(L_T)\right]}{\sqrt{V(L_T)}} \Rightarrow N(0,1) \tag{2.14}$$

as N increases (for finite T), as long as $E(L_T)$ and $V(L_T)$ exist.

Remark 1. The existence of $E(L_T)$ and $V(L_T)$ is a question of interest. The existence of these moments can be shown easily in our test statistic based on the LM approach. Schmidt and Phillips (1992) suggest estimating σ_i^2 based on the unrestricted SSE_i in (2.6). However, following the LM principle, we may estimate σ_i^2 alternatively by using the restricted SSE_i from the restricted regression (2.9); namely, $\tilde{\sigma}_i^2 = T^{-1}\sum_{t=1}^{T}(\Delta y_{it} - \tilde{\nu}_{2i})^2$. Then we can have another type of LM statistic: $LM_{iT}^{*} = \sqrt{T}\,\tilde{S}_{i,-1}' M_{(\iota_T)} \Delta Y_i \Big/ \sqrt{\left(\tilde{S}_{-1}' M_{(\iota_T)} \tilde{S}_{-1}\right)\left(\Delta Y_i' M_{(\iota_T)} \Delta Y_i\right)}$. Note that the square of LM_{iT}^{*} is the familiar TR^2 statistic in the regression (2.10). Since LM_{iT}^{*} cannot exceed $\sqrt{T}$, all of its moments will be finite so long as T is finite. We continue to use LM_{iT}, because the difference between LM_{iT} and LM_{iT}^{*} is negligible; it is easily seen that $LM_{iT} = LM_{iT}^{*} + O_p(T^{-1})$.

Remark 2. As noted in Schmidt and Phillips (1992), the statistic LM_{iT} is invariant numerically to the values of γ_{1i} and γ_{2i} under the null hypothesis. Therefore, without loss of generality, we can set $\gamma_{1i} = \gamma_{2i} = 0$ in the DGP (2.13).

We next assume that the errors ε_{it} in (2.2) follow an autoregressive process.[1]

Assumption 2.2. Assume that $\varepsilon_{it} = \sum_{j=1}^{p_i} \rho_{ij}\varepsilon_{it-j} + e_{it}$; $i = 1, \ldots, N$; $t = 1, \ldots, T$; where e_{it} are independent normal variables with mean zero and variance σ_i^2, and all the roots of $\rho_i(z) = 1 - \sum_{j=1}^{p_i} \rho_{ij} z^j$ lie outside the unit circle.

We follow Ahn (1993) and Amsler and Lee (1995) in suggesting an ADF type correction for serially correlated errors in (2.2). The LM statistic for the i-th time series is obtained as a t-statistic for $\beta_i = 0$ in the augmented regression:

$$\Delta y_{it} = \text{intercept} + \beta_i \tilde{S}_{i,t-1} + \sum_{j=1}^{p_i} \rho_{ij}\Delta y_{i,t-j} + \text{error}, \tag{2.15}$$

where p_i denotes the order of augmentations for the i-th times series, and $\tilde{S}_{i,t-1}$ is defined in (2.8).[2] See Remark 3 below for more details of the construction of

[1] In this paper, we do not pursue the asymptotics of the LM test when the errors are serially correlated. IPS present some simulation results that the IPS test is valid in the presence of AR(p) errors as long as N=T converges to any finite number as $N, T \to \infty$. Similarly, the panel LM test can be shown to be valid under the same conditions for AR(p) errors.

[2] Amsler and Lee (1995) suggest to augment $\Delta\tilde{S}_{i,t-j}$ rather than $\Delta y_{i,t-j}$. In the case where no structural break is involved, both terms produce numerically identical test statistics.

$\hat{S}_{i,t-1}$ in practice. Denoting the resulting t-statistic as $LM_{iT}(p_i)$, we obtain the average statisitc as:

$$\overline{LM}_{NT}(p) = \frac{1}{N}\sum_{i=1}^{N} LM_{iT}(p_i). \tag{2.16}$$

With regard to standardization, we follow the procedure as proposed by IPS. To do so, we utilize the following definitions of the expected value and variance of $LM_{iT}(p_i)$, when $\beta_i = 0$ and $\rho_{ij} = 0$, for $j = 1, ..., p_i$.

Definition 2. $E[L_T(p_i)]$ and $V[L_T(p_i)]$ are the mean and variance of the LM statistic obtained as the t-statistic for testing $\beta_i = 0$ in the regression (2.15), where y_{it} follows:

$$\Delta y_{it} = \varepsilon_{it}, \text{ with } \varepsilon_{it} \sim iidN\left(0, \sigma_i^2\right), \text{ for } t = 1, 2, ..., T, \text{ and } y_{i0} = 0. \tag{2.17}$$

Then, standardizing $\overline{LM}_{NT}(p)$ based on $E[L_T(p_i)]$ and $V[L_T(p_i)]$ yields:

$$\Gamma_{LM}(p) = \frac{\sqrt{N}\left\{\overline{LM}_{NT}(p) - \frac{1}{N}\sum_{i=1}^{N} E[L_T(p_i)]\right\}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} V[L_T(p_i)]}}. \tag{2.18}$$

As before, the statistic $LM_{iT}(p_i)$ is numerically invariant to the values of γ_{1i} and γ_{2i} (see Remark 2). The expected values and variances, $E[L_T(p)]$ and $V[L_T(p)]$, were computed for various combinations of T and p, via stochastic simulations using 500,000 replications. These are provided in Table 1 (see Remark 3 below for more details). Obviously, Γ_{LM} in (2.14) is a special case of $\Gamma_{LM}(p)$ when $p_i = 0$, for all i.

Remark 3. It will be helpful to practitioners to explain more details of Table 1.

(1) The value of T in Table 1 denotes the regression dimension excluding initial observations, rather than the actual number of observations in the sample. For example, suppose one has 25 observations for the i-th time series and wants to allow $p_i = 2$. Then, one loses the first three observations and the regression dimension becomes 22. Thus, the appropriate expected value and variance corresponding to $T = 22$ and $p = 2$ are -1.880 and 0.413, respectively.

(2) In computing $E[L_T(p)]$ and $V[L_T(p)]$ (with $T+p+1$ observations available), we construct the series $\hat{S}_{t-1}$ as $\hat{S}_{t-1} = y_{t-1} - (y_{T+p+1} - y_1)/(T+p)$.

(3) When $T \geq 50$, the expected value and variance are reported for selective values of T. For other values of T, they can be interpolated. For example, suppose that one has 63 observations for each time series and allows for $p = 4$. In this case, the regression dimension is 58. From Table 1, the expected value is -1.894 for $T = 55$, $p = 4$, and -1.902 for $T = 60$, $p = 4$, respectively. The interpolated expected value is obtained as: $\left(\frac{2}{5}\right)(-1.894) + \left(\frac{3}{5}\right)(-1.902) = -1.899$.

3. The Panel LM Test with Level Shifts

We now derive an extension test of our panel LM unit root test where a level shift occurs in each individual time series. We will show that the asymptotic distribution of the panel LM test with a shift in level is the same as that of the baseline panel LM test not allowing for any shifts. This is a generalization of the invariance result of Amsler and Lee (1995) who have shown that the asymptotic null distribution of the univariate LM statistic with a mean shift is the same as that of the LM statistic without a shift. Then, we are interested in examining how closely and under what conditions the invariance property of the univariate LM test for a single time series carries over to the panel LM unit root test. We demonstrate the relationship for the case of a one-time shift in the mean, but the result can be generalized for cases with a finite number of mean (level) shifts. We begin with the test for which the errors are serially uncorrelated.

