Panel Cointegration Tests of the Fisher Hypothesis

Westerlund, Joakim; Edgerton, David

2005

Panel Cointegration Tests with Deterministic Trends and Structural Breaks

Joakim Westerlund[*] and David Edgerton

October 11, 2005

Abstract

This paper proposes Lagrange multiplier (LM) based tests for the null hypothesis of no cointegration in panel data. The tests are general enough to allow for heteroskedastic and serially correlated errors, individual specific time trends, and a single structural break in both the intercept and slope of each regression, which may be located different dates for different individuals. The limiting distributions of the test statistics are derived, and are found to be standard normal and free of nuisance parameters under the null. In particular, the distributions are found to be invariant not only with respect to trend and structural break, but also with respect to the presence of stochastic regressors. A small Monte Carlo study is also conducted to investigate the small-sample properties of the tests. The results reveal that the tests have small size distortions and good power even in very small samples.

JEL Classification: C12; C32; C33.

Keywords: Panel Cointegration; Residual-Based Cointegration Test; Structural Break; Deterministic Trend; LM Principle.

1 Introduction

Tests based on the residuals from a static regression are undoubtedly the most popular class of cointegration tests, in which the null hypothesis of no cointegration is tested against the alternative that the variables are cointegrated in the sense of Engle and Granger (1987). Unlike system based tests for cointegration rank, these residual-based tests rely on economic theory to provide the set of cointegrated variables. However, while economic theory may well imply cointegration as a long-run economic equilibrium relationship, they do not take into account many of the important features that characterizes the actual data generating process.

One such feature is the presence of deterministic time trends. Indeed, many economic variables, such as GDP, consumption and price levels, which

are commonly considered to be nonstationary, are actually best described as nonstationary with drift. With such variables, unless the cointegration vector annihilates both the stochastic and deterministic trends, the test regression should be fitted with a linear time trend as an additional explanatory variable. Yet, most cointegration testing has traditionally been, and continues to be, preformed in regressions without such trends. In fact, most researchers prefer to run their regressions with a constant term only, and use critical values that are appropriate in the absence of deterministic trends. Unfortunately, the presence of deterministic trends affects the asymptotic distribution of the test statistic whether or not the trend has in fact been included in the test regression. Thus, these critical values are not valid when at least one of the nonstationary variables exhibits deterministic trending behavior.

Another feature not accounted for by economic theory is structural breaks, such as the great depression, oil price shocks and policy regime shifts. Because no linear combination of nonstationary variables can be stationary in the presence of structural change, the long-run economic equilibrium may no longer imply cointegration in the sense of Engle and Granger (1987). In such cases, the appropriate test of economic theory may not be a conventional cointegration test, but rather a test for cointegration with structural break. However, it is well known that the presence of such breaks may induce serial correlation properties in the residuals that are akin to those of a random walk. Therefore, the conventional tests may incorrectly accept the null hypothesis of no cointegration when there is a break under the alternative hypothesis.

In this paper, we propose two simple tests for the null hypothesis of no cointegration that allow for deterministic trends and structural change in panel data. The tests are derived from the very popular Lagrange multiplier (LM) based unit root tests developed by Schmidt and Phillips (1992), Ahn (1993) and Amsler and Lee (1996), and they are able to accommodate heteroskedastic and serially correlated errors, individual specific time trends, and a single break in both the intercept and slope of each regression, which may be located different dates for different individuals.[1] Yet, both tests are very straightforward and easy to implement.

The new LM based test statistics are analyzed when there is no break, when the locations of the breaks are known *a priori* and when they are estimated from the data. In all three cases, we allow of for the presence of an individual specific time trend. The asymptotic analysis reveals that the statistics have limiting normal distributions that are free of nuisance parameters under the null hypothesis. In particular, it is shown that the asymptotic null distributions are independent of the nuisance parameters associated with both trend and structural break. Thus, with there tests, there is no need to compute different critical values for different deterministic specifications or break point patterns as in e.g. Perron (1989). Moreover, in contrast to existing cointegration tests, we find that the asymptotic null distributions of the LM based test statistics are independent of the presence of stochastic regressors. Another surprising property of the new statistics is that their asymptotic null

---

[1]Westerlund (2005a) also adopted an LM approach but with cointegration as the null hypothesis.

distributions are unaffected by erroneous omission of the structural breaks in the intercept but not in the slope.

We also evaluate the small-sample performance of the tests via Monte Carlo simulations. The results, which generally coincide with what might be expected from the asymptotic theory, can be summarized as follows. First, although the size of the tests is not affected, we find that erroneous omission of existing breaks can lead to a substantial loss of power. Second, treating the breaks as unknown rather than known has very little effect on the performance of the tests, which is a great operational advantage. Third, the tests generally perform well even in very small samples.

The paper proceeds as follows. Section 2 derives the LM based test statistics and their asymptotic distributions under the assumption of no breaks. Section 3 then allows for the presence of known breaks, while Section 4 extends the results to the case with unknown breaks. Section 5 is devoted to the Monte Carlo study. Section 6 concludes.

2   The LM based tests with no breaks

In this section, we derive the panel LM based test statistics for the simple case when there is no structural breaks. For this purpose, we make the assumption that the data generating process (DGP) can be described by the following unobserved components representation

$$
\begin{align}
y_{it} &= \alpha_i + \tau_i t + x'_{it}\beta_i + z_{it}, & (1)\\
z_{it} &= \rho_i z_{it-1} + e_{it}, & (2)\\
x_{it} &= x_{it-1} + w_{it}, & (3)
\end{align}
$$

where $i = 1, ..., T$ and $i = 1, ..., N$ indexes the time series and cross-sectional units, respectively. The vector x_{it} has dimension K and contains the regressors. For convenience in deriving the new tests and their asymptotic distributions, the initial conditions z_{i0} and x_{i0} are treated as fixed, and we make the following assumption regarding the error processes e_{it} and w_{it}.

Assumption 1. (Error process.) The processes e_{it} and w_{it} satisfy the following set of conditions:

(i) $E(e_{it}e_{kj}) = 0$ and $E(w_{it}w'_{kj}) = 0$ for all $i \neq k$, t and j.

(ii) $E(e_{it}w_{kj}) = 0$ for all i, k, t and j.

(iii)   The process e_{it} is normal, independent and identically distributed (i.i.d.) with $E(e_{it}) = 0$ and $E(e_{it}^2) = \sigma^2 > 0$ for all i.

(iv) $E(w_{it}w'_{it}) = \Omega_i$ is positive definite for all i and t.

Assumption 1 establishes the basic conditions needed for developing the new cointegration tests. Many of these are quite restrictive but are made here to simplify the derivation of the tests and their limiting distributions, and will be relaxed later on. Consider first Assumption 1 (i). This type of independence assumption is typical for our panel data approach and we will use it in the

3

derivation of the asymptotic distribution of our test statistics. In applied work, however, it may be useful to be able to allow for at least some kind of dependence among the cross-sectional units. A very straightforward and common way to do this is to assume that the dependence can be approximated by means of common time effects, which will be discussed later on.

Assumption 1 (ii) and (iv) relate to the covariance matrix of the regressors. Assumption 1 (ii) states that e_{it} and w_{it} are uncorrelated, which is equivalent to imposing weak exogeneity on the regressors. Although quite restrictive, as we shall see, our tests can be easily modified to accommodate for endogenous regressors. Assumption 1 (iv) states that the $K \times K$ covariance matrix Ω_i is positive definite, which is tantamount to requiring that x_{it} is not cointegrated in case we have multiple regressors. This assumption is very standard in the related literature and will therefore be maintained throughout this study.

Assumption 1 (iii) requires that the innovations e_{it} are i.i.d. and normal, which is very standard for the likelihood based framework that we consider in this paper. It is also very convenient because it means that standard functional central limit arguments can be applied to each cross-sectional unit, which makes the asymptotic analysis relatively uncomplicated. A few different possibilities concerning how to modify the tests to allow for more general error dynamics will be discussed later.

Having laid out the key assumptions that characterize the DGP considered in this paper, we now proceed by deriving the test statistics. To this end, consider testing the null of no cointegration using equations (1) to (3). The error z_{it} is stationary when x_{it} and y_{it} are cointegrated and it has a unit root if they are not. Thus, the null hypothesis that all the cross-sectional units are not cointegrated can be stated equivalently as

$$H_0 : \rho_i = 1 \text{ for all } i = 1, ..., N,$$

versus

$$H_1 : |\rho_i| < 1 \text{ for } i = 1, ..., N_1 \text{ and } \rho_i = 1 \text{ for } i = N_1 + 1, ..., N.$$

This formulation of the alternative hypothesis allows ρ_i to differ across the cross-sectional units, and is more general than the common homogenous alternative hypothesis that $\rho_i = \rho$ and $|\rho| < 1$ for all i. For consistency of the tests, it is necessary to assume that the fraction of cointegrated units is nonvanishing. Thus, we require that $N_1/N \longrightarrow \delta \in (0, 1]$ as $N \longrightarrow \infty$.

The restriction that $\rho_i = 1$ for all i can be tested using the LM, or score, principle that the score vector, conditional on past information, has zero mean when evaluated at the vector of true parameters under the null hypothesis. Consider therefore the following pooled log-likelihood function

$$\log L = -\frac{NT}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{N} SSE_i, \tag{4}$$

where the functional expression for SSE_i, the sum of squared errors, can be found in the appendix. The LM based tests for the null hypothesis of $\rho_i = 1$ can be derived by first concentrating the log-likelihood function with respect to

4

σ^2 and then evaluating the resultiong score vector at the restricted maximum likelihood estimates. As shown in Proposition A.1 of the appendix, this score vector is given by

$$\frac{\partial \log L}{\partial \rho_i} = \widehat{\sigma}^{-2} \sum_{t=2}^{T} (\Delta \widehat{S}_{it} - \Delta \widehat{S}_{i.})(\widehat{S}_{it-1} - \widehat{S}_{i.-1}), \tag{5}$$

where $\widehat{\sigma}^2 = (TN)^{-1} \sum_{i=1}^{N} SSE_i$ is the maximum likelihood estimate of σ^2. The variables $\Delta \widehat{S}_{i.}$ and $\widehat{S}_{i.-1}$ are the sample means of $\Delta \widehat{S}_{it}$ and $\widehat{S}_{it-1}$, respectively, where $\widehat{S}_{it}$ can be written as $\widehat{S}_{it} = y_{it} - \widehat{\alpha}_i - \widehat{\tau}_i t - x_{it}' \widehat{\beta}_i$ with $\widehat{\alpha}_i = y_{i1} - \widehat{\tau}_i - x_{i1}' \widehat{\beta}_i$ being the restricted maximum likelihood estimate of $\widetilde{\alpha}_i = \alpha_i + z_{i0}$. The corresponding estimates $\widehat{\tau}_i$ and $\widehat{\beta}_i$ of τ_i and β_i, respectively, are obtained by running the following least squares regression

$$\Delta y_{it} = \widehat{\tau}_i + \Delta x_{it}' \widehat{\beta}_i + \widehat{e}_{it}.$$

Equation (5) shows that the score vector is proportional to the numerator of the least squares estimate of the autoregressive parameter $\phi_i = \rho_i - 1$ in the following auxiliary regression

$$\Delta \widehat{S}_{it} = \mu_i + \phi_i \widehat{S}_{it-1} + u_{it}, \tag{6}$$

where μ_i is an individual specific constant and u_{it} is a stationary error term. It follows that a test of the hypothesis of $\rho_i = 1$ versus $|\rho_i| < 1$ for cross-section unit i can be formulated equivalently as a test of the hypothesis of $\phi_i = 0$ against $\phi_i < 0$, which can be tested using either the least squares estimate of ϕ_i or its t-ratio. Thus, by looking at (4), it is not difficult to see that a panel test of H_0 versus H_1 can be constructed by using the cross-sectional sum of these statistics for each i. The precise form of the panel statistics considered in this paper is given as follows.

Definition 1. (The LM based statistics with no breaks.) Let $\widetilde{S}_{it} = (\widehat{S}_{it-1} - \widehat{S}_{i.-1}, \Delta \widehat{S}_{it} - \Delta \widehat{S}_{i.})'$ and $\widetilde{S}_i = \sum_{t=2}^{T} \widetilde{S}_{it} \widetilde{S}_{it}'$, then the panel LM based test statistics are defined in the following fashion

$$Z_\phi \equiv \sum_{i=1}^{N} \widetilde{S}_{i11}^{-1} \widetilde{S}_{i12} \quad \text{and} \quad Z_t \equiv \widehat{\sigma}^{-1} \sum_{i=1}^{N} \widetilde{S}_{i11}^{-1/2} \widetilde{S}_{i12}.$$

Remark 1. In this paper, we are not concerned with the LM test itself, but rather with LM based tests. The LM test statistic of H_0 versus H_1 under Assumption 1 is derived in Proposition A.1 of the appendix. It is shown that it takes the form of a sum of N terms corresponding to the square of a t-ratio of the hypothesis of a zero slope in a regression of $\Delta \widehat{S}_{it}$ on $\widehat{S}_{it-1}$ for each i. The proposed statistics are computationally more convenient and, as pointed out by Ahn (1993), asymptotically as efficient as this statistic.

Remark 2. Analogous to the unit root case studied by Schmidt and Phillips (1992), we see that the parameters used to compute $\widehat{S}_{it}$ are estimated from the first differentiated data. By contrast, the Dickey and Fuller (1979) versions of Z_ϕ and Z_t studied by e.g. Banerjee and Carrion-i-Silvestre (2005), Gutierrez

(2004) and Westerlund (2005c) are based on estimating the parameters using the data in levels. Thus, since the variables are nonstationary, a regression of y_{it} on x_{it}, intercept and trend in levels is spurious, so that the estimated regression parameters do not converge to constants, but remain random even asymptotically. As we shall see, this lower degree of randomness makes the asymptotic properties of Z_ϕ and Z_t relatively simple.