Suppose a structural shift in the mean occurs at time period $T_{B,i}$ in the i-th time series. Then, the data generating process is given as:

$$y_{it} = z_{it} + x_{it}, \quad z_{it} = \gamma_{1i} + \gamma_{2i} t + \delta_i D_{it}, \quad x_{it} = \phi_i x_{i,t-1} + \varepsilon_{it}, \tag{3.1}$$

for $t = 0, 1, 2, ..., T$; $i = 1, 2, ..., N$; where

$$D_{it} = \begin{cases} 0 & t < T_{B,i} \\ 1 & t \geq T_{B,i} + 1 \end{cases} . \tag{3.2}$$

The above equation can be expressed as:

$$\Delta y_{it} = \beta_i y_{i,t-1} - \beta_i \gamma_{1i} + [1 - (\beta_i + 1)(t-1)]\gamma_{2i} + (\Delta D_{it} - \beta_i D_{i,t-1})\delta_i + \varepsilon_{it}, \tag{3.3}$$

for $t = 1, 2, ..., T$; $i = 1, 2, ..., N$; where $\Delta D_{it} = D_{it} - D_{i,t-1}$, i.e.,

$$\Delta D_{it} = \begin{cases} 1 & t = T_{B,i} + 1 \\ 0 & \text{otherwise} \end{cases} . \tag{3.4}$$

>From (3.3), and under Assumption 2.1, we can similarly obtain the pooled likelihood function as in (2.5) with

$$SSE_i = \sum_{t=1}^{T} \{\Delta y_{it} - \beta_i y_{i,t-1} + \beta_i \gamma_{1i} - [1 - (\beta_i + 1)(t-1)]\gamma_{2i} - (\Delta D_{it} - \beta_i D_{i,t-1})\delta_i\}^2 . \tag{3.5}$$

As Amsler and Lee (1995) showed, the LM statistic for the i-th time series can be obtained as a t-statistic testing $\beta_i = 0$ in the regression:

$$\Delta y_{it} = \gamma_{2i} + \delta_i \Delta D_{it} + \beta_i \tilde{S}_{i,t-1} + \text{error}, \tag{3.6}$$

where

$$\tilde{S}_{i,t-1} = y_{i,t-1} - \tilde{\gamma}_{2i}(t-1) - \tilde{\delta}_i D_{i,t-1}, \tag{3.7}$$

and $\tilde{\gamma}_{2i}$ and $\tilde{\delta}_i$ are obtained as the OLS estimators of γ_{2i} and δ_i in the restricted regression:

$$\Delta y_{it} = \gamma_{2i} + \delta_i \Delta D_{it} + \varepsilon_{it}. \tag{3.8}$$

Letting $\tilde{S}_{i,-1} = \left(\tilde{S}_{i0}, \tilde{S}_{i1}, ..., \tilde{S}_{i,T-1}\right)'$ and $\Delta D_i = (\Delta D_{i1}, \Delta D_{i2}, ..., D_{iT})'$, we form the LM statistic for the i-th time series as:

$$LM_{iT}^{B} = \frac{\sqrt{T-3}\left(\tilde{S}_{i,-1}' M_{(\iota_T, \Delta D_i)} \Delta Y_i\right)}{\sqrt{\left(\tilde{S}_{i,-1}' M_{(\iota_T, \Delta D_i)} \tilde{S}_{i,-1}\right)\left(\Delta Y_i' M_{(\iota_T, \Delta D_i, \tilde{S}_{i,-1})} \Delta Y_i\right)}}. \tag{3.9}$$

Then, the panel LM statistic based on the pooled likelihood function is given by the sum of LM_{iT}^{B} so that $LM_{NT}^{B} = \sum_{i=1}^{N} LM_{iT}^{B}$. We consider:

$$\overline{LM}_{NT}^{B} = \frac{1}{N} \sum_{i=1}^{N} LM_{iT}^{B}. \tag{3.10}$$

Following Amsler and Lee (1995), we can show that the limiting distribution of LM_{iT}^{B} does not depend on the parameter indicating the location of the shift point:

$$\lambda_i = \frac{T_{B,i}}{T}. \tag{3.11}$$

As a result, the distribution of LM_{iT}^{B} asymptotically approaches that of LM_{iT}, implying that their asymptotic distributions are the same. Specifically, the difference between the LM statistic allowing for a shift and the LM statistic not allowing for a shift is asymptotically negligible: that is, $LM_{iT}^{B} - LM_{iT} = o_p(1)$. Only in strictly finite samples, the distribution of LM_{iT}^{B} can be seen as being dependent on λ_i. If we have the exact expected value and the exact variance of LM_{iT}^{B} under the null hypothesis, which we denote as $E\left[L_T^B(\lambda_i)\right]$ and $V\left[L_T^B(\lambda_i)\right]$, then it follows that:

$$\Gamma_{LM}^{B*} = \frac{\sqrt{N}\left\{\overline{LM}_{NT}^{B} - \frac{1}{N}\sum_{i=1}^{N} E\left[L_T^B(\lambda_i)\right]\right\}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} V\left[L_T^B(\lambda_i)\right]}} \Rightarrow N(0,1), \tag{3.12}$$

under the null hypothesis, as $N \to \infty$, as long as $V\left[L_T^B(\lambda_i)\right]$ exists for all i. However, the statistic Γ_{LM}^{B*}, as it stands, is not very practical since it requires $E\left[L_T^B(\lambda_i)\right]$ and $V\left[L_T^B(\lambda_i)\right]$ for all different combinations of λ_i in the sample. Due to this, we can consider a practical statistic:

$$\Gamma_{LM}^{B} = \frac{\sqrt{N}\left[\overline{LM}_{NT}^{B} - E(L_T)\right]}{\sqrt{V(L_T)}}. \tag{3.13}$$

where $E(L_T)$ and $V(L_T)$ are defined in Definition 1. The question of interest is under what condition the following result holds:

$$\Psi^B_{LM} - \Psi^{B*}_{LM} = o_p(1). \tag{3.14}$$

When this result holds, and upon applying the standardizations that use $E(L_T)$ and $V(L_T)$ as in the case of the panel LM statistic without shifts, we can show that the panel LM statistic with shifts follows a normal distribution:

$$\Psi^B_{LM} \Rightarrow N(0,1). \tag{3.15}$$

Thus, we are interested in finding the condition that guarantees (3.14). For this, we make the following assumption:

Assumption 3.1. The variance of L_T defined in Definition 1 and the variance of $L^B_T(\lambda)$ in (3.12), for all values of λ, are finite for all $T \geq T_0$ for some finite T_0. (See Remark 1.)

Note that

$$\begin{aligned}
\Psi^B_{LM} - \Psi^{B*}_{LM} &= \frac{\sqrt{N}\left[\overline{LM}^B_{NT} - E(L_T)\right]}{\sqrt{V(L_T)}} - \frac{\sqrt{N}\left\{\overline{LM}^B_{NT} - \frac{1}{N}\sum_{i=1}^N E\left[L^B_T(\lambda_i)\right]\right\}}{\sqrt{\frac{1}{N}\sum_{i=1}^N V[L^B_T(\lambda_i)]}} \\
&= \sqrt{N}\left(\overline{LM}^B_{NT} - \frac{1}{N}\sum_{i=1}^N E\left[L^B_T(\lambda_i)\right]\right)\frac{\sqrt{\frac{1}{N}\sum_{i=1}^N V[L^B_T(\lambda_i)]} - \sqrt{V(L_T)}}{\sqrt{\frac{1}{N}\sum_{i=1}^N V[L^B_T(\lambda_i)]}\sqrt{V(L^B_T)}} \\
&\quad + \frac{N^{-1/2}\sum_{i=1}^N E\left[L^B_T(\lambda_i) - L_T\right]}{\sqrt{V(L_T)}}.
\end{aligned} \tag{3.16}$$

It is shown in the Appendix that

$$L^B_T(\lambda) - L_T = O_p\left(T^{-1/2}\right), \text{ for all } \lambda. \tag{3.17}$$

Assumption 3.1 ensures that $V\left[L^B_T(\lambda)\right] - V[L_T] = O\left(T^{-1/2}\right)$ for all λ, from which we deduce $\sqrt{\frac{1}{N}\sum_{i=1}^N V[L^B_T(\lambda_i)]} - V(L_T) = O\left(T^{-1/2}\right)$. It is obvious that the first term of the last equation of (3.16) is $o_p(1)$. Now we need to find the condition under which the last term of (3.16) is negligible asymptotically, or $N^{-1/2}\sum_{i=1}^N E\left[L^B_T(\lambda_i) - L_T\right] = o_p(1)$. We show in the Appendix that

$$E\left[L^B_T(\lambda) - L_T\right] = O\left(T^{-1}\right), \text{ for all } \lambda. \tag{3.18}$$

Therefore, $N^{-1/2}\sum_{i=1}^N E\left[L^B_T(\lambda_i) - L_T\right] = O\left(\frac{\sqrt{N}}{T}\right)$, which is $o(1)$ as long as $\frac{\sqrt{N}}{T} \to 0$. Since $\frac{\sqrt{N}}{T} = \sqrt{\frac{N}{T}}T^{-1/2}$, it follows that $N^{-1/2}\sum_{i=1}^N E\left[L^B_T(\lambda_i) - L_T\right] = o_p(1)$ unless N/T diverges at the rate of $\sqrt{T}$ or faster as $N, T \to \infty$.

Remark 4. We show in the Appendix that the results in (3.17) and (3.18) hold: $E\left[L_T^B(\lambda) - L_T\right] = O(T^{-1})$ and $L_T^B(\lambda) - L_T = O(T^{-1/2})$. To confirm these asymptotic results, we have conducted simulations and examine how $TE\left[L_T^B(\lambda) - L_T\right]$ changes as T grows. In the following table, the expected value and standard deviation of $T\left[L_T^B(\lambda) - L_T\right]$ are computed based on 50,000 replications with $\lambda = 0.5$:

T	100	1,000	10,000	100,000
$TE\left[L_T^B(\lambda) - L_T\right]$	-0.012	0.287	-0.254	-0.331
$T\sqrt{V[L_T^B(\lambda) - L_T]}$	10.118	30.910	97.795	307.480

As can be seen in the above table, $TE\left[L_T^B(\lambda) - L_T\right]$ fluctuates within a very narrow interval, while the standard deviation of $T\left[L_T^B(\lambda) - L_T\right]$ shows the apparent tendency of creeping up at the rate of $\sqrt{T}$; which supports the claims in (3.17) and (3.18).

Remark 5. It is not necessary to impose the restriction that structural shifts in the series, should they exist, must occur at the same location. The shift point can occur at different locations for each cross section unit. Thus, our test can allow for heterogeneous structural changes, while using the same critical values in Table 1.

When the errors ε_{it} in (3.3) are serially correlated, their effects can be corrected by augmenting $\Delta S_{i,t-j}$; as suggested by Amsler and Lee (1995). Therefore, the LM statistic for the i-th time series is obtained as a t-statistic for $\beta_i = 0$ in the augmented regression:

$$\Delta y_{it} = \text{intercept} + \delta_i \Delta D_{it} + \beta_i S_{i,t-1} + \sum_{j=1} \rho_{ij} \Delta S_{i,t-j} + \text{error}, \tag{3.19}$$

where $S_{i,t-1}$ is defined in (3.7). We define $LM_{iT}^B(p_i)$ as the t-statistic for $\beta_i = 0$ in regression (3.19), and its average as:

$$\overline{LM}_{NT}^B(p) = \frac{1}{N}\sum_{i=1} LM_{iT}^B(p_i). \tag{3.20}$$

We then follow the standardization procedure described in Section 2.2 to derive

$$\Gamma_{LM}^B(p) = \frac{\sqrt{N}\left\{\overline{LM}_{NT}^B(p) - \frac{1}{N}\sum_{i=1}^N E[L_T(p_i)]\right\}}{\sqrt{\frac{1}{N}\sum_{i=1}^N V[L_T(p_i)]}}, \tag{3.21}$$

where $E[L_T(p_i)]$ and $V[L_T(p_i)]$ are similarly defined as in Definition 2 of Section 2.2. It also follows that

$$\Gamma_{LM}(p) \Rightarrow N(0,1), \tag{3.22}$$

unless N/T diverges as $N,T \to \infty$. It is worth noticing that we do not impose the same number of augmentation terms for each cross section unit. Thus, depending on the autocorrelation structure in each cross-section unit, the number of augmentation terms must be selected jointly along with shift point locations.

4. Simulation Results

We now investigate small sample properties of our panel LM unit root test via Monte Carlo simulations. Three different experiments are conducted. In the first experiment, we simulate the case where the data contain no structural shifts and the errors are serially independent. The first experiment is designed to compare the generic power of the panel LM test vis-a-vis the IPS test in the benchmark case. In the second experiment we allow for a structural shift in the data generation process while the errors are still serially independent. With this, we investigate the consequence of ignoring existing structural shifts as well as the performance of the panel LM test that controls for structural shifts. Special attention is paid to how the required condition of N/T derived in Section 3.1 works in finite samples. The third experiment investigates the case where the errors are serially correlated and a structural shift is present in each time series. For this, we consider only the panel LM test that controls for structural shifts.

In the first and second experiments where no serial correlations are involved, we report the results for 16 different combinations of $N, T = 10, 25, 50, 100$. But, in the third experiment, where the data contain structural shifts and the errors are serially correlated, we drop the case with $T = 10$ because the degrees of freedom become too small. Then, we are left with only 12 combinations of $N = 10, 25, 50, 100$ and $T = 25, 50, 100$. In each case the results are based on 2,000 replications. In each replication we generate N independent time series of $T + 51$ data points using pseudo-iid N(0,1) random numbers from the Gauss RNDNS procedure. The first 50 observations are discarded to avoid a possible initial value effect. Therefore, the regression dimension is $T - p$. All tests are conducted using a 5% nominal size.

Experiment 1: No Shifts, No Serial Correlations

Here, we consider the case where each time series contains no structural shift and the errors are serially uncorrelated. In particular, we wish to compare the generic power of the panel LM test vis-a-vis the IPS test in the baseline case where no structural shifts occur in the data. For this, N independent time series are generated with:

$$y_{it} = \phi_i y_{i,t-1} + \varepsilon_{it}, \quad t = 0, 1, ...T, \quad i = 1, 2, ..., N, \tag{4.1}$$

with $\varepsilon_{it} \sim iid\ N(0,1)$. We set $\phi_i = 1$ in examining the empirical size and $\phi_i = 0.9$ for computing the power of the test for all i. We compare the performance of our panel LM test statistic Γ_{LM} given in (2.14) and the IPS test, which is obtained by standardizing the average of the t-statistics from the DF regressions including an intercept and a linear trend (see IPS, 2003, equations 2.2 and 4.2). The results are reported in Table 2.

Since both the LM and IPS statistics follow the standard normal distribution under the null hypothesis as N increases (with fixed T), we expect the empirical size to be reasonably close to the 5% nominal size in both tests for relatively large N. As expected, all reported sizes are reasonably close to 5%, even for the case with $N = 10$. Comparing the power of the tests, we observe that the panel LM test is not less powerful than the IPS test in all cases we consider and it is often more powerful than the IPS test. For example, when $N = 100$; $T = 25$; the power of LM is 0.518; while the power of IPS is 0.415. This result is not likely due to possible size distortions since the corresponding sizes are 0.060 and 0.063. When $N = 25$; $T = 50$; the power of LM is 0.802 and the power of IPS is 0.622, while the corresponding sizes are 0.058 and 0.059; respectively.

We also observe that power increases more rapidly with T than with N for both tests. For instance, when $T = 10$ the power is close to the size even with $N = 100$ in both tests. A similar pattern is reported for the IPS test in IPS (2003).

Experiment 2: Structural Shifts, No Serial Correlations

Next, we investigate the effect of structural shifts in level. We assume that a structural shift occurs at $T_{B,i}$ for the i-th time series with the magnitude of shift being δ_i. The data generation process follows:

$$y_{it} = z_{it} + x_{it}; \quad z_{it} = \delta_i D_{it}, \quad x_{it} = \rho_i x_{i,t-1} + \varepsilon_{it}. \tag{4.2}$$

We consider two cases. In the first case, all structural shifts are assumed to occur in the middle of the series so that $\lambda_i = 0.5$ for all i. In the second case, the structural shifts are assumed to occur at $\lambda_i = 0.3$ for all i. In both cases, we set $\sigma_i = 1$ and $\delta_i = 5$ for all i.

We experiment with three tests: the panel LM test without shifts using the statistic Γ_{LM} as defined in (2.14), the IPS test based on the statistic described in the previous subsection, and the panel LM test with level shifts using the statistic Γ_{LM}^{B} as defined in (3.13). We refer to them as LM_N, IPS_N and LM_B, respectively. Note that both LM_N and IPS_N ignore existing shifts, but LM_B controls for structural shifts. The results are reported in Table 3.

First, we examine the size and power properties of LM_N and IPS_N. We begin with the size of the tests under the null. Amsler and Lee (1995, Theorem 2) show that in a single time series both the LM and DF tests ignoring existing structural shifts are still valid under the null hypothesis, but this result holds only asymptotically. When the sample size is small, the usual tests ignoring structural shifts may result in a size somewhat different from the asymptotic size, depending on the values of λ and T. In the panel data framework, we may expect this divergence to be magnified in small samples. Our simulation results are the same as expected. We first look at the results when $\lambda = 0.5$. As we can see in the first panel of Table 3, the size distortion gets serious as N increases. For example, for the sample size $T = 25$, the empirical sizes of IPS_N are 0.017; 0.016, 0.007; 0.005 when $N = 10, 25, 50, 100$, and the corresponding figures of LM_N are 0.033, 0.021, 0.013, 0.006. Although the empirical sizes get closer to the nominal size as T increases, a similar pattern of size distortion is observed at a moderate sample size with $T = 100$. The size distortion gets worse for both IPS_N and LM_N when $\lambda = 0.3$; the results are given in the second panel of Table 3. Obviously, any small size distortion in individual time series accumulates in the panel framework as N increases. On the contrary, the size of LM_B is shown to be more or less correct in most cases. For instance, when $T = 25$, the empirical sizes of LM_B are 0.074; 0.054, 0.064; 0.058 when $N = 10, 25, 50, 100$. No serious size distortion is found for LM_B; clearly, it controls effectively the effect of structural changes. This result implies that we need to control for possible structural shifts in finite samples even under the null.