Remark 3. Since Δy_{it} is a stationary process with intercept τ_i and slope β_i under the null hypothesis, the restricted maximum likelihood estimators of these parameters can easily be obtained using least squares as explained earlier, and are in fact consistent at rate $T^{1/2}$. However, while identifiable under the alternative, the parameters α_i and z_{i0} cannot be identified separately under the null. Therefore, we estimate $\widetilde{\alpha}_i$ rather than α_i, which entails no loss of generality in large samples. Moreover, as shown in the appendix, since the population intercept is zero, equation (6) could be specified without the constant term. However, as pointed out by Ahn (1993), because α_i and z_{i0} are not separately identifiable, $\widehat{\alpha}_i$ may be biased in small samples, which implies that $\widehat{S}_{it}$ may be biased too. To ease this problem, we fit (6) with a constant. This specification also performed best in the simulations, which corroborates the finding of Schmidt and Lee (1991) that the exclusion of the constant term makes the tests dependent on z_{i0}.

Remark 4. Relaxing Assumption 1 (iii) mean that e_{it} could be both heteroskedastic and serially correlated as well as nonnormal. This implies the LM based statistics are no longer asymptotically similar and that they need to be modified to account for the temporal dependence in the DGP. To facilitate this, we must replace Assumption 1 (iii) with something else. Towards this end, we make the assumption that the linear process conditions of Phillips and Solo (1992) are satisfied, which allows for a very general class of errors including all stationary autoregressive moving average processes. Under these conditions, similar statistics may be obtained using either the semiparametric correction of Schmidt and Phillips (1992) or the parametric correction of Ahn (1993) and Amsler and Lee (1996). The former involves multiplying Z_ϕ by $\widehat{\sigma}_i^2 \widehat{\omega}_i^{-2}$ and Z_t by $\widehat{\sigma}_i \widehat{\omega}_i^{-1}$, where $\widehat{\sigma}_i^2$ and $\widehat{\omega}_i^2$ are consistent estimates of the contemporaneous and long-run variances of u_{it}, respectively. The latter involves augmenting the right-hand side of (6) with, say p, lagged values of $\Delta \widehat{S}_{it}$, where p should be sufficiently large to whiten the remaining error. Our Monte Carlo results suggest that the parametric correction method works best and it will therefore be applied in this paper.

Remark 5. Relaxing Assumption 1 (ii) is equivalent to allowing for endogenous regressors. As with the relaxation of Assumption 1 (iii), this implies that the LM based statistics are no longer similar with respect to the DGP unless they can be appropriately modified to account for the endogeneity of the regressors. One simple way to accomplish this is by estimating β_i using instrumental variables rather than least squares. In this case, the instruments should be uncorrelated with e_{it} but highly correlated with Δx_{it}. Obvious candidates are lagged values of Δx_{it}.

Remark 6. Assumption 1 (i) requires that the cross-sectional units are in-

dependent of each other. Although the potential effects of the breakdown of this assumption have been stressed by e.g. Banerjee *et al.* (2004), the accommodation of such dependence has yet to become standard in the panel cointegration literature. One simple solution is to use data that has been demeaned with respect to common time effects, in which case the statistics are calculated as before but with $\widetilde{x}_{it} = x_{it} - N^{-1}\sum_{i=1}^{N} x_{it}$ in place of x_{it} and $\widetilde{y}_{it} = y_{it} - N^{-1}\sum_{i=1}^{N} y_{it}$ in place of y_{it}. This approach not only preserves the simple structure of our test statistics but also leaves their asymptotic distributions unaltered, which is very convenient. Another advantage of this approach is that subtracting the cross-sectional average may be quite effective even against very general forms of cross-sectional correlation structures (see, e.g. Westerlund, 2005b).

The asymptotic distributions of the test statistics are derived next. For this purpose, sequential limit theory is employed, which is a very convenient method for obtaining the limiting distribution of a double indexed sequence. In particular, the following theorem shows that both Z_ϕ and Z_t have limiting normal distributions when standardized by the first two moments of the following Brownian motion functionals

$$ U_i \equiv -\left(4\int_0^1 \overline{V}_i(r)^2 \mathrm{d}r\right)^{-1/2} \text{ and } K_i \equiv -\left(2\int_0^1 \overline{V}_i(r)^2 \mathrm{d}r\right)^{-1}, $$

where $W_i(r)$ is a standard Brownian motion defined on the unit interval $r \in [0,1]$, $V_i(r) = W_i(r) - rW_i(1)$ is a standard Brownian bridge and $\overline{V}_i(r) = V_i(r) - \int_0^1 V_i(s)\mathrm{d}s$ is a demeaned standard Brownian bridge. Let $\Rightarrow$ signify weak convergence, and let Θ and Σ denote the expected value and the covariance of the vector $(U_i, K_i)'$, respectively. The following result now holds.

Theorem 1. (Asymptotic distribution with no breaks.) Under H_0 and Assumption 1, as $T \longrightarrow \infty$ prior to $N \longrightarrow \infty$

$$ \begin{aligned} N^{-1/2}Z_t - N^{1/2}\Theta_1 &\Rightarrow N(0, \Sigma_{11}), \\ TN^{-1/2}Z_\phi - N^{1/2}\Theta_2 &\Rightarrow N(0, \Sigma_{22}). \end{aligned} $$

Remark 7. The proof of Theorem 1 is given in the appendix but it is constructive to consider briefly why it holds. For this purpose, let $e_{i.}$ denote the time series average of e_{it}. The proof begins by showing that $T^{-1/2}\widehat{S}_{it}$ is equal to $T^{-1/2}\sum_{k=2}^{t}(e_{ik} - e_{i.})$ plus a term that is asymptotically negligible suggesting that $T^{-1/2}\widehat{S}_{it}$ converges in distribution to σ times a standard Brownian bridge as $T \longrightarrow \infty$ for a fixed N. This implies that $T^{-2}\sum_{t=2}^{T}(\widehat{S}_{it-1} - \widehat{S}_{i.-1})^2$ converges to σ^2 times the square of an integral demeaned Brownian bridge. Also, because $T^{-1}\sum_{t=2}^{T}(\widehat{S}_{it-1} - \widehat{S}_{i.-1})(\Delta\widehat{S}_{it} - \Delta\widehat{S}_{i.})$ converges to $-\sigma^2/2$, we obtain the required intermediate limit distributions described by K_i and U_i. Therefore, since these functionals are i.i.d. across i, the asymptotic results follow directly by standard Lindberg-Lévy central limit arguments as T and N grow large in sequence.

Remark 8. Since the limit of $T^{-1/2}\widehat{S}_{it}$ does not depend on x_{it}, this means that the panel LM based statistics have the surprising property that their limiting

distributions are unaffected by the presence of stochastic regressors. Hence, the same set of moments can be applied regardless of the number of regressors, which make the statistics computationally very convenient. On the other hand, the limit of $T^{-1/2}\widehat{S}_{it}$ does depend on the estimation of the trend parameter τ_i, which is seen by the fact that the distributions of the tests involve the Brownian bridge $V_i(r)$ instead of the Brownian motion $W_i(r)$. Therefore, if it is known that $\tau_i = 0$ for all i, then the moments of the LM based test statistics should be based on the demeaned Brownian motion $W_i(r) - \int_0^1 W_i(s)\mathrm{d}s$ rather than the demeaned Brownian bridge $\overline{V}_i(r)$. However, if it is not known whether $\tau_i = 0$ for all i or not, the model with a trend is recommended. In addition, preliminary simulation results indicate that the tests with the trend included performs best in small samples.

The moments of the test are obtained by simulation methods. That is, the elements of the Brownian motion functionals K_i and U_i are approximated by functions of sums of partial sums of T independent draws from the standard normal distribution. Specifically, we generate $100,000$ samples of length $T = 1,000$, which are then used to obtain numerical values of the moments in Theorem 1. The simulated values of Θ_1 and Σ_{11} using this method are -1.9675 and 0.3301, respectively. The corresponding values of Θ_2 and Σ_{22} are -8.4376 and 25.8964, respectively. Because the normalized statistics diverge towards negative infinity under the alternative hypothesis, the computed value of the test statistic should be compared with the left tail of the normal distribution. If the computed value is greater than the appropriate left tail critical value, we reject the null hypothesis.

Moreover, although this paper is concerned with panel data, it is obvious that our approach also applies to the pure time series case, in which Z_ϕ and Z_t reduce to their corresponding time series quantities for each individual. The asymptotic distributions for these statistics are easily deducible from the appendix, where it is shown that $Z_t \Rightarrow U_i$ and $Z_\phi \Rightarrow K_i$ as $T \longrightarrow \infty$. The appropriate critical values for these distributions are tabulated by Schmidt and Phillips (1992).

3 The LM based tests with known breaks

We now extend the tests derived in the previous section to the more general case when there is a structural break in both the intercept and slope of each individual regression. Many of the results given here follow directly from the same kind of arguments used in Section 2 and hence only essential details will be given.

We begin with the level shift case, in which the slope parameter β_i is not subject to structural change. The DGP can be characterized by equations (2) and (3) but with equation (1) replaced with the following expression

$$y_{it} = \alpha_i + \tau_i t + \delta_i D_{it} + x'_{it}\beta_i + z_{it}. \tag{7}$$

In this parametrization of the DGP, α_i represents the intercept before the break and δ_i represent the change in the intercept at the time of the shift. If

we let T_i denote the location of this shift for cross-section unit i, then $D_{it} = 1$ if $t > T_i$ and zero otherwise. The location of the break is assumed to be a fixed fraction $\lambda_i \in (0,1)$ of T such that $T_i = \lambda_i T$ for all i, where the limit is taken as $T \longrightarrow \infty$ in a sequence that ensures an integer value of T_i. For now, it is assumed that the location of each break is known, which simplifies the presentation. The case with unknown breaks will be dealt with in Section 4.

As in the case when there is no break, the LM based statistics can be obtained on the basis of the following auxiliary regression

$$\Delta \widehat{S}_{it}^c = \mu_i + \phi_i \widehat{S}_{it-1}^c + u_{it}. \tag{8}$$

In this case, however, in order to account for the level break, the variable $\widehat{S}_{it}^c$ is constructed as $\widehat{S}_{it}^c = y_{it} - \widehat{\alpha}_i - \widehat{\tau}_i t - \widehat{\delta}_i D_{it} - x_{it}' \widehat{\beta}_i$, where the new restricted maximum likelihood estimate of $\widetilde{\alpha}_i$ is given by $\widehat{\alpha}_i = y_{i1} - \widehat{\tau}_i - \widehat{\delta}_i D_{i1} - x_{i1}' \widehat{\beta}_i$. Analogous to the no break case, the restricted maximum likelihood estimates $\widehat{\tau}_i$, $\widehat{\delta}_i$ and $\widehat{\beta}_i$ of τ_i, δ_i and β_i, respectively, can be obtained by using the following least squares regression

$$\Delta y_{it} = \widehat{\tau}_i + \widehat{\delta}_i \Delta D_{it} + \Delta x_{it}' \widehat{\beta}_i + \widehat{e}_{it}.$$

As before, a test of H_0 versus H_1 can be based on the least squares estimate of ϕ_i in (8) or its t-ratio. The test statistics in the level break case can thus be constructed exactly as in Definition 1 with $\widehat{S}_{it}^c$ in place of $\widehat{S}_{it}$. The resulting modified versions of Z_ϕ and Z_t will henceforth be denoted Z_ϕ^c and Z_t^c, respectively. If follows that the LM based statistics in the case with level breaks are computationally as convenient as in the case with no breaks. In fact, the only novelty here is the break dummy D_{it}, which is needed in order to obtain tests that are unaffected by the break under the alternative hypothesis of cointegration. The asymptotic results for the level break model is given next.

Theorem 2. (Asymptotic distribution with level shifts.) Under H_0 and Assumption 1, as $T \longrightarrow \infty$ prior to $N \longrightarrow \infty$, then the following results hold:

(i) The modified statistics Z_ϕ^c and Z_t^c are distributed as Z_ϕ and Z_t in Theorem 1, respectively.

(ii) The asymptotic distributions of the unmodified statistics Z_ϕ and Z_t are unaltered by the presence of a level break.

Remark 9. Theorem 2 (i) states that allowing for a single known structural break in the intercept of each individual does not affect the asymptotic distributions of the modified LM based statistics under the null hypothesis of no cointegration. This is true whether there are breaks or not. Hence, the asymptotic distributions of the statistics are unaffected by allowing for breaks even when there are no breaks. This is, of cause, very convenient as it means that we can use the same moments as derived in Section 2, and proceed with the modified statistics as if there were no breaks at all. Thus, there is no need to tabulate different critical values for different break structures as in e.g. Perron (1989).

9

Remark 10. Analogous with the unit root case studied by Amsler and Lee (1995), Theorem 2 (ii) shows that the asymptotic null distributions of the LM based statistics computed under the assumption of no break are unaffected by the presence of structural breaks. Hence, the null distributions of these statistics do not depend on breaks that are misplaced or even excluded in the estimation. The intuition behind this result follows from the fact that the LM based statistics are obtained using a regression in first differences, and ΔD_{it} equals one only at one point, so its inclusion has no effect asymptotically. The problem is that incorrect placement, or exclusion, of the breaks make the tests biased towards accepting the null hypothesis. Thus, although the presence of breaks does not affect the null distribution, it does affect the tests by reducing their power, which is why accounting for breaks is important.