A more serious problem of ignoring existing structural shifts is the loss of power under the alternative hypothesis. Perron (1989) and Amsler and Lee (1995) report the same problem for the cases with a single time series. Our simulation results indicate that a similar pattern carries over to the panel unit root test. For example, as we can see in the first panel of Table 3, the power of IPS_N and LM_N is reported as 0.048 (size $= 0.005$) and 0.050 (size $= 0.006$), respectively, when $T = 25; N = 100$. The power of LM_B, however, is 0.473 (size $= 0.058$), exhibiting a decent power when allowing for a structural shift. The power of IPS_N and LM_N improves as T increases. Thus, the power loss is more serious when T is small. It appears that the low power of IPS_N and LM_N largely reflects the downward size distortion. But, the power of LM_N is still lower than that of LM_B even after controlling for the size. It is apparent that panel unit root tests lose power when they fail to control for structural shifts.

It is encouraging to see that the power of LM_B controlling for a level shift remains quite close to the corresponding figures (reported in Table 2) of the panel LM test without shifts. For example, when $T = N = 25$, the power of LM_B is 0.201 in the first panel (size $= 0.054$) when $\lambda = 0.5$, and 0.204 in the second panel (size $= 0.056$) when $\lambda = 0.3$. The corresponding figure in Table 2 is 0.207 (size $= 0.054$). This result has an important implication for practitioners. That is, there is almost no loss of power if one mistakenly adopts LM_B (instead of LM_N or IPS_N) if there are, in fact, no structural shifts in the time series. Also, the power of LM_B in both panels remains more or less the same for corresponding N and T cases, which implies that the power of LM_B is not affected by the value of λ.[3]

[3] When the DF regression includes a dummy variable to control for a structural break, the

Another interesting analysis is to examine how the size of LM_B varies when using different combinations of the N/T ratios. As we discussed in the previous section, the asymptotic validity of LM_B requires that N/T should not diverge as $N, T \to \infty$. It will be useful to examine how closely this asymptotic result holds in finite samples. It appears that this asymptotic result is well reflected in our simulation results using finite samples. Looking at the first panel in Table 3, for instance, the size of LM_B is reported as $0.083, 0.060, 0.062, 0.068$, respectively, when we vary the sample size $N = 10, 25, 50, 100$ for a fixed value of $T = 10$. These values appear stable. In addition, the sizes remain quite stable in other cases with different N/T ratios. As a result, it appears that the N/T ratio is not restrictive for applying LM_B under most practical situations.

Experiment 3: Structural Shifts, Serial Correlations

We now allow for both serial correlations and a structural shift in each time series. Two types of serial correlations are considered:

$$AR(1) \;:\; \varepsilon_{it} = 0.3\varepsilon_{i,t-1} + e_{it}, \tag{4.3}$$

$$MA(1) \;:\; \varepsilon_{it} = e_{it} - 0.3e_{i,t-1}, \tag{4.4}$$

where $e_{it} \sim iidN(0, 1)$. We maintain the same pattern of structural changes used in the second experiment in (4.2) with $\lambda_i = 0.5$ and $\delta_i = 5$ for all i. We simulate only the panel LM test based on the statistic $\Gamma_{LM}^{B}(p)$ in (3.21), where the number of augmented terms is fixed at $p_i = 0, 1, 2, 3, 4$, for all i in the regression (3.19).[4]

The results for the case of the $AR(1)$ error are reported in the first panel of Table 4. When p is under-specified ($p < 1$), we would expect size distortions. As expected, when $p = 0$ is used in the estimation procedure, all reported sizes are zero. The adverse bias effect of using a value of p smaller than necessary is more serious in the panel test since the bias accumulates as N grows. A similar pattern was reported for the IPS test in their paper. However, when the true value of p

asymptotic distribution of the resuting DF statistic depends on λ. Therefore the IPS test requires the expected values and variances of the ADF t-statisitcs for all different values of λ in the sample. This is quite an onerous job and was excluded from our simulation. However, we simulated the IPS-type statistic constructed from the DF regression including a break dummy variable, but using incorrectly the expected values and variances reported in IPS. The size is reported as 1.0 in many cases studied in our simulation.

[4] IPS (2003) observed a very serious size distortion when the lag orders are selected by Akaike or Schwarz criterion. They conjectured that the information criteria, as is well known, tend to choose too few lags when T is small, and this effect accumulates in the panel unit root test. A similar pattern is expected in the panle LM test. The data dependent method of selecting the lag order is an important issue, and it remains interesting to see the performance of the panel unit root test when the order of lags are chosen, for example, by the sequential method advocated by Ng and Perron (1995). But, in this paper we restrict ourselves to the study of fixed p for all i, leaving the data dependent selection of p to future research. By fixing p across individuals we could examine the effect of having too many or too few lagged terms.

is used at $p = 1$, the empirical sizes are reasonably close to the nominal 5% size for all cases using different values of N and T. Since the required condition of the N/T ratio is potentially important, we pay special attention to the changes in empirical size as we vary the N/T ratio at different values of p when the errors are serially correlated. For example, when $T = 25$ and $N = 10; 25; 50; 100;$ the empirical sizes are reported as $0.059; 0.064; 0.061; 0.058$, respectively. This result is viewed as evidence that the test remains valid unless N/T diverges as both N and $T \to \infty$. Similar results are obtained for other cases using $p = 2; 3$ or 4; no systematic pattern of size distortions is observed for any different combinations of N/T.

It is not surprising that the power of the test declines when the employed value of p is bigger than the true value of p in the DGP. Note that the true value of p in the DGP is $p = 1$. The power loss is more evident when T is relatively small. As an illustration, we compare the cases when $T = 25$ and $T = 50$. We use the same value $N = 100$ for both cases. We observe that the power of the test shrinks more rapidly when T is smaller. As we vary the values of p so that $p = 1; 2; 3; 4$, the powers of the test are given as 0.357; 0.283; 0.207; 0.165 when $T = 25$, but they are $0.993; 0.972; 0.916; 0.823$ when $T = 50$.

The results for the MA(1) errors are reported in the second panel of Table 4. Given MA errors, it is usually difficult to find a correct number of augmentation terms. All we can hope for is that the selected value of p is large enough to ensure that any left over errors in the regression (3.19) are nearly uncorrelated, so that the actual size of the test is reasonably close to the nominal size. At the same time, we also hope that the selected value of p is not too big so that the adverse effect on the power remains minimal. It is well known that the optimal value of p that serves this purpose depends on T as well as on the moving average parameter value. There is no asymptotic result available for the validity of any panel unit root tests with moving average errors. Despite the lack of theoretical underpinning, it is reasonable to expect that a bigger value of p is needed when T is larger. Also there is no reason to believe that the optimal value of p is associated with the size of N.

It is clear that under MA errors, using a smaller value of p leads to a size distortion. When using $p = 1$, the size distortion is apparent. It is also apparent that the optimal value of p depends on T. For instance, when $T = 25$; a value of p equal to 2 yields a size reasonably close to the nominal 5% level. However, when $T = 50$; the value of p must be 3 or larger. For the case when $p = 3$ or 4, the empirical size is tolerably close to the nominal size in all sample sizes we consider. A systematic effect of N on the size of the test is rather obvious when a smaller value of p is selected. But, the effect quickly becomes negligible when p is selected properly. For example, looking at the results with $T = 50$; we observe that the size of the test is quite stable if a reasonably large value of $p = 3$ is used; i.e., $0.052; 0.055; 0.049; 0.050$ when $N = 10; 25; 50; 100$. On the other hand, the size of the test tends to increase if a smaller value of p is used with $p = 2$;

i.e., $0.081; 0.098; 0.110; 0.138$ when the same values of N are used. Looking at the power of the test, we see that the power decreases when the chosen value of p is larger than necessary. For example, when $T = 25$ and $N = 50$; the power drops from 0.160 to 0.066 as p changes from 2 to 3. When $T = 50$ and $N = 10$, the power decreases from 0.356 to 0.233 as p changes from 3 to 4.

5. Empirical Application

As a practical application, we now apply our panel LM unit root test to analyze the empirical validity of long run purchasing power parity, PPP. This theory has been tested at length using univariate unit root tests, with largely disappointing results. Much of the failure of these tests to support the notion of PPP in real exchange rates has been attributed to the relatively low power of univariate unit root tests, especially given the relatively short time span of the post-Bretton Woods float.

However, recent developments in panel unit root tests, such as the 'Fisher test' proposed by Maddala and Wu (1998), the IPS test, and the test proposed in Choi (2001), have allowed for the possibility of increasing the power of the test for PPP. Recent examples of empirical tests for PPP using the above-mentioned panel unit root tests include: Coakley and Fuertes (1997), Wu and Wu (2001) and Choi (2001). The results of testing for PPP with these panel unit root tests has generally been more favorable than those in the univariate setting. Still, none of these applications allowed for the possibility of structural shifts in level in the exchange rate series under consideration, a factor which itself can lead to reduced power and spurious rejections of the null. As such, our application should provide a more accurate assessment of this issue.

To test for relative PPP, we form the natural log of the real exchange rate, y, as follows: $y = s + p - p^*$; where s is the natural log of the nominal exchange rate, expressed in units of foreign currency per U.S. dollar; p is the natural log of the U.S. price level, as measured by the consumer price index; and, p^* is the natural log of the foreign price level, also measured by the consumer price index. The null hypothesis of our test states that each country's real exchange rate contains a unit root. Failure to reject the null indicates the absence of a long run PPP relationship. Thus, rejection of the null provides evidence in favor of the PPP hypothesis.

Our test provides the following features. We allow for up to two structural shifts, as identified by the dummy variables D_{1it} and D_{2it}, as described in equation (3.19). Our panel LM test allows for country-specific fixed effects due to the fact that the t-statistics that are used to calculate the panel test statistic are estimated with heterogeneous intercept terms in the univariate LM tests. Finally, our test allows for heterogeneous measures of persistence among each of the countries of the panel. In this way, the rate of convergence to long run PPP can vary from country to country.

In order to test the robustness of the PPP hypothesis in the panel setting, we applied our test to the real exchange rates of four different panels of countries. The smallest panel follows the work of Choi (2001) and is comprised of the following six countries: Canada, France, Germany, Japan, Switzerland and the U.K. This group will be referred to as 'Choi's selection.' The second group is comprised of the following 12 countries from the European Monetary Union (EMU): Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal and Spain. The third group is comprised of the following 15 countries from the European Community (EC): Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden and the U.K. The fourth group includes the following 21 OECD countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the U.K. In all cases, both monthly and quarterly real exchange rates were constructed over the period from April 1973 through December 1999. The data used in this application was taken from the International Monetary Fund's International Financial Statistics CD ROM.

We begin with univariate LM tests, with no shift, one shift and two shifts. We determine jointly the number of augmentation terms and the optimal shift points. To so do, we first determine the number of augmentations at all shift combinations by following a 'general to specific' procedure; see Ng and Perron (1995). Then, the shift locations are determined via a grid search, at the minimum of the test statistics over the time interval [.1T, .9T] (to eliminate end points), by utilizing the optimal number of augmentation terms for all combinations of shift locations for each of the models with different numbers of structural shifts.

Next, we determine the optimal number of structural changes for each country. We adopt the approach of examining the significance of the dummy coefficients ΔD_{1it} and ΔD_{2it} in equation (3.19), using the usual t-test. The model-selection procedure is as follows. First, we estimate a two-shift model and determine whether or not such a model is appropriate by testing the significance of the dummy coefficients in this model. We select the two-shift model if both dummy coefficients are significant at the 5% level. If one of these coefficients is insignificant, we move to a one-shift model. If the dummy coefficient in this model is insignificant, we then select the model with no shift. It is important to note that the same expected value and variance of the panel LM test with no shift can be still used when T is big enough, even if the number of shifts as well as locations of shifts differ over different countries. This is due to the invariance property of the panel LM tests, as discussed in Section 3. It is necessary to assume that the locations of shifts are known in applying the panel unit root tests.

Table 5 contains the results of our panel LM unit root test applied to each of the four panels described above. In all four cases, several countries in the panel were found to have at least one structural shift (details of this analysis are available from the authors upon request). The results in Table 5 provide overwhelming support in favor of the PPP hypothesis for both monthly and quarterly exchange rates for all of the panels under consideration here. The null of a unit root is rejected at the 5% significance level in all cases. With regard to the EMU, EC and OECD panels, our results are consistent with those of Wu and Wu (1998).

With regard to the panel of Choi's selection of countries, however, we find evidence in favor of PPP where Choi (2001) was only able to find support for this hypothesis when Japan was excluded from the panel. Since Choi's test can be generalized to the case that allows for structural shifts, it would be interesting to see if his lack of evidence in favor of PPP when the panel includes Japan is attributable to the failure to allow for structural shifts in the series.

6. Concluding Remarks

In this paper we develop a new panel unit root test based on the LM principle. We apply this test in an empirical example and show strong support for long run PPP in four different panels of countries. Our proposed panel LM test is robust to the presence of heterogeneous structural shifts. This property bears important implications for empirical work, since no other panel unit root test has yet been developed with relevant asymptotics that can handle structural shifts in a practical way. In the benchmark case where no structural shifts are involved, the panel LM test is shown to be more powerful than the popular IPS test. Further, since the LM test is not subject to significant power loss when controlling for possibly spurious structural shifts when they do not exist, it will be a reasonable strategy to control for shifts even when their existence is questionable.

We have shown asymptotic results with the case of one structural shift in the mean in each time series, but extensions to the case with multiple shifts should be straightforward. When there is more than one shift in each time series, we have multiple dummy variables in the data generating process (3.1). The asymptotic results presented in this paper should also extend to the case of a finite number of multiple shifts.

Other extensions of the panel unit root test may also be of interest, but are left for future research. For example, one might be interested in a possible extension to the case of a trend shift. This case appears to pose practical difficulties, however, since the invariance property of the univariate LM test does not hold in the case of the panel LM test. In addition, recent developments in panel unit root tests have examined how to control the effect of the possible cross-sectional dependency. This issue is not yet considered in our paper, mainly due to the difficulty of deriving relevant asymptotic results.

A. Appendix

In this appendix we show that

$$L_T^B(\lambda) - L_T = O\left(T^{-1/2}\right), \tag{A.1}$$
$$E\left[L_T^B(\lambda) - L_T\right] = O\left(T^{-1}\right), \tag{A.2}$$

which we claimed in (3.17) and (3.18). We assume uniform integrability of $L_T^B(\lambda)$ and L_T. (See Remarks 1 and 3 in the text.) We drop subscript i for simplicity. The following lemmas will be used in the proof.

Lemma 1. Consider a regression:

$$y_t = x_t\beta + \delta D_t + \text{error}, \quad t = 1, 2, \ldots, T, \tag{A.3}$$

where

$$D_t = \begin{cases} 1 & \text{for } t = \tau, \\ 0 & \text{otherwise,} \end{cases}$$

and x_t is $1 \times k$ vector. Let (Y, X, D) be the $T \times (k+2)$ data matrix and (Y_*, X_*) be $(T-1) \times (k+1)$ matrix obtained after eliminating the τ-th observations (y_τ, x_τ) from (Y, X). Then,

$$X'M_{(D)}X = X_*'X_*, \quad X'M_{(D)}Y = X_*'Y_*,$$

where $M_{(\cdot)}$ is the projection onto the null space of $(\cdot)$. The first $\tau - 1$ and the last $T - \tau$ OLS residuals from the regression (A.3) is identical to the residuals obtained from regression of Y_* on X_* and the τ-th residual is zero.

Lemma 2. Let $\varepsilon_t \sim iidN(0, \sigma^2)$, for $t = 1, 2, \ldots, T$, and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)'$. Then

$$\sqrt{\frac{T}{\varepsilon'\varepsilon}} = \frac{1}{\sigma} + O_p\left(T^{-1/2}\right). \tag{A.4}$$

Proof: The result follows from a Nagar-type expansion: $\sqrt{\frac{T}{\varepsilon'\varepsilon}} = \sqrt{\frac{T}{E(\varepsilon'\varepsilon) + [\varepsilon'\varepsilon - E(\varepsilon'\varepsilon)]}}$. Since $\varepsilon'\varepsilon - E(\varepsilon'\varepsilon) = O_p\left(T^{1/2}\right)$, dividing both the numerator and the denominator by $\sqrt{T}$ yields the result.