Remark 11. Since the breaks are allowed under both the null and alternative hypotheses, there is no confusion about the interpretation of the test outcome. As an example, consider the univariate unit root test of Zivot and Andrews (1992), which allows for a single unknown break to affect the level and trend of the series. The problem with this test is that the break is only permitted under the alternative hypothesis of stationarity. Thus, a rejection of the null does not necessarily imply a rejection of a unit root *per se* but rather a rejection of a unit root without breaks. This outcome calls for a careful interpretation of the test result in applied work. Particularly, in the presence of breaks under the null, researchers might incorrectly conclude that a rejection of the null indicates evidence of stationarity with a break, when in fact the series nonstationary with breaks.

Next, we extend the results of the level break model to the case when there is a structural break in both the intercept and slope of each individual regression. As before, we assume that (2) and (3) holds but that y_{it} evolves according to the following equation

$$y_{it} = \alpha_i + \tau_i t + \delta_i D_{it} + D_{it} x'_{it} \gamma_i + x'_{it} \beta_i + z_{it}. \tag{9}$$

In this case, the parameters α_i, τ_i and δ_i are as in the level shift model, β_i is the slope parameter before the shift and γ_i represents the change in the slope. Thus, the relation in (9) is permitted to rotate as well as shift parallel. Following the terminology introduced by Gregory and Hansen (1996), we shall refer to this DGP as the regime shift model. As before, the LM based statistics can be derived from the of the following auxiliary regression

$$\Delta \widehat{S}^r_{it} = \mu_i + \phi_i \widehat{S}^r_{it-1} + u_{it}. \tag{10}$$

Analogous to the level shift case, the variable $\widehat{S}^r_{it}$ may be written as $\widehat{S}^r_{it} = y_{it} - \widehat{\alpha}_i - \widehat{\tau}_i t - \widehat{\delta}_i D_{it} - D_{it} x'_{it} \widehat{\gamma}_i - x'_{it} \widehat{\beta}_i$, where we use $\widehat{\alpha}_i = y_{i1} - \widehat{\tau}_i - \widehat{\delta}_i D_{i1} - D_{i1} x'_{i1} \widehat{\gamma}_i - x'_{i1} \widehat{\beta}_i$ to denote the new restricted maximum likelihood estimate of $\widetilde{\alpha}_i$. As in the previous section, the restricted maximum likelihood estimates of τ_i, δ_i, γ_i and β_i can be obtained from the following least squares regression

$$\Delta y_{it} = \widehat{\tau}_i + \widehat{\delta}_i \Delta D_{it} + \Delta (D_{it} x_{it})' \widehat{\gamma}_i + \Delta x'_{it} \widehat{\beta}_i + \widehat{e}_{it}.$$

As expected, the LM based statistics in the regime shift case can be constructed as in Definition 1 with $\widehat{S}^r_{it}$ in place of $\widehat{S}_{it}$. The modified versions of Z_ϕ and Z_t will henceforth be written in an obvious notation as Z^r_ϕ and Z^r_t, respectively. Their asymptotic distributions are given in the following theorem.

Theorem 3. (Asymptotic distribution with regime shifts.) Under H_0 and Assumption 1, as $T \longrightarrow \infty$ prior to $N \longrightarrow \infty$, the modified statistics Z^r_ϕ and Z^r_t are distributed as Z_ϕ and Z_t in Theorem 1, respectively.

Remark 12. Analogous with the level break case, Theorem 3 shows that the asymptotic distributions of the modified statistics Z^r_ϕ and Z^r_t are unaffected by the presence of regime shifts. Thus, in this sense, the modified LM based statistics derived from the level and regime shift models are very similar. One important difference is that the unmodified statistics in the regime shift model are not asymptotically invariant with respect to the breaks as in the level shift model. Thus, it is no longer possible to disregard or misspecify the structural breaks without affecting the asymptotic null distributions of the test statistics. The intuition behind this result is that, although ΔD_{it} vanishes asymptotically when normalized by $T^{-1/2}$, the first difference of $x_{it} D_{it}$ does not. Thus, unless the breaks have been correctly accounted for, the distributions of the statistics will depend on both the regressors and the locations of the structural breaks.

4   The LM based tests with unknown breaks

In the previous section, we assumed that the locations of the structural breaks are known. This is not necessary. In fact, if there is no *a priori* knowledge about the breakpoints, one can treat them as endogenous variables that can be estimated from the data.

Towards this end, note that, because the LM based test statistics are asymptotically similar with respect to the breaks, their null distributions will be unaffected by dispensing with the assumption of known breaks, which is likely to be unduly restrictive for most empirical purposes. In fact, the breaks could be misplaced or even ignored without affecting the distributions of the test statistics. Hence, the properties of the tests under the null remain unaltered even if we employ an inconsistent estimator of the breakpoints. The problem here is that incorrect placement of the breaks makes the tests biased towards accepting the null hypothesis. Thus, employing a poor estimator is expected to result in a loss of power.

Arguably, the single most popular cointegration testing procedure with unknown breaks is that employed by e.g. Gregory and Hansen (1996) and Westerlund (2005c), in which the location of a single break can be estimated via grid search at the minimum of the individual test statistics. However, our simulation results suggest that the size of our tests based on this procedure can be highly unreliable with serious distortions in many cases. Another drawback of this procedure is that it cannot be easily extended to allow for more than one break. Indeed, extending this grid search beyond the single break case

would be computationally extremely demanding and practically infeasible for more than three breaks, say.

An alternative approach, that is computationally very convenient and readily extendable to the case with multiple breaks, is that developed by Bai (1994, 1997) and Bai and Perron (1998), in which the break locations are estimated by minimizing the sum of squared residuals among all possible sample splits. Let $\widehat{u}_{it}$ denote the least squares residual from (8) or (10), then the break point estimator is defined as follows

$$\widehat{T}_i = \operatorname*{argmin}_{T_i \in [n, T-n]} \sum_{t=1}^{T} \widehat{u}_{it}^2,$$

where n is a trimming parameter that imposes a minimum length for each subsample. Consider the level break model, in which δ_i denotes the magnitude of the break for each i. Bai (1994) shows that $\widehat{T}_i - T_i$ is $O_p(\delta_i^{-2})$, which implies that $\widehat{\lambda}_i - \lambda_i$ must be $O_p(T^{-1}\delta_i^{-2})$. Thus, even though $\widehat{\lambda}_i$ is consistent for λ_i as T grows, the break date estimator is by itself not consistent for the break date. Fortunately, consistent estimation of the model under both the null and alternative hypotheses requires only that the break fraction λ_i can be estimated consistently. Thus, this procedure is expected to generate tests with good power. Note also that the accuracy of the estimated breakpoints depend on the parameter δ_i, which governs the size of the break. This is to be expected since a smaller break is more difficult to discern. The same argument applies to the regime shift model.

5 Monte Carlo simulations

In this section, we investigate the small-sample properties of the LM based tests through Monte Carlo simulations. The DGP used for this purpose is given by the following system of equations

$$
\begin{aligned}
y_{it} &= \delta D_{it} + D_{it}x_{it}\gamma + x_{it}\beta + z_{it}, \\
z_{it} &= \rho z_{it-1} + e_{it}, \\
x_{it} &= x_{it-1} + w_{it}.
\end{aligned}
$$

For the error process, we have two scenarios. In the first, $e_{it} = u_{it} + \theta u_{it-1}$ so e_{it} follows a first order moving average process. In the second, we have $e_{it} = \gamma e_{it-1} + u_{it}$, in which case e_{it} follows a first order autoregressive process. In both cases, we have $u_{it} \sim N(0, 1)$ and $w_{it} \sim N(0, 1)$, and we use the value zero to initiate x_{it}, z_{it}, e_{it} and u_{it}.

The DGP is parameterized as follows. For the structural breaks, we have two different configurations, each of which correspond to one of our two model specifications. For the level break model, we have $\gamma = 0$ and $\delta = 5$, whereas, for the regime shift model, $\gamma = \delta = 5$. For convenience, we assume a common break fraction λ for all i. The parameter ρ determines whether the null hypothesis it true or not. As with the other parameters, we make the simplifying assumption that ρ takes on a common value for all i. Under the null hypothesis, we have $\rho = 1$, while, under the alternative hypothesis, we have $\rho < 1$.

The data is generated for $1,000$ panels with N cross-sectional and $T + 50$ time series observations. The first 50 observations for each cross-section is then disregarded in order to avoid possible initial value effect.

In constructing the LM based test statistics, we need to decide on how to remove the serial correlation in the regression errors. Towards this end, consistent with the results of e.g. Schwert (1989) and DeJong *et al.* (1992), we find that the semiparametric adjustment method proposed by Schmidt and Phillips (1992) does not work very well for typical sample sizes. Therefore, in this section we report only the results based on the parametric long autoregression method of Ahn (1993) and Amsler and Lee (1995). In so doing, we follow the recommendation of Zivot and Andrews (1992) and let the order of the lag augmentation to increase with T. In particular, since there is no obvious choice, we follow Newey and West (1994) and set the lag order to the largest integer less than $4(T/100)^{2/9}$.

For brevity, we only report the size and size-adjusted power of a nominal 5% level test. The results on the accuracy of the estimated breakpoints are not presented but will be discussed when appropriate. These results are available from the corresponding author upon request. All computational work was performed in GAUSS.

Consider first the results of the size presented in Table 1 for the level shift model and Table 2 for the regime shift model. There are several things that are noteworthy. First, it has been well documented in the earlier literature that negative moving average errors may cause substantial size distortions when testing the null hypothesis of no cointegration (see, e.g. Haug, 1996). In agreement with these results, Tables 1 and 2 suggest that the Z_ϕ type tests tend to reject the null hypothesis too frequently when $\gamma < 0$ or $\theta < 0$. In most cases, there is even a compounding effect as the distortions tend to accumulate and to become very serious as N grows. By contrast, the Z_t type tests generally maintains the nominal level well and do not suffer from this kind of distortions in the presence of negative autoregressive or negative moving average errors. Thus, the parametric correction method seem to work very well for this test.

Second, the best size accuracy in the level break model is obtained by using the test based on no breaks. The tests based on estimated and known breakpoints are about equally accurate, and perform only slightly worse. This is not unexpected given that the tests based on the assumption of no breaks are asymptotically independent of the breaks under the null in the level shift case. Hence, even the tests based on no breaks are expected to perform well here. The ranking of the tests is reversed in the regime shift model, in which case the tests based on estimated and known breakpoints perform best.

Third, the results for the level break model reveal that the performance of the LM based tests under the null hypothesis is not affected much by the location of the structural break, which corroborates the results given in Theorem 2. By contrast, the size of the unmodified LM based tests in the regime shift model is seen to be highly dependent on the location of the break. This accords well with the asymptotic theory suggesting that these tests should depend on both the regressors and the location of the breaks.

The results of the power of the tests, which are presented in Tables 3 to 6, generally coincides with what might be expected from the asymptotic theory. First, the power increases faster when increasing T rather than N, which is presumably a reflection of the sequential limit method used in the derivation of the test. Second, the power increases as ρ departs form its hypothesized value of one. Third, the tests based on the assumption of no break are generally the least powerful, thus corroborating the result that erroneous omission of structural breaks should affect the tests by lowering their power. This effect is particularly prominent in the regime shift model, in which case the results suggest the power of the no break tests need not be much larger than their size.

Fourth, the tests based on estimated breakpoints together with those based on known breakpoints produce the best power. In particular, it is seen that there is generally no loss of power involved in estimating the breaks rather than treating them as known. This is due to the accuracy of the estimated breakpoints, which is almost perfect in all cases considered. In fact, the results on the estimated breakpoints suggest that the correct selection frequency rarely falls below 95%. Of course, as indicated in Section 4, the accuracy of these estimates will in general depend on the magnitude of the structural breaks. Thus, smaller values of the break parameters δ and γ are expected to lead to lower accuracy and vice versa.

Overall, the simulations lead us to the conclusion that the LM based tests perform well in general with good power and small size distortions in most experiments. However, in the presence of negative autoregressive or moving average errors, we find that the Z_ϕ statistic can perform quite poorly. As a practical matter, we therefore recommend using the Z_t statistic.

6   Conclusions

This paper proposes two new panel cointegration tests that are appropriate when the cross-sectional units are independent of each other. The tests are based on the univariate unit root tests developed by Schmidt and Phillips (1992), Ahn (1993) and Amsler and Lee (1996), and they are derived by applying the LM, or score, principle to an unobserved components representation of the data. Allowable features include heteroskedastic and serially correlated errors, individual specific time trends, and a single break in both the intercept and slope of each regression, which may be located different dates for different individuals. Yet, both tests are shown to be surprisingly straightforward and very easy to implement.

The new statistics are analyzed when there is no break, when the locations of the breaks are assumed to be known and when they are unknown. By using sequential limit arguments, we show that the statistics have limiting normal distributions, which are free of nuisance parameters under the null hypothesis. In particular, we show that the asymptotic null distributions are independent of the both the trend and structural break. This makes the tests very simple to implement since there is no need to consult a table for the critical values for every possible combination of breakpoint and deterministic specification,

which may be a very tedious undertaking.

Secondly, we show that the asymptotic distributions of the test statistics are independent of the presence of stochastic regressors, which make them even simpler to implement. Finally, we show that erroneous omission of structural breaks does not affect the asymptotic null distribution of the statistics in the intercept break case.

The small-sample performance of the tests is evaluated via Monte Carlo simulations. The results, which generally accords well with the asymptotic theory, suggest that the tests generally perform well with small size distortions and good power even in very small samples.

Appendix: Mathematical proofs

In this appendix, we derive the LM statistic and the limiting distributions of the LM based statistics under the null hypothesis. For brevity, the notations introduced in the main text are taken as given and are thus not repeated.