>From Lemma 1, ν_2 (suppressing subscripts i) in regression (3.8) is the average of Δy_t after eliminating Δy_{T_B+1}, and $\mu = \Delta y_{T_B+1} - \nu_2$, namely:

$$\begin{aligned} \nu_2 &= \frac{\sum_{t=1}^{T} \Delta y_t - \Delta y_{T_B+1}}{T-1} = \frac{y_T - y_0}{T-1} - \frac{\Delta y_{T_B+1}}{T-1} \\ &= \frac{y_T - y_0}{T} + \frac{y_T - y_0}{T(T-1)} - \frac{\Delta y_{T_B+1}}{T-1}, \end{aligned} \tag{A.5}$$

$$\mu = \Delta y_{T_B+1} - \frac{y_T - y_0 - \Delta y_{T_B+1}}{T-1} = \frac{T\Delta y_{T_B+1} - (y_T - y_0)}{T-1}. \tag{A.6}$$

Therefore, if we let

$$q_{t-1} = \begin{cases} \left[\frac{y_T - y_0}{T(T-1)} - \frac{\Delta y_{T_B+1}}{T-1}\right](t-1) & \text{for } t \leq T_B; \\ \left[\frac{y_T - y_0}{T(T-1)} - \frac{\Delta y_{T_B+1}}{T-1}\right](t-1) + \frac{T\Delta y_{T_B+1} - (y_T - y_0)}{T-1} & \text{for } t \geq T_B + 1; \end{cases} \tag{A.7}$$

we have

$$S_{t-1} = \hat{S}_{t-1} - q_{t-1}. \tag{A.8}$$

Since $L_T^B(\lambda)$ is invariant numerically to γ_1, γ_2 and δ under the null hypothesis, we assume, without loss of generality, $\gamma_1 = \gamma_2 = \delta = 0$. Therefore, in the following, y_t follows a simple random walk with 0 initial value, and is denoted as

$$S_t = \sum_{j=1}^{t} \varepsilon_t; \tag{A.9}$$

where $\varepsilon_t \sim iidN(0, \sigma^2)$. We then have, under the null hypothesis,

$$\hat{S}_{t-1} = S_{t-1} - \frac{S_T}{T}(t-1); \tag{A.10}$$

$$\Delta y_t = \varepsilon_t; \tag{A.11}$$

and

$$q_{t-1} = \begin{cases} \left[\frac{S_T}{T(T-1)} - \frac{\varepsilon_{T_B+1}}{T-1}\right](t-1) & \text{for } t \leq T_B; \\ \left[\frac{S_T}{T(T-1)} - \frac{\varepsilon_{T_B+1}}{T-1}\right](t-1) + \frac{T\varepsilon_{T_B+1} - S_T}{T-1} & \text{for } t \geq T_B + 1. \end{cases} \tag{A.12}$$

>From (2.11) and (3.9), it is seen that

$$\begin{aligned} L_T &= \frac{\sqrt{T}\left(\hat{S}'_{-1} M_{(\iota_T)} \varepsilon\right)}{\sqrt{\left(\hat{S}'_{-1} M_{(\iota_T)} \hat{S}_{-1}\right)\left(\varepsilon' M_{(\iota_T, \hat{S}_{-1})} \varepsilon\right)}} + O_p\left(T^{-1}\right) \\ &= \frac{\sqrt{T}\left(\hat{S}'_{-1} M_{(\iota_T)} \varepsilon\right)}{\sqrt{\left(\hat{S}'_{-1} M_{(\iota_T)} \hat{S}_{-1}\right)(\varepsilon'\varepsilon)}} + O_p\left(T^{-1}\right); \end{aligned} \tag{A.13}$$

and

$$\begin{aligned} L_T^B(\lambda) &= \frac{\sqrt{T}\left(S'_{-1} M_{(\iota_T, \Delta D)} \varepsilon\right)}{\sqrt{\left(S'_{-1} M_{(\iota_T, \Delta D)} S_{-1}\right)\left(\varepsilon' M_{(\iota_T, \Delta D, S_{-1})} \varepsilon\right)}} + O_p\left(T^{-1}\right) \\ &= \frac{\sqrt{T}\left(S'_{-1} M_{(\iota_T, \Delta D)} \varepsilon\right)}{\sqrt{\left(S'_{-1} M_{(\iota_T, \Delta D)} S_{-1}\right)(\varepsilon'\varepsilon)}} + O_p\left(T^{-1}\right). \end{aligned} \tag{A.14}$$

Define

$$a_T = \hat{S}'_{-1} M_{(\iota_T)} \hat{S}_{-1} - \tilde{S}'_{-1} M_{(\iota_T;\Delta D)} \tilde{S}_{-1}; \; b_T = \hat{S}'_{-1} M_{(\iota_T)} \varepsilon - \tilde{S}'_{-1} M_{(\iota_T;\Delta D)} \varepsilon; \tag{A.15}$$

and

$$A_T = \frac{a_T}{\hat{S}'_{-1} M_{(\iota_T)} \hat{S}_{-1}}; \; B_T = \frac{b_T}{\sqrt{\left(\hat{S}'_{-1} M_{(\iota_T)} \hat{S}_{-1}\right)(\varepsilon'\varepsilon)}}. \tag{A.16}$$

Dividing both the numerator and the denominator of $L_T^B(\lambda)$ by $\sqrt{\left(\hat{S}'_{-1} M_{(\iota_T)} \hat{S}_{-1}\right)(\varepsilon'\varepsilon)}$; we have

$$\begin{aligned} L_T^B(\lambda) &= \frac{\left(\frac{\sqrt{T}\left(\hat{S}'_{-1} M_{(\iota_T)} \varepsilon\right)}{\sqrt{\left(\hat{S}'_{-1} M_{(\iota_T)} \hat{S}_{-1}\right)(\varepsilon'\varepsilon)}} - \sqrt{T} B_T \right)}{\sqrt{(1 - A_T)}} + O_p\left(T^{-1}\right) \\ &= L_T \sqrt{\left(1 + A_T + \frac{A_T^2}{1 - A_T}\right)} - \sqrt{T} B_T \sqrt{\left(1 + \frac{A_T}{1 - A_T}\right)} + O_p\left(T^{-1}\right). \end{aligned} \tag{A.17}$$

After some algebra (available from the authors upon request), we obtain

$$a_T = 2\varepsilon_{T_B+1} \left\{ -\frac{1}{T} \sum_{t=1}^{T} (t-1) S_{t-1} + \left(\frac{1}{2} + \lambda\right) \sum_{t=1}^{T} S_{t-1} + \left(\frac{1}{3} + \frac{1}{2}\lambda^2\right) T S_T \right\} + O_p(T); \tag{A.18}$$

and

$$b_T = \varepsilon_{T_B+1} \left[\frac{1}{2T} \sum_{t=1}^{T} t\varepsilon_t + \frac{1}{T} \sum_{t=1}^{T} S_{t-1} \right] + O_p(1); \tag{A.19}$$

where S_{t-1} is the partial sum process defined in (A.9), and $\lambda = \frac{T_B}{T}$. It now is obvious that $a_T = O_p\left(T^{3/2}\right)$; $b_T = O_p\left(T^{1/2}\right)$; and

$$A_T = O_p(T^{-1/2}) \text{ and } \sqrt{T} B_T = O_p(T^{-1/2}). \tag{A.20}$$

Therefore, it is seen from (A.17) that

$$L_T^B(\lambda) - L_T = \frac{1}{2} L_T A_T - \sqrt{T} B_T + O_p\left(T^{-1}\right); \tag{A.21}$$

which is $O_p(T^{-1/2})$.

Now we prove the claim in (A.2). From Assumption 3.1 of the uniform integrability of $L_T^B(\lambda)$ and L_T we have

$$E\left[L_T^B(\lambda) - L_T\right] = \frac{1}{2} E(L_T A_T) - \sqrt{T} E(B_T) + O\left(T^{-1}\right). \tag{A.22}$$

This result follows if we show $E(L_T A_T) = O(T^{-1})$ and $E(C_T) = O\left(T^{-3/2}\right)$. Combining (A.16) and (A.18) and applying Lemma 2, we have:

$$L_T A_T = \frac{2\varepsilon_{T_B+1}\hat{S}'_{-1} M_{(i_T)}\varepsilon\left[-\frac{1}{T}\sum_{t=1}^{T}(t-1)S_{t-1} + \left(\frac{1}{2}+\lambda\right)\sum_{t=1}^{T}S_{t-1} + \left(-\frac{1}{3}+\frac{1}{2}\lambda^2\right)TS_T\right]}{\sigma\left(\hat{S}'_{-1} M_{(i_T)}\hat{S}_{-1}\right)^{3/2}} + O_p\left(T^{-1}\right): \tag{A.23}$$

It is straightforward to see that

$$\hat{S}'_{-1} M_{(i_T)}\hat{S}_{-1} = \sum_{t=1}^{T} S_{t-1}^2 - \frac{2S_T}{T}\sum_{t=1}^{T}(t-1)S_{t-1} + \frac{T}{12}S_T^2 - \frac{1}{T}\left(\sum_{t=1}^{T} S_{t-1}\right)^2 + S_T\sum S_{t-1} + O_p(T) \tag{A.24}$$

Let $\hat{S}'_{-1} M_{(i_T)}\hat{S}_{-1} = V_{1T} + V_{2T}$; where V_{1T} is the sum of all the terms that include ε_{T_B+1}, and V_{2T} is the sum of the rest of the terms. Thus, no ε_{T_B+1} is associated with V_{2T}. Then, it is straightforward to see that $V_{1T} = O_p\left(T^{3/2}\right)$ and $V_{2T} = O_p(T^2)$. For example, the term associated with ε_{T_B+1} in the leading term, $\sum_{t=1}^{T} S_{t-1}^2$; is the squared terms of order T plus the sum of the mean zero cross product terms, the order of which is $O_p\left(T^{3/2}\right)$. The order of the other terms associated with ε_{T_B+1} is similarly obtained. From the normality, ε_{T_B+1} is independent of V_{2T}:

We define the numerator of (A.23) to be $2\varepsilon_{T_B+1}U_T$; and divide both the numerator and the denominator of (A.23) by $(V_{2T})^{3/2}$ to have

$$L_T A_T = \frac{2\varepsilon_{T_B+1}U_T/(V_{2T})^{3/2}}{\sigma(1+V_{1T}/V_{2T})^{3/2}} = \frac{2\varepsilon_{T_B+1}U_T}{(V_{2T})^{3/2}} + O_p\left(T^{-1}\right): \tag{A.25}$$

Now let $U_T = U_{1T} + U_{2T}$; where U_{1T} is the sum of the terms associated with ε_{T_B+1} and U_{2T} is the sum of the rest. Therefore, we have;

$$E(L_T A_T) = 2E\frac{\varepsilon_{T_B+1}U_{1T}}{(V_{2T})^{3/2}} + 2E\frac{\varepsilon_{T_B+1}U_{2T}}{(V_{2T})^{3/2}} + O\left(T^{-1}\right): \tag{A.26}$$

From the normality assumption, $U_{2T}/(V_{2T})^{3/2}$ is independent of ε_{T_B+1} so that the second term on the right hand side vanishes. Also, it is straightforward to show that $U_{1T} = O_p(T^2)$. We therefore have $\varepsilon_{T_B+1}U_{1T}/(V_{2T})^{3/2} = O_p(T^{-1})$ and $E\left[\varepsilon_{T_B+1}U_{1T}/(V_{2T})^{3/2}\right] = O(T^{-1})$:

$E(B_T) = O\left(T^{-3/2}\right)$ is proved similarly. From (A.16) and (A.19), we have:

$$B_T = \frac{\varepsilon_{T_B+1}\left(\frac{1}{2T}\sum_{t=1}^{T} t\varepsilon_t + \frac{1}{T}\sum_{t=1}^{T} S_{t-1}\right)}{\sqrt{\left(\hat{S}'_{-1}M_{(\iota_T)}\hat{S}_{-1}\right)\left(\varepsilon'\varepsilon\right)}} + O_p\left(T^{-3/2}\right)$$

$$= \frac{\varepsilon_{T_B+1}\left(\frac{1}{2}\sum_{t=1}^{T} t\varepsilon_t + \sum_{t=1}^{T} S_{t-1}\right)}{\sigma\sqrt{T^3\hat{S}'_{-1}M_{(\iota_T)}\hat{S}_{-1}}} + O_p\left(T^{-3/2}\right) \qquad \text{(A.27)}$$

$$= \frac{\varepsilon_{T_B+1}\left(\frac{1}{2}\sum_{t=1}^{T} t\varepsilon_t + \sum_{t=1}^{T} S_{t-1}\right)}{\sqrt{T^3V_{2T}}} + O_p\left(T^{-3/2}\right). \qquad \text{(A.28)}$$

The second equality follows from Lemma 2, and the third equality is obtained by dividing the numerator and the denominator by $\sqrt{T^3V_{2T}}$. Let

$$W_T = W_{1T} + W_{2T} = \frac{1}{2}\sum_{t=1} t\varepsilon_t + \sum_{t=1} S_{t-1}; \qquad \text{(A.29)}$$

where W_{1T} is the sum of the terms associated with ε_{T_B+1} and W_{2T} is the sum of the rest of the terms. From the independence of ε_{T_B+1} and $W_{2T}/\sqrt{T^3V_{2T}}$, we have

$$E(B_T) = E\frac{\varepsilon_{T_B+1}(W_{1T}+W_{2T})}{\sqrt{T^3V_{2T}}} + O\left(T^{-3/2}\right) = E\frac{\varepsilon_{T_B+1}W_{1T}}{\sqrt{T^3V_{2T}}} + O\left(T^{-3/2}\right). \qquad \text{(A.30)}$$

But, $\varepsilon_{T_B+1}W_{1T} = O_p(T)$ and $\varepsilon_{T_B+1}W_{1T}/\sqrt{T^3V_{2T}} = O_p\left(T^{-3/2}\right)$, so that we have $E(B_T) = O\left(T^{-3/2}\right)$, which completes the proof.

References

[1] Amsler, C. and J. Lee (1995), "An LM Test for a Unit-Root in the Presence of a Structural Change," Econometric Theory, 11, 359-368.

[2] Ahn, K. Sung (1993). "Some Tests for Unit Roots in Autoregressive-Integrated-Moving-Average Models with Deterministic Trends," Biometrika, 80, 855-868

[3] Baltagi, B.H. and C. Kao (2000), "Nonstationary Panels, Cointegrations in Panels and Dynamic Panels: A Survey," unpublished manuscript, Texas A&M University.

[4] Choi, I. (2001), "Unit Root Tests for Panel Data," Journal of International Money and Finance, 20 (2), 249-72

[5] Coakley, J. and A. Fuertes (1997), "New Panel Unit Root Tests of PPP," Economics Letters, 57 (1), 17-22.

[6] Im, K.S., M.H. Pesaran and Y. Shin (2003), "Testing for Unit Roots in Heterogeneous Panels," forthcoming, Journal of Econometrics.

[7] Maddala, G. S. and Wu, S (1999), "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," Oxford Bulletin of Economics and Statistics, 61, 631-52.

[8] Levin, A., C.F. Lin, and C. J. Chu (2002), "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties,"Journal of Econometrics, 108, 1-24.

[9] Ng, S. and P. Perron (1995), "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," Journal of the American Statistical Association, 90, 269-281.

[10] Perron, P. (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," Econometrica, 57, 1361-1401.

[11] Schmidt, P. and P.C.B. Phillips (1992), "LM Tests for a Unit Root in the Presence of Deterministic Trends. Oxford Bulletin of Economics and Statistics, 54, 257-287

[12] Wu, J. and S. Wu (2001), "Is Purchasing Power Parity Overvalued?" Journal of Money, Credit, and Banking, 33 (3), 804-812.

Addendum to:
"PANEL LM UNIT ROOT TESTS WITH LEVEL SHIFTS"
by Kyung So Im, Junsoo Lee and Margie Tieslau

In this note we show the algebra for deriving a_T and b_T in (A.18) and (A.19) of the Appendix. Note that

$$\hat{S}_{-1}' M_{(\iota_T)} \hat{S}_{-1} = \sum_{t=1}^{T} \hat{S}_{t-1}^2 - \frac{1}{T}\left(\sum_{t=1}^{T} \hat{S}_{t-1}\right)^2 . \tag{1}$$

We also have from Lemma 1

$$\begin{aligned}
\tilde{S}_{-1}' M_{(\iota_T;\cdot D)} \tilde{S}_{-1} &= \sum_{t=1}^{T} \tilde{S}_{t-1}^2 - \tilde{S}_{T_B}^2 - \frac{1}{T-1}\left(\sum_{t=1}^{T} \tilde{S}_{t-1} - \tilde{S}_{T_B}\right)^2 \\
&= \sum_{t=1}^{T} \hat{S}_{t-1}^2 - 2\sum_{t=1}^{T} \hat{S}_{t-1} q_{t-1} + \sum_{t=1}^{T} q_{t-1}^2 \qquad (2)\\
&\quad - \frac{1}{T-1}\left\{\left(\sum_{t=1}^{T} \hat{S}_{t-1}\right)^2 - 2\sum_{t=1}^{T} \hat{S}_{t-1} \sum_{t=1}^{T} q_{t-1} + \left(\sum_{t=1}^{T} q_{t-1}\right)^2\right. \\
&\quad \left. + 2\left(\hat{S}_{T_B} - q_{T_B}\right)\sum_{t=1}^{T}\left(\hat{S}_{t-1} - q_{t-1}\right) + T\left(\hat{S}_{T_B} - q_{T_B}\right)^2\right\} .
\end{aligned}$$