Proposition A.1. (The panel LM test statistic.) Under Assumption 1, the panel LM test statistic of H_0 versus H_1 is given by

$$\xi_{NT} \;=\; \sum_{i=1}^{N} \left(\sum_{t=2}^{T} \Delta \widehat{S}_{it} \widehat{S}_{it-1} \right)^2 \left(\widehat{\sigma}^2 \sum_{t=2}^{T} \widehat{S}_{it-1}^2 \right)^{-1}.$$

Proof of Proposition A.1

We begin by rewriting equations (1) and (2) in the text as follows

$$
\begin{aligned}
y_{i1} &= \alpha_i + \tau_i + x_{i1}'\beta_i + \rho_i z_{i0} + e_{i1}, & \text{(A1)}\\
y_{it} &= \rho_i y_{it-1} + \alpha_i(1-\rho_i) + \tau_i(t + \rho_i(1-t)) \\
&+ (x_{it} - \rho_i x_{it-1})'\beta_i + e_{it}, & \text{(A2)}
\end{aligned}
$$

where the last equality holds for $t = 2, ..., T$. Assume that z_{i0} is fixed and that Assumption 1 holds, then (A1) and (A2) imply the following log-likelihood function

$$\log L \;=\; -\frac{NT}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{N} SSE_i, \qquad \text{(A3)}$$

where $SSE_i = \sum_{t=1}^{T} e_{it}^2$ and e_{it} is defined by (A1) and (A2). It easy to verify that this log-likelihood is maximized for $\widehat{\sigma}^2 = \frac{1}{NT} \sum_{i=1}^{N} SSE_i$ and so the concentrated log-likelihood function becomes

$$\log L \;=\; -\frac{NT}{2} \log 2\pi - \frac{NT}{2} - \frac{NT}{2} \log \widehat{\sigma}^2. \qquad \text{(A4)}$$

Let γ_i denote the vector containing the true parameters $\widetilde{\alpha}_i$, τ_i and β_i, then the score of the concentrated log-likelihood is given by

$$\frac{\partial \log L}{\partial \gamma_i} \;=\; -\frac{1}{2\widehat{\sigma}^2} \frac{\partial SSE_i}{\partial \gamma_i}. \qquad \text{(A5)}$$

We begin by deriving the maximum likelihood estimators of $\widetilde{\alpha}_i$, τ_i and β_i under the null hypothesis. These can be obtained by minimizing SSE_i under the null. Towards this end, let $S_{it} = y_{it} - \widetilde{\alpha}_i - \tau_i t - x_{it}'\beta_i$ and assume that $\rho_i = 1$, in which case SSE_i reduces to

$$SSE_i \;=\; S_{i1}^2 + \sum_{t=2}^{T} \Delta S_{it}^2.$$

Direct calculation reveal that the elements of the score of SSE_i with respect to γ_i are given by

$$\frac{\partial SSE_i}{\partial \widetilde{\alpha}_i} = -2S_{i1}, \tag{A6}$$

$$\begin{aligned}
\frac{\partial SSE_i}{\partial \tau_i} &= -2S_{i1} - 2\sum_{t=2}^{T} \Delta S_{it} \\
&= -2S_{i1} - 2(S_{iT} - S_{i1}) \\
&= -2S_{iT}, \tag{A7}
\end{aligned}$$

$$\frac{\partial SSE_i}{\partial \beta_i} = -2x_{i1}S_{i1} - 2\sum_{t=2}^{T} \Delta x_{it}\Delta S_{it}. \tag{A8}$$

By the first order condition, and some algebra, we obtain the restricted maximum likelihood estimators as explained in the text.

Now, to derive the LM test statistic, we need to obtain the score of SSE_i with respect to ρ_i. This is given by

$$\frac{\partial SSE_i}{\partial \rho_i} = -2z_{i0}e_{i1} - 2\sum_{t=2}^{T} e_{it}(S_{it-1} + z_{i0}). \tag{A9}$$

If we let $\rho_i = 1$, then (A9) reduces to the following expression when evaluated at the restricted maximum likelihood estimators

$$\begin{aligned}
\frac{\partial SSE_i}{\partial \rho_i} &= -2z_{i0}\widehat{S}_{i1} - 2z_{i0}\sum_{t=2}^{T} \Delta\widehat{S}_{it} - 2\sum_{t=2}^{T} \Delta\widehat{S}_{it}\widehat{S}_{it-1} \\
&= -2z_{i0}\widehat{S}_{i1} - 2z_{i0}(\widehat{S}_{iT} - \widehat{S}_{i1}) - 2\sum_{t=2}^{T} \Delta\widehat{S}_{it}\widehat{S}_{it-1} \\
&= -2z_{i0}\widehat{S}_{iT} - 2\sum_{t=2}^{T} \Delta\widehat{S}_{it}\widehat{S}_{it-1} \\
&= -2\sum_{t=2}^{T} \Delta\widehat{S}_{it}\widehat{S}_{it-1}. \tag{A10}
\end{aligned}$$

The last equality follows from the fact that $\widehat{S}_{iT}$ is equal to zero, which can be seen by writing

$$\begin{aligned}
\widehat{S}_{iT} &= y_{iT} - \widehat{\alpha}_i - \widehat{\tau}_i T - x'_{iT}\widehat{\beta}_i \\
&= y_{iT} - (y_{i1} - \widehat{\tau}_i - x'_{i1}\widehat{\beta}_i) - \widehat{\tau}_i T - x'_{iT}\widehat{\beta}_i \\
&= y_{iT} - y_{i1} - \widehat{\tau}_i(T-1) - (x_{iT} - x_{i1})'\widehat{\beta}_i.
\end{aligned}$$

If we let $\Delta y_{i.}$ and $\Delta x_{i.}$ denote the sample averages of Δy_{it} and Δx_{it}, respectively, then this expression becomes

$$\begin{aligned}
\widehat{S}_{iT} &= \sum_{t=2}^{T} \Delta y_{it} - \widehat{\tau}_i(T-1) - \sum_{t=2}^{T} \Delta x'_{it}\widehat{\beta}_i \\
&= \sum_{t=2}^{T} \Delta y_{it} - (\Delta y_{i.} - \Delta x'_{i.}\widehat{\beta}_i)(T-1) - \sum_{t=2}^{T} \Delta x'_{it}\widehat{\beta}_i,
\end{aligned}$$

17

which is zero.

Next, we show that the Hessian of the restricted log-likelihood is asymptotically block-diagonal between ρ_i and γ_i, in which case the LM statistic of H_0 versus H_1 has the simple form

$$\xi_{NT} \;\equiv\; \sum_{i=1}^{N} \left(\frac{\partial \log L}{\partial \rho_i}\right)^2 \left(\frac{\partial^2 \log L}{\partial \rho_i^2}\right)^{-1}. \tag{A11}$$

From (A5), we can see that

$$\frac{\partial^2 \log L}{\partial \gamma_i \partial \gamma_i'} \;=\; -\frac{1}{2\widehat{\sigma}^2} \frac{\partial^2 SSE_i}{\partial \gamma_i \partial \gamma_i'}. \tag{A12}$$

Straightforward calculations yield the following second order partial derivatives when evaluated at the restricted maximum likelihood estimators

$$\frac{\partial^2 SSE_i}{\partial \rho_i \partial \widetilde{\alpha}_i} \;=\; 2z_{i0}, \tag{A13}$$

$$\frac{\partial^2 SSE_i}{\partial \rho_i \partial \tau_i} \;=\; 2T z_{i0} + 2\sum_{t=2}^{T}(\Delta \widehat{S}_{it}(t-1) + \widehat{S}_{it-1})$$

$$=\; 2T z_{i0} + 2(\widehat{S}_{iT} - \widehat{S}_{i1})(t-1) + 2\sum_{t=2}^{T} \widehat{S}_{it-1}$$

$$=\; 2T z_{i0} + 2\sum_{t=2}^{T} \widehat{S}_{it-1}, \tag{A14}$$

where the last equality holds because $\widehat{S}_{i1}$ is zero, which follows by writing

$$\widehat{S}_{i1} \;=\; y_{i1} - \widehat{\alpha}_i - \widehat{\tau}_i - x_{i1}'\widehat{\beta}_i$$
$$=\; y_{i1} - (y_{i1} - \widehat{\tau}_i - x_{i1}'\widehat{\beta}_i) - \widehat{\tau}_i - x_{i1}'\widehat{\beta}_i.$$

For $\partial^2 SSE_i/\partial \rho_i \partial \beta_i$, we have

$$\frac{\partial^2 SSE_i}{\partial \rho_i \partial \beta_i} \;=\; 2x_{i1}z_{i0} + 2\sum_{t=2}^{T}(\Delta x_{it}(\widehat{S}_{it-1} + z_{i0}) + x_{it-1}\Delta\widehat{S}_{it})$$

$$=\; 2x_{i1}z_{i0} + 2(x_{iT} - x_{i1})z_{i0} + 2\sum_{t=2}^{T}(\Delta x_{it}\widehat{S}_{it-1} + x_{it-1}\Delta\widehat{S}_{it})$$

$$=\; 2x_{iT}z_{i0} + 2\sum_{t=2}^{T}(\Delta x_{it}\widehat{S}_{it-1} + x_{it-1}\Delta\widehat{S}_{it}). \tag{A15}$$

The appropriate normalizing order of these derivatives is determined by the order of the second derivative with respect to ρ_i. Similar calculations yield

$$\frac{\partial^2 SSE_i}{\partial \rho_i^2} \;=\; 2z_{i0}^2 + 2\sum_{t=2}^{T}(S_{it-1} + z_{i0})^2$$

$$=\; 2\sum_{t=2}^{T} \widehat{S}_{it-1}^2. \tag{A16}$$

18

The second equality follows by evaluating the second term at the restricted maximum likelihood estimators, and by ignoring the first, which is $o_p(1)$.

In the proof of Theorem 1, we show that $T^{-1/2}\widehat{S}_{it}$ is $O_p(1)$, which implies that $T^{-2}\sum_{t=2}^{T}\widehat{S}_{it-1}^2$ is $O_p(1)$ too. Hence, the appropriate normalizing order for the Hessian is T^{-2}. Moreover, by standard asymptotic arguments, we can show that the scaled quantities $T^{-3/2}\sum_{t=2}^{T}\widehat{S}_{it-1}$, $T^{-1}\sum_{t=2}^{T}\Delta x_{it}\widehat{S}_{it-1}$ and $T^{-1}\sum_{t=2}^{T}x_{it-1}\Delta\widehat{S}_{it}$ are $O_p(1)$. This shows that the expressions appearing in (A13) to (A15), when normalized by T^{-2}, must be $o_p(1)$. Hence, the Hessian is indeed asymptotically block-diagonal.

By using (A10) and (A16), we can thus show that the expression in (A11) has the following asymptotically equivalent representation

$$\xi_{NT} = \sum_{i=1}^{N}\left(\sum_{t=2}^{T}\Delta\widehat{S}_{it}\widehat{S}_{it-1}\right)^2\left(\widehat{\sigma}^2\sum_{t=2}^{T}\widehat{S}_{it-1}^2\right)^{-1}. \qquad (A17)$$

This completes the derivation of the LM test statistic. ∎

Proof of Theorem 1.

We begin by considering the limit of $\widehat{S}_{it}$. This variable can be written as

$$\begin{aligned}\widehat{S}_{it} &= y_{it} - \widehat{\alpha}_i - \widehat{\tau}_i t - x'_{it}\widehat{\beta}_i \\ &= y_{it} - y_{i1} - \widehat{\tau}_i(t-1) - (x_{it} - x_{i1})'\widehat{\beta}_i \\ &= \sum_{k=2}^{t}e_{ik} - (x_{it}-x_{i1})'(\widehat{\beta}_i - \beta_i) - (\widehat{\tau}_i - \tau_i)(t-1). \qquad (A18)\end{aligned}$$

For the third term, we have

$$\begin{aligned}\widehat{\tau}_i - \tau_i &= \Delta y_{i.} - \Delta x'_{i.}\widehat{\beta}_i - \tau_i \\ &= e_{i.} - \Delta x'_{i.}(\widehat{\beta}_i - \beta_i),\end{aligned}$$

where $\Delta y_{i.}$, $\Delta x_{i.}$ and $e_{i.}$ are the sample means of Δy_{it}, Δx_{it} and e_{it}, respectively. Further, let $w_{i.}$ be the mean of w_{it}, then $\widehat{\tau}_i - \tau_i = e_{i.} - w'_{i.}(\widehat{\beta}_i - \beta_i)$. This implies that (A18) can be rewritten as follows

$$\begin{aligned}\widehat{S}_{it} &= \sum_{k=2}^{t}e_{ik} - e_{i.}(t-1) - \sum_{k=2}^{t}w'_{ik}(\widehat{\beta}_i - \beta_i) + w'_{i.}(\widehat{\beta}_i - \beta_i)(t-1) \\ &= \sum_{k=2}^{t}(e_{ik}-e_{i.}) - \sum_{k=2}^{t}(w_{ik}-w_{i.})'(\widehat{\beta}_i - \beta_i). \qquad (A19)\end{aligned}$$

As shown by Schmidt and Phillips (1992), we have $T^{-1/2}\sum_{k=2}^{t}(e_{ik}-e_{i.}) \Rightarrow \sigma V_i(r)$ as $T \longrightarrow \infty$ for a fixed N. For the second term in (A19), note that $T^{-1/2}\sum_{k=2}^{t}(w_{ik}-w_{i.}) \Rightarrow \Omega_i^{1/2}\widetilde{V}_i(r)$ as $T \longrightarrow \infty$, where $\widetilde{V}_i(r)$ is a standard Brownian bridge process. Also, if we let $\Delta\widetilde{x}_{it}$ and $\widetilde{e}_{it}$ denote the deviations of Δx_{it} and e_{it} from their respective mean values, then $T^{1/2}(\widehat{\beta}_i - \beta_i)$ can be

written as

$$T^{1/2}(\widehat{\beta}_i - \beta_i) = T^{1/2}\left(\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\Delta\widetilde{x}'_{it}\right)^{-1}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\widetilde{e}_{it}$$

$$= \left(T^{-1}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\Delta\widetilde{x}'_{it}\right)^{-1}T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\widetilde{e}_{it}.$$

If we let $\xrightarrow{p}$ signify convergence in probability, then the denominator of this expression is $T^{-1}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\Delta\widetilde{x}'_{it} \xrightarrow{p} \Omega_i > 0$ by Assumption 1 (iv) and standard asymptotic arguments for stationary processes. For the numerator, we have $T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\widetilde{e}_{it} = O_p(1)$ since w_{it} and e_{it} are orthogonal by Assumption 1 (ii). It follows that $T^{1/2}(\widehat{\beta}_i - \beta_i) = O_p(1)$. Therefore, we get

$$T^{-1/2}\widehat{S}_{it} = T^{-1/2}\sum_{k=2}^{t}(e_{ik} - e_{i.}) - T^{-1}\sum_{k=2}^{t}(w_{ik} - w_{i.})'T^{1/2}(\widehat{\beta}_i - \beta_i)$$

$$= T^{-1/2}\sum_{k=2}^{t}(e_{ik} - e_{i.}) + o_p(1)$$

$$\Rightarrow \sigma V_i(r). \tag{A20}$$

It is useful to rewrite $\widehat{S}_{it}$ as

$$\widehat{S}_{it} = y_{it} - y_{i1} - \widehat{\tau}_i(t-1) - (x_{it} - x_{i1})'\widehat{\beta}_i$$

$$= y_{it} - y_{i1} - (\Delta y_{i.} - \Delta x'_{i.}\widehat{\beta}_i)(t-1) - (x_{it} - x_{i1})'\widehat{\beta}_i$$

$$= \sum_{k=2}^{t}(\Delta\widetilde{y}_{it} - \Delta\widetilde{x}'_{it}\widehat{\beta}_i). \tag{A21}$$

It follows that $\Delta\widehat{S}_{it} = \Delta\widetilde{y}_{it} - \Delta\widetilde{x}'_{it}\widehat{\beta}_i$. The model in first differences is given by $\Delta y_{it} = \widehat{\tau}_i + \Delta x'_{it}\widehat{\beta}_i + \widehat{e}_{it}$ or $\Delta\widetilde{y}_{it} = \Delta\widetilde{x}'_{it}\widehat{\beta}_i + \widetilde{e}_{it}$, where $\widetilde{e}_{it}$ now denotes the deviation of $\widehat{e}_{it}$ from its mean value. Thus, by using (A21), we can deduce that $\Delta\widehat{S}_{it}$ is equal to $\widetilde{e}_{it}$.

Now, consider Z_ϕ. This statistic can be written as

$$TZ_\phi = T\sum_{i=1}^{N}\widetilde{S}_{i11}^{-1}\widetilde{S}_{i12}$$

$$= \sum_{i=1}^{N}\left(T^{-2}\widetilde{S}_{i11}\right)^{-1}T^{-1}\widetilde{S}_{i12}. \tag{A22}$$

For the denominator, we have the following limit passing $T \longrightarrow \infty$

$$T^{-2}\widetilde{S}_{i11} = T^{-2}\sum_{t=2}^{T}(\widehat{S}_{it-1} - \widehat{S}_{i.-1})^2$$

$$\Rightarrow \sigma^2\int_0^1\overline{V}_i(r)^2\mathrm{d}r. \tag{A23}$$

Similarly, for the denominator, we have

$$
\begin{aligned}
T^{-2}\widetilde{S}_{i12} &= T^{-1}\sum_{t=2}^{T}(\widehat{S}_{it-1} - \widehat{S}_{i.-1})(\Delta\widehat{S}_{it} - \Delta\widehat{S}_{i.}) \\
&= -\frac{1}{2}\left(T^{-1}\sum_{t=2}^{T}(\Delta\widehat{S}_{it} - \Delta\widehat{S}_{i.})^2\right) \\
&= -\frac{1}{2}\left(T^{-1}\sum_{t=2}^{T}\widetilde{e}_{it}^2\right) \\
&\xrightarrow{p} -\frac{1}{2}\sigma^2.
\end{aligned}
\tag{A24}
$$

For the second equality, we use Lemma 1 of Schmidt and Phillips (1992), from which it follows that $\sum_{t=2}^{T}(\widehat{S}_{it-1} - \widehat{S}_{i.})(\Delta\widehat{S}_{it} - \Delta\widehat{S}_{i.}) = -\frac{1}{2}\sum_{t=2}^{T}(\Delta\widehat{S}_{it} - \Delta\widehat{S}_{i.})^2$. Equations (A23) and (A24) imply that the limit of TZ_ϕ as $T \longrightarrow \infty$ with N held fixed is equal to

$$
\begin{aligned}
TZ_\phi &= \sum_{i=1}^{N}\left(T^{-2}\widetilde{S}_{i11}\right)^{-1}T^{-1}\widetilde{S}_{i12} \\
&= \sum_{i=1}^{N}-\frac{1}{2}\left(\int_0^1\overline{V}_i(r)^2\mathrm{d}r\right)^{-1}.
\end{aligned}
\tag{A25}
$$

Now, to derive the sequential limit of Z_ϕ, define $E_i \equiv \left(T^{-2}\widetilde{S}_{i11}\right)^{-1}T^{-1}\widetilde{S}_{i12}$ and then expand the statistic as follows

$$
TN^{-1/2}Z_\phi - N^{1/2}\Theta_2 = N^{1/2}\left(N^{-1}\sum_{i=1}^{N}E_i - \Theta_2\right).
$$

Assume that $E(K_i^2) = \Sigma_{11} < \infty$ exist. Then, by Assumption 1 (i) and the Lindberg-Lévy central limit theorem, as $T \longrightarrow \infty$ and then N sequentially, $TN^{-1/2}Z_\phi - N^{1/2}\Theta_2 \Rightarrow N(0, \Sigma_{22})$. This establishes the first part of the proof.

Consider next the limiting distribution for Z_t. Similar to the analysis Z_ϕ, since $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$, we obtain the following limit as $T \longrightarrow \infty$

$$
\begin{aligned}
Z_t &= \sum_{i=1}^{N}\left(\widehat{\sigma}^{-2}\widetilde{S}_{i11}\right)^{1/2}E_i \\
&= \sum_{i=1}^{N}\left(\widehat{\sigma}^{-2}T^{-2}\widetilde{S}_{i11}\right)^{1/2}TE_i \\
&= \sum_{i=1}^{N}-\frac{1}{2}\left(\int_0^1\overline{V}_i(r)^2\mathrm{d}r\right)^{-1/2}.
\end{aligned}
\tag{A26}
$$

If we define $Q_i \equiv \widetilde{S}_{i11}^{1/2}E_i$, then Z_t can be rewritten as

$$
N^{-1/2}Z_t - N^{1/2}\Theta_1 = N^{1/2}\left(N^{-1}\sum_{i=1}^{N}\widehat{\sigma}^{-1}Q_i - \Theta_1\right).
$$

Given that $E(U_i^2) = \Sigma_{22} < \infty$ exist, then $TN^{-1/2}Z_\phi - N^{1/2}\Theta_2 \Rightarrow N(0, \Sigma_{22})$ as $T \longrightarrow \infty$ and then $N \longrightarrow \infty$ by similar arguments used for Z_ϕ. This establishes the second part of the proof. ∎

Proof of Theorem 2.

To prove Theorem 2 (i), we need to show that $T^{-1/2}\widehat{S}_{it}$ is asymptotically unaffected by the inclusion of the break dummy D_{it}. In so doing, we begin by rewriting $\widehat{S}_{it}^c$ as follows

$$
\begin{aligned}
\widehat{S}_{it}^c &= y_{it} - \widehat{\alpha}_i - \widehat{\tau}_i t - \widehat{\delta}_i D_{it} - x_{it}'\widehat{\beta}_i \\
&= y_{it} - y_{i1} - \widehat{\tau}_i(t-1) - \widehat{\delta}_i(D_{it} - D_{i1}) - (x_{it} - x_{i1})'\widehat{\beta}_i \\
&= \sum_{k=2}^t e_{ik} - (\widehat{\delta}_i - \delta_i)(D_{it} - D_{i1}) - (x_{it} - x_{i1})'(\widehat{\beta}_i - \beta_i) \\
&\quad - (\widehat{\tau}_i - \tau_i)(t-1). \tag{A27}
\end{aligned}
$$

Let $\Delta y_{i\cdot}$, $\Delta x_{i\cdot}$ and $e_{i\cdot}$ be the sample averages of Δy_i, Δx_{it} and e_{it}, respectively, and note that the sample average of ΔD_{it} is $(T-1)^{-1}$. Then, the fourth term on the right-hand side of (A27) can be written as

$$
\begin{aligned}
\widehat{\tau}_i - \tau_i &= \Delta y_{i\cdot} - (T-1)^{-1}\widehat{\delta}_i - \Delta x_{i\cdot}'\widehat{\beta}_i - \tau_i \\
&= e_{i\cdot} - (T-1)^{-1}(\widehat{\delta}_i - \delta_i) - \Delta x_{i\cdot}'(\widehat{\beta}_i - \beta_i) \\
&= e_{i\cdot} - (T-1)^{-1}(\widehat{\delta}_i - \delta_i) - w_{i\cdot}'(\widehat{\beta}_i - \beta_i).
\end{aligned}
$$

Hence, by using (A27), we get

$$
\begin{aligned}
T^{-1/2}\widehat{S}_{it}^c &= T^{-1/2}\sum_{k=2}^t e_{ik} - e_{i\cdot}(t-1) - T^{-1/2}\sum_{k=2}^t w_{ik}'(\widehat{\beta}_i - \beta_i) \\
&\quad + w_{i\cdot}'(\widehat{\beta}_i - \beta_i)(t-1) - T^{-1/2}(\widehat{\delta}_i - \delta_i)\left(D_{it} - D_{i1} - \frac{t-1}{T-1}\right) \\
&= T^{-1/2}\sum_{k=2}^t (e_{ik} - e_{i\cdot}) - T^{-1/2}\sum_{k=2}^t (w_{ik} - w_{i\cdot})'(\widehat{\beta}_i - \beta_i) \\
&\quad - T^{-1/2}(\widehat{\delta}_i - \delta_i)\left(D_{it} - D_{i1} - \frac{t-1}{T-1}\right) \\
&= I - II - III. \tag{A28}
\end{aligned}
$$

Now, part II is $o_p(1)$ by similar arguments used in the proof of Theorem 1. Therefore, to prove part (i) of Theorem 2, we only need to show that III is $o_p(1)$, in which case $T^{-1/2}\widehat{S}_{it}^c$ has the same distribution as $T^{-1/2}\widehat{S}_{it}$.

Let $\Delta\widetilde{x}_{it}$, $\Delta\widetilde{D}_{it}$ and $\widetilde{e}_{it}$ be the deviations of Δx_{it}, ΔD_{it} and e_{it} from their respective mean values. Also, let $\Delta\widetilde{D}_{it}^X = \Delta\widetilde{D}_{it} - \Delta\widetilde{x}_{it}'b_i$ and $\widetilde{e}_{it}^X = \widetilde{e}_{it} - \Delta\widetilde{x}_{it}'a_i$ be the errors from projecting $\Delta\widetilde{D}_{it}$ and $\widetilde{e}_{it}$, respectively, onto $\Delta\widetilde{x}_{it}$. Consider $T^{1/2}(\widehat{\delta}_i - \delta_i)$. This term may be written as

$$
\begin{aligned}
T^{1/2}(\widehat{\delta}_i - \delta_i) &= T^{1/2}\left(\sum_{t=2}^T \Delta\widetilde{D}_{it}^{X2}\right)^{-1}\sum_{t=2}^T \Delta\widetilde{D}_{it}^X \widetilde{e}_{it}^X \\
&= \left(T^{-1}\sum_{t=2}^T \Delta\widetilde{D}_{it}^{X2}\right)^{-1}T^{-1/2}\sum_{t=2}^T \Delta\widetilde{D}_{it}^X \widetilde{e}_{it}^X. \tag{A29}
\end{aligned}
$$

The denominator of this expression can be expanded as follows

$$
\begin{aligned}
T^{-1}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}^{X2} &= T^{-1}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}^{2} - 2T^{-1}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}\Delta\widetilde{x}_{it}'b_i \\
&\quad + T^{-1}\sum_{t=2}^{T}b_i'\Delta\widetilde{x}_{it}\Delta\widetilde{x}_{it}'b_i \\
&= T^{-1}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}^{2} - T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}\Delta\widetilde{x}_{it}' \\
&\quad \cdot \left(T^{-1}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\Delta\widetilde{x}_{it}'\right)^{-1}T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\Delta\widetilde{D}_{it}. \quad\text{(A30)}
\end{aligned}
$$

Now, since ΔD_{it} equals one only at one point, it is obvious that all sums involving $\Delta\widetilde{D}_{it}$ must be $o_p(1)$ when normalized by $T^{-1/2}$ or T^{-1}. Hence, we can deduce that $T^{-1}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}^{2}$ and $T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\Delta\widetilde{D}_{it}$ must be $o_p(1)$. Also, since $T^{-1}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\Delta\widetilde{x}_{it}' \xrightarrow{p} \Omega_i$ as $T \longrightarrow \infty$, we get $T^{-1}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}^{X2} = o_p(1)$.