Inserting the results of (1) and (2) into a_T in (A.15) of the Appendix yields:

$$\begin{aligned}
a_T &= \hat{S}_{-1}' M_{(\iota_T)} \hat{S}_{-1} - \tilde{S}_{-1}' M_{(\iota_T;\cdot D)} \tilde{S}_{-1} \\
&= 2\sum_{t=1}^{T} \hat{S}_{t-1} q_{t-1} - \sum_{t=1}^{T} q_{t-1}^2 + \frac{1}{T(T-1)}\left(\sum_{t=1}^{T} \hat{S}_{t-1}\right)^2 \\
&\quad - \frac{1}{T-1}\left\{2\sum_{t=1}^{T} \hat{S}_{t-1} \sum_{t=1}^{T} q_{t-1} - \left(\sum_{t=1}^{T} q_{t-1}\right)^2 + 2\left(\hat{S}_{T_B} - q_{T_B}\right)\sum_{t=1}^{T}\left(\hat{S}_{t-1} - q_{t-1}\right)\right\} \\
&\quad - \frac{T}{T-1}\left(\hat{S}_{T_B} - q_{T_B}\right)^2 .
\end{aligned}$$

It is straightforward to see that $\sum_{t=1}^{T} q_{t-1}^2 = O_p(T)$; $\frac{1}{T(T-1)}\left(\sum_{t=1}^{T} \hat{S}_{t-1}\right)^2 = O_p(T)$; $\frac{1}{T-1}\left(\sum_{t=1}^{T} q_{t-1}\right)^2 = O_p(T)$; $\frac{1}{T-1}\left(\hat{S}_{T_B} - q_{T_B}\right)\sum_{t=1}^{T}\left(\hat{S}_{t-1} - q_{t-1}\right) = O_p(T)$;

and $\frac{T}{T-1}\left(\hat{S}_{T_B} - q_{T_B}\right)^2 = O_p(T)$. Therefore, we have

$$a_T = 2\left(\sum_{t=1}^{T}\hat{S}_{t-1}q_{t-1} - \frac{1}{T}\sum_{t=1}^{T}\hat{S}_{t-1}\sum_{t=1}^{T}q_{t-1}\right) + O_p(T)$$
$$= 2\varepsilon_{T_B+1}\left\{-\frac{1}{T}\sum_{t=1}^{T}(t-1)S_{t-1} + \left(\frac{1}{2}+\lambda\right)\sum_{t=1}^{T}S_{t-1} + \left(\frac{1}{3}+\frac{1}{2}\lambda^2\right)TS_T\right\} + O_p(T);$$

which is what we have in (A.18) in the Appendix. Following a visual inspection, the order of a_T is $O_p\left(T^{3/2}\right)$; so that $A_T = O_p\left(T^{-1/2}\right)$.

Now we derive the expression in (A.19). Note that

$$\tilde{S}'_{-1}M_{(i_T;\Delta D)}\varepsilon = \sum_{t=1}^{T}\tilde{S}_{t-1}\varepsilon_t - \tilde{S}_{T_B}\varepsilon_{T_B+1} - \frac{1}{T-1}\left(\sum_{t=1}^{T}\tilde{S}_{t-1} - \tilde{S}_{T_B}\right)\left(\sum_{t=1}^{T}\varepsilon_t - \varepsilon_{T_B+1}\right)$$
$$= \sum_{t=1}^{T}\tilde{S}_{t-1}\varepsilon_t - \frac{1}{T-1}\left(\sum_{t=1}^{T}\tilde{S}_{t-1}\sum_{t=1}^{T}\varepsilon_t - \tilde{S}_{T_B}\sum_{t=1}^{T}\varepsilon_t - \varepsilon_{T_B+1}\sum_{t=1}^{T}\tilde{S}_{t-1} + T\tilde{S}_{T_B}\varepsilon_{T_B+1}\right)$$
$$= \sum_{t=1}^{T}\hat{S}_{t-1}\varepsilon_t - \sum_{t=1}^{T}q_{t-1}\varepsilon_t$$
$$-\frac{1}{T-1}\left(\sum_{t=1}^{T}\hat{S}_{t-1}\sum_{t=1}^{T}\varepsilon_t - \sum_{t=1}^{T}q_{t-1}\sum_{t=1}^{T}\varepsilon_t - \tilde{S}_{T_B}\sum_{t=1}^{T}\varepsilon_t - \varepsilon_{T_B+1}\sum_{t=1}^{T}\tilde{S}_{t-1} + T\tilde{S}_{T_B}\varepsilon_{T_B+1}\right);$$

and

$$\hat{S}'_{-1}M_{(i_T)}\varepsilon = \sum_{t=1}^{T}\hat{S}_{t-1}\varepsilon_t - \frac{1}{T}\sum_{t=1}^{T}\hat{S}_{t-1}\sum_{t=1}^{T}\varepsilon_t.$$

Therefore,

$$b_T = \hat{S}'_{-1}M_{(i_T)}\varepsilon - \tilde{S}'_{-1}M_{(i_T;\Delta D)}\varepsilon$$
$$= \sum_{t=1}^{T}q_{t-1}\varepsilon_t - \frac{1}{T(T-1)}\sum_{t=1}^{T}\hat{S}_{t-1}\sum_{t=1}^{T}\varepsilon_t$$
$$+\frac{1}{T-1}\left(-\sum_{t=1}^{T}q_{t-1}\sum_{t=1}^{T}\varepsilon_t - \hat{S}_{T_B}\sum_{t=1}^{T}\varepsilon_t - \varepsilon_{T_B+1}\sum_{t=1}^{T}\hat{S}_{t-1} + T\hat{S}_{T_B}\varepsilon_{T_B+1}\right).$$

Examining each term, we have $\frac{1}{T(T-1)}\sum_{t=1}^{T}\hat{S}_{t-1}\sum_{t=1}^{T}\varepsilon_t = O_p(1)$; $\frac{1}{T-1}\tilde{S}_{T_B}\sum_{t=1}^{T}\varepsilon_t = O_p(1)$. Therefore,

$$\sum_{t=1}^{T}q_{t-1}\varepsilon_t = \left[\frac{S_T}{T(T-1)} - \frac{\varepsilon_{T_B+1}}{T-1}\right]\sum_{t=1}^{T}(t-1)\varepsilon_t + \frac{T\varepsilon_{T_B+1} - S_T}{T-1}\sum_{t=T_B+1}\varepsilon_t$$
$$= -\frac{1}{2T}\varepsilon_{T_B+1}\sum_{t=1}(t-1)\varepsilon_t + \varepsilon_{T_B+1}\sum_{t=T_B+1}\varepsilon_t + O_p(1);$$

$$
\begin{aligned}
\frac{1}{T-1}\sum_{t=1} q_{t-1}\sum_{t=1}\varepsilon_t &= \frac{1}{T-1}\left\{\frac{S_T}{T(T-1)} - \frac{\varepsilon_{T_B+1}}{T-1}\lambda\sum_{t=1}(t-1) + (1-\lambda)\frac{T\varepsilon_{T_B+1}-S_T}{T-1}\right\}\sum_{t=1}\varepsilon_t \\
&= \left(\frac{1}{2}-\lambda\right)\varepsilon_{T_B+1}\sum_{t=1}\varepsilon_t + O_p(1);
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{T-1}\varepsilon_{T_B+1}\sum_{t=1}\tilde{S}_{t-1} &= \frac{1}{T-1}\varepsilon_{T_B+1}\sum_{t=1}\hat{S}_{t-1} + O_p(1) \\
&= \frac{1}{T-1}\varepsilon_{T_B+1}\sum_{t=1}\left(S_{t-1} - \frac{S_T}{T}(t-1)\right) + O_p(1) \\
&= \frac{1}{T}\varepsilon_{T_B+1}\sum_{t=1}S_{t-1} - \frac{1}{2}\varepsilon_{T_B+1}S_T + O_p(1);
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{T}{T-1}\tilde{S}_{T_B}\varepsilon_{T_B+1} &= \hat{S}_{T_B}\varepsilon_{T_B+1} + O_p(1) \\
&= S_{T_B}\varepsilon_{T_B+1} + \lambda S_T\varepsilon_{T_B+1} + O_p(1).
\end{aligned}
$$

Combining these results, we obtain

$$
\begin{aligned}
b_T &= \frac{1}{2T}\varepsilon_{T_B+1}\sum_{t=1}(t-1)\varepsilon_t - \varepsilon_{T_B+1}\sum_{t=T_B+1}\varepsilon_t + \frac{1}{T}\varepsilon_{T_B+1}\sum_{t=1}S_{t-1} - S_{T_B}\varepsilon_{T_B+1} + O_p(1) \\
&= \varepsilon_{T_B+1}\left[\frac{1}{2T}\sum_{t=1}t\varepsilon_t + \frac{1}{T}\sum_{t=1}S_{t-1}\right] + O_p(1);
\end{aligned}
$$

which is the expression in (A.19) of the Appendix. It is obvious that $b_T = O_p\left(T^{1/2}\right)$. Hence, $\sqrt{T}C_T = O_p\left(T^{-1/2}\right)$.