Similarly, for the numerator of (A29), we have

$$
\begin{aligned}
T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}^{X}\widetilde{e}_{it}^{X} &= T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}\widetilde{e}_{it} - T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}\Delta\widetilde{x}_{it}'a_i \\
&\quad - T^{-1/2}\sum_{t=2}^{T}\widetilde{e}_{it}\Delta\widetilde{x}_{it}'b_i + T^{-1/2}\sum_{t=2}^{T}a_i'\Delta\widetilde{x}_{it}\Delta\widetilde{x}_{it}'b_i \\
&= T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}\widetilde{e}_{it} - T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}\Delta\widetilde{x}_{it}' \\
&\quad \cdot \left(T^{-1}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\Delta\widetilde{x}_{it}'\right)^{-1}T^{-1}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\widetilde{e}_{it}. \quad\text{(A31)}
\end{aligned}
$$

We have that $T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}\widetilde{e}_{it} = o_p(1)$ and $T^{-1}\sum_{t=2}^{T}\Delta\widetilde{x}_{it}\widetilde{e}_{it} \xrightarrow{p} 0$ as $T \longrightarrow \infty$, where the latter follows from the fact that $\Delta\widetilde{x}_{it}$ and $\widetilde{e}_{it}$ are independent by Assumption 1 (ii). Thus, we can show that $T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{D}_{it}^{X}\widetilde{e}_{it}^{X} = o_p(1)$.

These results imply that $T^{1/2}(\widehat{\beta}_i - \beta_i) = o_p(1)$. Hence, by using (A28), it follows that

$$
\begin{aligned}
III &= T^{1/2}(\widehat{\delta}_i - \delta_i)\left(T^{-1}(D_{it} - D_{i1}) - \frac{t-1}{T(T-1)}\right) \\
&= o_p(1) \cdot o_p(1).
\end{aligned}
$$

We have thus shown that III is $o_p(1)$ as required for the proof of (i).

Next, consider Theorem 2 (ii). From (A18), the variable $\widehat{S}_{it}$ assuming no breaks can be written as

$$
\begin{aligned}
\widehat{S}_{it} &= y_{it} - y_{i1} - \widehat{\tau}_i(t-1) - (x_{it} - x_{i1})'\widehat{\beta}_i \\
&= \sum_{k=2}^{t}e_{ik} - \sum_{k=2}^{t}w_{ik}'(\widehat{\beta}_i - \beta_i) - (\widehat{\tau}_i - \tau_i)(t-1) + \delta_i(D_{it} - D_{i1}), \quad\text{(A32)}
\end{aligned}
$$

23

where the equality follows from the fact that $y_{it} - y_{i1} = \tau_i(t-1) + \sum_{k=2}^{t} w'_{ik}\beta_i + \delta_i(D_{it} - D_{i1}) + \sum_{k=2}^{t} e_{ik}$. With a break in the DGP, $\widehat{\tau}_i - \tau_i$ can be written as

$$
\begin{aligned}
\widehat{\tau}_i - \tau_i &= \Delta y_{i.} - \Delta x'_{i.}\widehat{\beta}_i - \tau_i \\
&= e_{i.} - \Delta w'_{i.}(\widehat{\beta}_i - \beta_i) + (T-1)^{-1}\delta_i.
\end{aligned}
$$

This suggests that (A32) can be rewritten as

$$
\begin{aligned}
T^{-1/2}\widehat{S}_{it} &= T^{-1/2}\sum_{k=2}^{t}(e_{ik} - e_{i.}) - T^{-1/2}\sum_{k=2}^{t}(w_{ik} - w_{i.})'(\widehat{\beta}_i - \beta_i) \\
&\quad + T^{-1/2}\delta_i(D_{it} - D_{i1}) - T^{-1/2}\delta_i\left(\frac{t-1}{T-1}\right) \\
&= T^{-1/2}\sum_{k=2}^{t}(e_{ik} - e_{i.}) + o_p(1). \tag{A33}
\end{aligned}
$$

Thus, the distribution of $T^{-1/2}\widehat{S}_{it}$ does not depend on the presence of the level break. This establishes (ii). ∎

Proof of Theorem 3.

In this section, we show that the limit of $T^{-1/2}\widehat{S}_{it}^{r}$ is equal to the limit of $T^{-1/2}\widehat{S}_{it}$, which implies that the symptotic distributions of the LM based statistics in the regime shift case are unaffected by the shifts. The proof is very similar to that of Theorem 2 and hence only essential details are given.

For notational convenience, let $E'_{it} = D_{it}x'_{it}$. As in equation (A18), $\widehat{S}_{it}^{r}$ can be written as

$$
\begin{aligned}
\widehat{S}_{it}^{r} &= \sum_{k=2}^{t} e_{ik} - (\widehat{\delta}_i - \delta_i)(D_{it} - D_{i1}) - (x_{it} - x_{i1})'(\widehat{\beta}_i - \beta_i) \\
&\quad - (E_{it} - E_{i1})'(\widehat{\gamma}_i - \gamma_i) - (\widehat{\tau}_i - \tau_i)(t-1). \tag{A34}
\end{aligned}
$$

If we let $\Delta E_{i.}$ denote the sample average of ΔE_{it}, then we have

$$
\widehat{\tau}_i - \tau_i = e_{i.} - (T-1)^{-1}(\widehat{\delta}_i - \delta_i) - w'_{i.}(\widehat{\beta}_i - \beta_i) - \Delta E'_{i.}(\widehat{\gamma}_i - \gamma_i). \tag{A35}
$$

Equation (A35) imply that (A34) can be reformulated as

$$
\begin{aligned}
T^{-1/2}\widehat{S}_{it}^{r} &= T^{-1/2}\sum_{k=2}^{t} e_{ik} - T^{-1/2}e_{i.}(t-1) - T^{-1/2}\sum_{k=2}^{t} w'_{ik}(\widehat{\beta}_i - \beta_i) \\
&\quad + T^{-1/2}w'_{i.}(\widehat{\beta}_i - \beta_i)(t-1) - T^{-1/2}(\widehat{\delta}_i - \delta_i)\left(D_{it} - D_{i1} - \frac{t-1}{T-1}\right) \\
&\quad - T^{-1/2}\left(E_{it} - E_{i1} - \frac{t-1}{T-1}(E_{iT} - E_{i1})\right)'(\widehat{\gamma}_i - \gamma_i) \\
&= T^{-1/2}\sum_{k=2}^{t}(e_{ik} - e_{i.}) - T^{-1/2}\sum_{k=2}^{t}(w_{ik} - w_{i.})'(\widehat{\beta}_i - \beta_i) \\
&\quad - T^{-1/2}(\widehat{\delta}_i - \delta_i)\left(D_{it} - D_{i1} - \frac{t-1}{T-1}\right) \\
&\quad - T^{-1/2}\left(E_{it} - E_{i1} - \frac{t-1}{T-1}(E_{iT} - E_{i1})\right)'(\widehat{\gamma}_i - \gamma_i) \\
&= I - II - III - IV. \tag{A36}
\end{aligned}
$$

24

As with Theorem 2, proving Theorem 3 requires showing that II, III and IV are $o_p(1)$. Part II is an immediate consequence of the proof of Theorem 1 and is thus omitted.

Consider III. Let $\Delta \widetilde{E}_{it} = \Delta E_{it} - \Delta E_{i.}$, and let $\Delta \widetilde{D}_{it}^X = \Delta \widetilde{D}_{it} - X_{it}'b_i$ and $\widetilde{e}_{it}^X = \widetilde{e}_{it} - X_{it}'a_i$ be the errors from projecting $\Delta \widetilde{D}_{it}$ and $\widetilde{e}_{it}$, respectively, onto the generic projection vector $X_{it} = (\Delta \widetilde{x}_{it}', \Delta \widetilde{E}_{it}')'$. Then, $T^{1/2}(\widehat{\delta}_i - \delta_i)$ may be written as

$$T^{1/2}(\widehat{\delta}_i - \delta_i) = \left(T^{-1} \sum_{t=2}^{T} \Delta \widetilde{D}_{it}^{X2} \right)^{-1} T^{-1/2} \sum_{t=2}^{T} \Delta \widetilde{D}_{it}^X \widetilde{e}_{it}^X. \qquad \text{(A37)}$$

The denominator of this expression can be expanded as

$$\begin{aligned} T^{-1} \sum_{t=2}^{T} \Delta \widetilde{D}_{it}^{X2} &= T^{-1} \sum_{t=2}^{T} \Delta \widetilde{D}_{it}^2 - T^{-1} \sum_{t=2}^{T} \Delta \widetilde{D}_{it} X_{it}' \\ &\quad \cdot \left(T^{-1} \sum_{t=2}^{T} X_{it} X_{it}' \right)^{-1} T^{-1} \sum_{t=2}^{T} X_{it} \Delta \widetilde{D}_{it}. \qquad \text{(A38)} \end{aligned}$$

For the second term in (A38), we have

$$T^{-1} \sum_{t=2}^{T} X_{it} \Delta \widetilde{D}_{it} = T^{-1} \sum_{t=2}^{T} \begin{pmatrix} \Delta \widetilde{x}_{it} \Delta \widetilde{D}_{it} \\ \Delta \widetilde{E}_{it} \Delta \widetilde{D}_{it} \end{pmatrix}.$$

We have that $T^{-1} \sum_{t=2}^{T} \Delta \widetilde{x}_{it} \Delta \widetilde{D}_{it} = o_p(1)$ and $T^{-1} \sum_{t=2}^{T} \Delta \widetilde{E}_{it} \Delta \widetilde{D}_{it} = o_p(1)$ so $T^{-1} \sum_{t=2}^{T} X_{it} \Delta \widetilde{D}_{it} = o_p(1)$.

Next, consider $T^{-1} \sum_{t=2}^{T} X_{it} X_{it}'$. This term can be written as

$$T^{-1} \sum_{t=2}^{T} X_{it} X_{it}' = T^{-1} \sum_{t=2}^{T} \begin{pmatrix} \Delta \widetilde{x}_{it} \Delta \widetilde{x}_{it}' & \Delta \widetilde{x}_{it} \Delta \widetilde{E}_{it}' \\ \Delta \widetilde{E}_{it} \Delta \widetilde{x}_{it}' & \Delta \widetilde{E}_{it} \Delta \widetilde{E}_{it}' \end{pmatrix}.$$

Since $T^{-1} \sum_{t=2}^{T} \Delta \widetilde{x}_{it} \Delta \widetilde{x}_{it}' \xrightarrow{p} \Omega_i$ as $T \longrightarrow \infty$, while the remaining three terms in $T^{-1} \sum_{t=2}^{T} X_{it} X_{it}'$ converge in probability to Ω_i if $t > T_i$ and zero otherwise, we have that $T^{-1} \sum_{t=2}^{T} X_{it} X_{it}' = O_p(1)$. Therefore, as the first term on the right hand side of (A38) is $o_p(1)$, we can show that $T^{-1} \sum_{t=2}^{T} \Delta \widetilde{D}_{it}^{X2}$ is also $o_p(1)$.

For the numerator of (A37), we have

$$\begin{aligned} T^{-1/2} \sum_{t=2}^{T} \Delta \widetilde{D}_{it}^X \widetilde{e}_{it}^X &= T^{-1/2} \sum_{t=2}^{T} \Delta \widetilde{D}_{it} \widetilde{e}_{it} - T^{-1} \sum_{t=2}^{T} \Delta \widetilde{D}_{it} X_{it}' \\ &\quad \cdot \left(T^{-1} \sum_{t=2}^{T} X_{it} X_{it}' \right)^{-1} T^{-1/2} \sum_{t=2}^{T} X_{it} \widetilde{e}_{it}. \qquad \text{(A39)} \end{aligned}$$

Because $T^{-1/2} \sum_{t=2}^{T} \Delta \widetilde{D}_{it} \widetilde{e}_{it}$, $T^{-1} \sum_{t=2}^{T} \widetilde{e}_{it} X_{it}$ and $T^{-1/2} \sum_{t=2}^{T} \Delta \widetilde{D}_{it} X_{it}'$ are all $o_p(1)$, and $T^{-1} \sum_{t=2}^{T} X_{it} X_{it}'$ is $O_p(1)$, it follows that $T^{-1/2} \sum_{t=2}^{T} \Delta \widetilde{D}_{it}^X \widetilde{e}_{it}^X = o_p(1)$. Hence, since both $T^{-1} \sum_{t=2}^{T} \Delta \widetilde{D}_{it}^{X2}$ and $T^{-1/2} \sum_{t=2}^{T} \Delta \widetilde{D}_{it}^X \widetilde{e}_{it}^X$ are $o_p(1)$, we can deduce that $T^{1/2}(\widehat{\delta}_i - \delta_i) = o_p(1)$.

By using this result, we can show that

$$
\begin{aligned}
III & = T^{1/2}(\widehat{\delta}_i - \delta_i)\left(T^{-1}(D_{it} - D_{i1}) - \frac{t-1}{T(T-1)}\right) \\
& = o_p(1) \cdot o_p(1).
\end{aligned}
$$

Next, consider IV. The term $T^{1/2}(\widehat{\gamma}_i - \gamma_i)$ appearing in this expression can be written as

$$
T^{1/2}(\widehat{\gamma}_i - \gamma_i) = \left(T^{-1}\sum_{t=2}^{T}\Delta\widetilde{E}_{it}^X\Delta\widetilde{E}_{it}^{X\prime}\right)^{-1}T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{E}_{it}^{X\prime}\widetilde{e}_{it}^X,
$$

where $X_{it} = (\Delta\widetilde{x}_{it}', \Delta\widetilde{D}_{it})'$, $\Delta\widetilde{E}_{it}^X = \Delta\widetilde{E}_{it} - X_{it}'b_i$ and $\widetilde{e}_{it}^X = \widetilde{e}_{it} - X_{it}'a_i$. As in (A37), the denominator of $T^{1/2}(\widehat{\gamma}_i - \gamma_i)$ can be expanded as

$$
\begin{aligned}
T^{-1}\sum_{t=2}^{T}\Delta\widetilde{E}_{it}^X\Delta\widetilde{E}_{it}^{X\prime} & = T^{-1}\sum_{t=2}^{T}\Delta\widetilde{E}_{it}\Delta\widetilde{E}_{it}' - T^{-1}\sum_{t=2}^{T}\widetilde{E}_{it}X_{it}' \\
& \quad \cdot \left(T^{-1}\sum_{t=2}^{T}X_{it}X_{it}'\right)^{-1}T^{-1}\sum_{t=2}^{T}X_{it}\Delta\widetilde{E}_{it}'. \quad \text{(A40)}
\end{aligned}
$$

By using the results derived earlier, it is obvious that both terms on the right hand side of (A40) are $O_p(1)$ if $t > T_i$ and $o_p(1)$ otherwise, which implies that $T^{-1}\sum_{t=2}^{T}\Delta\widetilde{E}_{it}^X\Delta\widetilde{E}_{it}^{X\prime}$ is also $O_p(1)$.

The numerator of $T^{1/2}(\widehat{\gamma}_i - \gamma_i)$ is qual to

$$
\begin{aligned}
T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{E}_{it}^{X\prime}\widetilde{e}_{it}^X & = T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{E}_{it}'\widetilde{e}_{it} - T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{E}_{it}'X_{it}' \\
& \quad \cdot \left(T^{-1}\sum_{t=2}^{T}X_{it}X_{it}'\right)^{-1}T^{-1}\sum_{t=2}^{T}X_{it}\widetilde{e}_{it}. \quad \text{(A41)}
\end{aligned}
$$

In this expression, the terms with normalizing order $T^{-1/2}$ are $O_p(1)$ by standard arguments for stationary processes. Therefore, since $T^{-1}\sum_{t=2}^{T}\Delta X_{it}\widetilde{e}_{it} = o_p(1)$ and $T^{-1}\sum_{t=2}^{T}X_{it}X_{it}' = O_p(1)$, we can deduce that $T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{E}_{it}^{X\prime}\widetilde{e}_{it}^X = o_p(1)$. Consequently, as $T^{-1/2}\sum_{t=2}^{T}\Delta\widetilde{E}_{it}^X\widetilde{e}_{it}^X$ and $T^{-1}\sum_{t=2}^{T}\Delta\widetilde{E}_{it}^X\Delta\widetilde{E}_{it}^{X\prime}$ are both $O_p(1)$, $T^{1/2}(\widehat{\gamma}_i - \gamma_i)$ must also be $O_p(1)$.

This shows that IV must be $o_p(1)$ as seen by writing

$$
\begin{aligned}
IV & = \left(T^{-1}(E_{it} - E_{i1}) - \left(\frac{t-1}{T(T-1)}\right)(E_{iT} - E_{i1})\right)'T^{1/2}(\widehat{\gamma}_i - \gamma_i) \\
& = o_p(1) \cdot O_p(1),
\end{aligned}
$$

where we have used the fact that $E_{it} - E_{i1} = \sum_{k=2}^{t}w_{ik}$ if $t > T_i$ and zero otherwise. We have thus shown that II, III and IV are $o_p(1)$, which implies that (A36) reduces to

$$
T^{-1/2}\widehat{S}_{it}^r = T^{-1/2}\sum_{k=2}^{t}(e_{ik} - e_{i.}) + o_p(1).
$$

This completes the proof. ∎

Table 1: Empirical size in the level shift model

γ	θ	T	N	$\lambda=0.3$						$\lambda=0.7$					
				$\widehat{Z}_t^c$	$\widehat{Z}_\phi^c$	Z_t^c	Z_ϕ^c	Z_t	Z_ϕ	$\widehat{Z}_t^c$	$\widehat{Z}_\phi^c$	Z_t^c	Z_ϕ^c	Z_t	Z_ϕ
0.0	0.0	50	5	0.105	0.122	0.090	0.090	0.082	0.108	0.129	0.136	0.103	0.106	0.086	0.103
0.0	0.0	50	10	0.140	0.163	0.110	0.129	0.092	0.108	0.114	0.149	0.094	0.115	0.064	0.089
0.0	0.0	100	5	0.064	0.101	0.052	0.084	0.053	0.086	0.074	0.095	0.066	0.078	0.059	0.073
0.0	0.0	100	10	0.073	0.109	0.062	0.089	0.067	0.098	0.085	0.118	0.068	0.095	0.070	0.091
0.0	0.3	50	5	0.131	0.028	0.116	0.019	0.096	0.043	0.107	0.024	0.092	0.017	0.078	0.028
0.0	0.3	50	10	0.154	0.027	0.130	0.018	0.110	0.021	0.146	0.014	0.130	0.008	0.099	0.017
0.0	0.3	100	5	0.079	0.013	0.074	0.011	0.076	0.027	0.082	0.014	0.079	0.012	0.071	0.023
0.0	0.3	100	10	0.083	0.009	0.072	0.006	0.064	0.013	0.090	0.009	0.089	0.008	0.072	0.012
0.0	−0.3	50	5	0.101	0.418	0.088	0.397	0.044	0.190	0.101	0.418	0.088	0.388	0.049	0.235
0.0	−0.3	50	10	0.107	0.576	0.087	0.545	0.047	0.287	0.122	0.579	0.096	0.531	0.054	0.321
0.0	−0.3	100	5	0.074	0.404	0.059	0.383	0.035	0.274	0.078	0.443	0.059	0.421	0.047	0.290
0.0	−0.3	100	10	0.075	0.600	0.062	0.575	0.047	0.428	0.068	0.637	0.056	0.604	0.049	0.453
0.3	0.0	50	5	0.110	0.014	0.099	0.011	0.088	0.018	0.117	0.012	0.108	0.008	0.082	0.021
0.3	0.0	50	10	0.124	0.006	0.116	0.003	0.115	0.014	0.152	0.009	0.137	0.004	0.110	0.009
0.3	0.0	100	5	0.103	0.011	0.098	0.011	0.086	0.014	0.074	0.008	0.072	0.008	0.079	0.011
0.3	0.0	100	10	0.079	0.004	0.076	0.002	0.087	0.001	0.075	0.002	0.066	0.001	0.049	0.003
−0.3	0.0	50	5	0.102	0.319	0.088	0.288	0.052	0.180	0.092	0.299	0.076	0.275	0.051	0.169
−0.3	0.0	50	10	0.091	0.475	0.076	0.424	0.059	0.265	0.107	0.460	0.088	0.421	0.046	0.250
−0.3	0.0	100	5	0.075	0.347	0.068	0.328	0.041	0.219	0.078	0.347	0.062	0.318	0.053	0.247
−0.3	0.0	100	10	0.064	0.481	0.054	0.445	0.057	0.349	0.066	0.482	0.049	0.450	0.057	0.345

Notes: The value λ refers to the location of the structural break, γ refers to the autoregressive parameter and θ refers to the moving average parameter. The $\widehat{Z}_t^c$ and $\widehat{Z}_\phi^c$ statistics are based on estimated breakpoints, Z_t^c and Z_ϕ^c are based on known breakpoints, and Z_t and Z_ϕ are based on no breaks.

Table 2: Empirical size in the regime shift model

γ	θ	T	N	$\lambda = 0.3$						$\lambda = 0.7$					
				$\widehat{Z}_t^r$	$\widehat{Z}_\phi^r$	Z_t^r	Z_ϕ^r	Z_t	Z_ϕ	$\widehat{Z}_t^r$	$\widehat{Z}_\phi^r$	Z_t^r	Z_ϕ^r	Z_t	Z_ϕ
0.0	0.0	50	5	0.109	0.097	0.101	0.087	0.281	0.271	0.105	0.103	0.096	0.089	0.041	0.060
0.0	0.0	50	10	0.099	0.120	0.088	0.105	0.421	0.389	0.096	0.109	0.088	0.105	0.030	0.063
0.0	0.0	100	5	0.078	0.101	0.079	0.099	0.225	0.254	0.069	0.090	0.065	0.089	0.024	0.049
0.0	0.0	100	10	0.076	0.111	0.069	0.107	0.361	0.393	0.074	0.096	0.067	0.092	0.019	0.039
0.0	0.3	50	5	0.126	0.017	0.111	0.012	0.274	0.217	0.107	0.021	0.104	0.018	0.039	0.045
0.0	0.3	50	10	0.139	0.012	0.122	0.010	0.390	0.287	0.145	0.019	0.139	0.015	0.048	0.053
0.0	0.3	100	5	0.081	0.016	0.080	0.015	0.203	0.194	0.065	0.013	0.065	0.012	0.019	0.034
0.0	0.3	100	10	0.080	0.011	0.078	0.011	0.296	0.292	0.083	0.005	0.083	0.005	0.018	0.032
0.0	−0.3	50	5	0.096	0.397	0.091	0.375	0.308	0.318	0.089	0.334	0.089	0.325	0.029	0.065
0.0	−0.3	50	10	0.088	0.558	0.079	0.525	0.406	0.491	0.083	0.558	0.076	0.538	0.027	0.062
0.0	−0.3	100	5	0.063	0.449	0.063	0.444	0.243	0.316	0.074	0.411	0.073	0.407	0.021	0.056
0.0	−0.3	100	10	0.074	0.599	0.072	0.595	0.363	0.505	0.052	0.619	0.054	0.612	0.022	0.050
0.3	0.0	50	5	0.118	0.015	0.109	0.014	0.239	0.168	0.107	0.015	0.103	0.014	0.042	0.052
0.3	0.0	50	10	0.137	0.007	0.127	0.003	0.361	0.262	0.116	0.005	0.115	0.004	0.038	0.041
0.3	0.0	100	5	0.086	0.007	0.087	0.007	0.217	0.186	0.080	0.003	0.076	0.002	0.025	0.031
0.3	0.0	100	10	0.086	0.001	0.081	0.001	0.281	0.246	0.083	0.003	0.083	0.002	0.018	0.032
−0.3	0.0	50	5	0.096	0.308	0.092	0.281	0.274	0.290	0.103	0.302	0.096	0.290	0.032	0.064
−0.3	0.0	50	10	0.094	0.452	0.076	0.417	0.452	0.493	0.089	0.441	0.085	0.420	0.027	0.069
−0.3	0.0	100	5	0.071	0.302	0.065	0.291	0.226	0.304	0.072	0.300	0.070	0.293	0.033	0.064

Notes: See Table 1 for an explanation of the various features of the table.

Table 3: Size-adjusted power in the level shift model when $\rho = 0.9$

| | | | | $\lambda = 0.3$ | | | | | | $\lambda = 0.7$ | | | | | |
γ	θ	T	N	$\widehat{Z}_t^c$	$\widehat{Z}_\phi^c$	Z_t^c	Z_ϕ^c	Z_t	Z_ϕ	$\widehat{Z}_t^c$	$\widehat{Z}_\phi^c$	Z_t^c	Z_ϕ^c	Z_t	Z_ϕ
0.0	0.0	50	5	0.172	0.167	0.166	0.190	0.114	0.109	0.143	0.181	0.140	0.198	0.099	0.137
0.0	0.0	50	10	0.231	0.276	0.240	0.302	0.120	0.165	0.299	0.366	0.292	0.366	0.175	0.172
0.0	0.0	100	5	0.568	0.674	0.618	0.719	0.375	0.442	0.589	0.696	0.602	0.738	0.323	0.457
0.0	0.0	100	10	0.859	0.932	0.866	0.947	0.538	0.654	0.830	0.942	0.838	0.946	0.485	0.605
0.0	0.3	50	5	0.156	0.186	0.140	0.194	0.140	0.148	0.161	0.202	0.156	0.196	0.130	0.185
0.0	0.3	50	10	0.234	0.304	0.239	0.291	0.153	0.198	0.271	0.373	0.270	0.385	0.174	0.223
0.0	0.3	100	5	0.517	0.666	0.564	0.707	0.355	0.528	0.566	0.709	0.566	0.729	0.434	0.503
0.0	0.3	100	10	0.846	0.936	0.864	0.948	0.717	0.853	0.846	0.955	0.834	0.956	0.631	0.761
0.0	-0.3	50	5	0.174	0.211	0.181	0.211	0.121	0.104	0.141	0.166	0.179	0.162	0.066	0.069
0.0	-0.3	50	10	0.203	0.323	0.212	0.332	0.133	0.132	0.245	0.267	0.264	0.313	0.085	0.083
0.0	-0.3	100	5	0.513	0.662	0.580	0.694	0.248	0.307	0.523	0.558	0.581	0.645	0.190	0.201
0.0	-0.3	100	10	0.813	0.892	0.836	0.915	0.343	0.440	0.773	0.884	0.813	0.908	0.269	0.388
0.3	0.0	50	5	0.178	0.202	0.167	0.184	0.123	0.169	0.180	0.246	0.170	0.231	0.164	0.161
0.3	0.0	50	10	0.254	0.328	0.246	0.358	0.179	0.202	0.233	0.331	0.242	0.360	0.151	0.227
0.3	0.0	100	5	0.476	0.648	0.491	0.633	0.409	0.491	0.517	0.729	0.536	0.728	0.363	0.527
0.3	0.0	100	10	0.834	0.953	0.827	0.955	0.646	0.796	0.851	0.939	0.860	0.943	0.751	0.847
-0.3	0.0	50	5	0.160	0.177	0.155	0.186	0.112	0.109	0.145	0.199	0.174	0.210	0.083	0.112
-0.3	0.0	50	10	0.265	0.284	0.266	0.311	0.125	0.121	0.208	0.275	0.238	0.300	0.105	0.119
-0.3	0.0	100	5	0.539	0.634	0.540	0.671	0.276	0.299	0.545	0.639	0.557	0.680	0.205	0.284
-0.3	0.0	100	10	0.844	0.896	0.853	0.927	0.401	0.463	0.809	0.903	0.867	0.929	0.322	0.409

Notes: See Table 1 for an explanation of the various features of the table.

Table 4: Size-adjusted power in the regime shift model when $\rho = 0.9$

γ	θ	T	N	$\lambda = 0.3$						$\lambda = 0.7$					
				$\widehat{Z}_t^r$	$\widehat{Z}_\phi^r$	Z_t^r	Z_ϕ^r	Z_t	Z_ϕ	$\widehat{Z}_t^r$	$\widehat{Z}_\phi^r$	Z_t^r	Z_ϕ^r	Z_t	Z_ϕ
0.0	0.0	50	5	0.138	0.209	0.165	0.224	0.059	0.063	0.145	0.180	0.142	0.192	0.055	0.054
0.0	0.0	50	10	0.283	0.305	0.288	0.294	0.078	0.068	0.280	0.396	0.289	0.387	0.052	0.053
0.0	0.0	100	5	0.500	0.658	0.515	0.656	0.069	0.058	0.550	0.711	0.550	0.707	0.061	0.056
0.0	0.0	100	10	0.829	0.893	0.844	0.899	0.087	0.092	0.828	0.926	0.828	0.923	0.063	0.061
0.0	0.3	50	5	0.144	0.224	0.143	0.223	0.062	0.068	0.177	0.238	0.181	0.240	0.065	0.064
0.0	0.3	50	10	0.242	0.318	0.249	0.332	0.107	0.118	0.233	0.312	0.237	0.314	0.038	0.046
0.0	0.3	100	5	0.554	0.675	0.556	0.679	0.089	0.084	0.616	0.758	0.620	0.767	0.081	0.095
0.0	0.3	100	10	0.846	0.936	0.844	0.943	0.148	0.145	0.821	0.925	0.828	0.923	0.076	0.080
0.0	−0.3	50	5	0.167	0.182	0.167	0.222	0.046	0.066	0.137	0.188	0.137	0.188	0.061	0.056
0.0	−0.3	50	10	0.245	0.329	0.258	0.345	0.063	0.075	0.205	0.259	0.227	0.265	0.052	0.057
0.0	−0.3	100	5	0.498	0.631	0.510	0.630	0.065	0.063	0.492	0.608	0.482	0.600	0.067	0.071
0.0	−0.3	100	10	0.773	0.888	0.777	0.885	0.090	0.062	0.805	0.916	0.804	0.911	0.038	0.060
0.3	0.0	50	5	0.155	0.206	0.146	0.224	0.083	0.088	0.179	0.258	0.178	0.265	0.062	0.059
0.3	0.0	50	10	0.237	0.361	0.255	0.359	0.085	0.089	0.244	0.367	0.278	0.362	0.057	0.065
0.3	0.0	100	5	0.545	0.707	0.550	0.702	0.097	0.126	0.595	0.706	0.610	0.698	0.066	0.072
0.3	0.0	100	10	0.856	0.970	0.859	0.970	0.143	0.161	0.802	0.926	0.822	0.928	0.079	0.082
−0.3	0.0	50	5	0.157	0.162	0.168	0.186	0.061	0.049	0.167	0.161	0.155	0.168	0.065	0.043
−0.3	0.0	50	10	0.214	0.255	0.232	0.272	0.058	0.038	0.211	0.266	0.222	0.269	0.060	0.058
−0.3	0.0	100	5	0.493	0.659	0.507	0.675	0.051	0.051	0.482	0.641	0.478	0.634	0.049	0.050
−0.3	0.0	100	10	0.819	0.887	0.820	0.894	0.059	0.068	0.763	0.900	0.770	0.907	0.050	0.044

Notes: See Table 1 for an explanation of the various features of the table.

Table 5: Size-adjusted power in the level shift model when $\rho = 0.8$

γ	θ	T	N	$\lambda = 0.3$						$\lambda = 0.7$					
				$\widehat{Z}_t^c$	$\widehat{Z}_\phi^c$	Z_t^c	Z_ϕ^c	Z_t	Z_ϕ	$\widehat{Z}_t^c$	$\widehat{Z}_\phi^c$	Z_t^c	Z_ϕ^c	Z_t	Z_ϕ
0.0	0.0	50	5	0.414	0.561	0.418	0.621	0.165	0.227	0.324	0.542	0.371	0.620	0.122	0.200
0.0	0.0	50	10	0.603	0.834	0.648	0.890	0.214	0.390	0.654	0.870	0.667	0.885	0.248	0.333
0.0	0.0	100	5	0.932	0.991	0.951	0.998	0.534	0.770	0.942	0.992	0.946	0.993	0.498	0.802
0.0	0.0	100	10	0.999	1.000	1.000	1.000	0.794	0.981	0.998	1.000	0.998	1.000	0.730	0.971
0.0	0.3	50	5	0.382	0.643	0.382	0.660	0.225	0.295	0.411	0.615	0.406	0.640	0.204	0.370
0.0	0.3	50	10	0.636	0.883	0.666	0.895	0.358	0.580	0.659	0.902	0.671	0.921	0.338	0.553
0.0	0.3	100	5	0.926	0.996	0.945	0.999	0.646	0.942	0.926	0.999	0.930	1.000	0.652	0.891
0.0	0.3	100	10	1.000	1.000	1.000	1.000	0.928	0.998	0.999	1.000	1.000	1.000	0.908	0.999
0.0	−0.3	50	5	0.349	0.524	0.369	0.556	0.095	0.101	0.336	0.461	0.376	0.489	0.088	0.131
0.0	−0.3	50	10	0.598	0.837	0.628	0.872	0.150	0.195	0.541	0.773	0.585	0.829	0.091	0.133
0.0	−0.3	100	5	0.876	0.984	0.915	0.986	0.253	0.504	0.874	0.966	0.908	0.978	0.223	0.365
0.0	−0.3	100	10	0.997	1.000	0.998	1.000	0.415	0.744	0.989	1.000	0.993	1.000	0.332	0.698
0.3	0.0	50	5	0.418	0.636	0.406	0.632	0.240	0.394	0.396	0.665	0.389	0.662	0.277	0.361
0.3	0.0	50	10	0.639	0.907	0.630	0.931	0.405	0.580	0.651	0.932	0.652	0.942	0.329	0.630
0.3	0.0	100	5	0.901	0.996	0.911	0.996	0.677	0.920	0.934	1.000	0.945	1.000	0.656	0.947
0.3	0.0	100	10	0.998	1.000	0.998	1.000	0.939	1.000	0.999	1.000	0.999	1.000	0.949	1.000
−0.3	0.0	50	5	0.342	0.473	0.368	0.520	0.116	0.146	0.340	0.557	0.379	0.606	0.093	0.150
−0.3	0.0	50	10	0.631	0.818	0.637	0.850	0.150	0.221	0.506	0.767	0.596	0.830	0.115	0.194
−0.3	0.0	100	5	0.890	0.975	0.893	0.983	0.277	0.489	0.886	0.982	0.895	0.987	0.259	0.482
−0.3	0.0	100	10	0.987	1.000	0.988	1.000	0.477	0.824	0.992	1.000	0.997	1.000	0.426	0.742

Notes: See Table 1 for an explanation of the various features of the table.

Table 6: Size-adjusted power in the regime shift model when $\rho = 0.8$

γ	θ	T	N	$\lambda = 0.3$						$\lambda = 0.7$					
				$\widehat{Z}_t^r$	$\widehat{Z}_\phi^r$	Z_t^r	Z_ϕ^r	Z_t	Z_ϕ	$\widehat{Z}_t^r$	$\widehat{Z}_\phi^r$	Z_t^r	Z_ϕ^r	Z_t	Z_ϕ
0.0	0.0	50	5	0.331	0.591	0.357	0.608	0.056	0.081	0.346	0.562	0.349	0.579	0.050	0.075
0.0	0.0	50	10	0.685	0.865	0.685	0.860	0.102	0.102	0.670	0.886	0.685	0.886	0.054	0.068
0.0	0.0	100	5	0.902	0.993	0.915	0.994	0.087	0.095	0.908	0.995	0.905	0.995	0.061	0.076
0.0	0.0	100	10	0.995	1.000	0.996	1.000	0.098	0.112	0.997	1.000	0.997	1.000	0.069	0.084
0.0	0.3	50	5	0.421	0.643	0.414	0.649	0.077	0.092	0.417	0.644	0.440	0.656	0.077	0.075
0.0	0.3	50	10	0.620	0.904	0.645	0.914	0.139	0.176	0.657	0.888	0.659	0.895	0.050	0.067
0.0	0.3	100	5	0.937	0.997	0.936	0.998	0.112	0.131	0.948	0.999	0.951	0.999	0.088	0.103
0.0	0.3	100	10	0.996	1.000	0.996	1.000	0.158	0.186	0.999	1.000	0.998	1.000	0.076	0.101
0.0	-0.3	50	5	0.365	0.501	0.357	0.554	0.053	0.069	0.335	0.531	0.316	0.526	0.077	0.084
0.0	-0.3	50	10	0.574	0.796	0.608	0.829	0.051	0.070	0.543	0.756	0.559	0.760	0.053	0.067
0.0	-0.3	100	5	0.829	0.967	0.838	0.967	0.071	0.089	0.815	0.969	0.814	0.968	0.060	0.067
0.0	-0.3	100	10	0.981	1.000	0.982	1.000	0.091	0.089	0.984	0.999	0.984	0.999	0.059	0.074
0.3	0.0	50	5	0.367	0.628	0.365	0.655	0.102	0.114	0.388	0.654	0.383	0.663	0.056	0.052
0.3	0.0	50	10	0.582	0.912	0.614	0.915	0.121	0.166	0.627	0.910	0.680	0.915	0.064	0.080
0.3	0.0	100	5	0.916	0.994	0.920	0.995	0.114	0.150	0.934	0.997	0.940	0.997	0.075	0.089
0.3	0.0	100	10	0.999	1.000	0.999	1.000	0.162	0.239	0.998	1.000	0.998	1.000	0.077	0.104
-0.3	0.0	50	5	0.341	0.457	0.339	0.480	0.058	0.079	0.332	0.481	0.317	0.481	0.069	0.069
-0.3	0.0	50	10	0.518	0.767	0.548	0.782	0.054	0.058	0.512	0.760	0.524	0.759	0.059	0.073
-0.3	0.0	100	5	0.872	0.988	0.881	0.989	0.050	0.070	0.840	0.976	0.835	0.977	0.043	0.049
-0.3	0.0	100	10	0.989	1.000	0.990	1.000	0.061	0.093	0.979	1.000	0.984	1.000	0.065	0.071

Notes: See Table 1 for an explanation of the various features of the table.

References

Ahn, S. K. (1993). Some Tests for Unit Roots in Autoregressive-Integrated-Moving Average Models with Deterministic Trends, *Biometrica*, Vol. 80, pp. 855-868.

Amsler, C., and J. Lee (1995). An LM Test for a Unit Root in the Presence of a Structural Break, *Econometric Theory*, Vol. 11, pp. 359-368.

Bai, J. (1997). Estimating Multiple Breaks One at a Time, *Econometric Theory*, Vol. 13, pp. 315-52.

Bai, J. (1994). Least Squares Estimation of a Shift in Linear Processes, *Journal of Time Series Analysis*, Vol. 15, pp. 453-472.

Bai, J., and P. Perron (1998). Estimating and Testing Linear Models with Multiple Structural Changes, *Econometrica*, Vol. 66, pp. 47-78.

Banerjee, A., M. Marcellino and C. Osbat (2004). Some Cautions on the use of Panel Methods for Integrated Series of Macroeconomic Data, *Econometrics Journal*, Vol. 7, pp. 322-340.

DeJong, D. N., J. C. Nankervis, N. E. Savin and C. H. Whiteman (1992). The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors, *Journal of Econometrics*, Vol. 53, pp. 323-343.

Dickey, D. A., and W. A. Fuller (1979). Distribution of the Estimator for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, Vol. 74, pp. 427-431.

Engle, R., and C. Granger (1987). Cointegration and Error Correction: Representation, Estimation and Testing, *Econometrica*, Vol. 55, pp. 251-276.

Gregory, A. M., and B. E. Hansen (1996). Residual-Based Tests for Cointegration in Models with Regime Shifts, *Journal of Econometrics*, Vol. 70, pp. 99-126.

Haug, A. A. (1996). Tests for Cointegration: A Monte Carlo Comparison, *Journal of Econometrics*, Vol. 71, pp. 89115.

Im, K. S., J. Lee, and M. Tieslau (2005). Panel LM Unit Root Tests with Level Shifts, *Oxford Bulletin of Economics and Statistics*, Vol. 67, pp. 393-419.

Newey, W., and K. West (1994). Autocovariance Lag Selection in Covariance Matrix Estimation, *Review of Economic Studies*, Vol. 61, pp. 613-653.

Perron, P. (1989). The Great Crash, the Oil Price Shock and the Unit Root Hypothesis, *Econometrica*, Vol. 57, pp. 1361-1401.

Phillips, P. C. B., and V. Solo (1992). Asymptotics for Linear Processes, *Annals of Statistics*, Vol. 20, pp. 971-1001.

Schmidt, P. and J. Lee (1991). A Modification to the Schmidts-Phillips Unit Root Test, *Economics Letters*, Vol. 36, pp. 285-289.

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*, Vol. 54, pp. 257-287.

Schwert, G. W. (1989). Tests for Unit Root: A Monte Carlo Investigation, *Journal of Business and Economic Statistics*, Vol. 7, pp. 147-159.

Zivot, E., and D. W. K. Andrews (1992). Further Evidence on the Great Crash, the Oil-Price Shock and the Unit-Root Hypothesis, *Journal of Business and Economics Statistics*, Vol. 10, pp. 251-270.

Westerlund, J. (2005a). Testing for Panel Cointegration with Multiple Structural Breaks. *Oxford Bulletin of Economics and Statistics*, forthcoming.

Westerlund, J. (2005b). New Simple Tests for Panel Cointegration, *Econometrics Reviews*, Vol. 24, pp. 297-316.

Westerlund, J. (2005c). Testing for Panel Cointegration with a Level Break, *Economics Letters*, forthcoming.