END-OF-SAMPLE COINTEGRATION BREAKDOWN TESTS

By

Donald W.K. Andrews and Jae-Young Kim

March 2003

COWLES FOUNDATION DISCUSSION PAPER NO. 1404

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

End-of-Sample Cointegration Breakdown Tests

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

Jae-Young Kim
Department of Economics
SUNY - Albany

June 2002
Revision: February 2003

Abstract

This paper introduces tests for cointegration breakdown that may occur over a relatively short time period, such as at the end of the sample. The breakdown may be due to a shift in the cointegrating vector or due to a shift in the errors from being $I(0)$ to being $I(1)$. Tests are introduced based on the post-breakdown sum of squared residuals and the post-breakdown sum of squared reverse partial sums of residuals. Critical values are provided using a parametric subsampling method.

The regressors in the model are taken to be arbitrary linear combinations of deterministic, stationary, and integrated random variables. The tests are asymptotically valid when the number of observations in the breakdown period, m, is fixed and finite as the total sample size, $T + m$, goes to infinity. The tests are asymptotically valid under weak conditions.

Simulation results indicate that the tests work well in the scenarios considered.

Use of the tests is illustrated by testing for interest rate parity breakdown during the Asian financial crisis of 1997.

Keywords: Cointegration, least squares estimator, model breakdown, parameter change test, structural change.

JEL Classification Numbers: C12, C52.

1 Introduction

This paper addresses the problem of cointegration breakdown over a short period of time. We are interested in breakdown due to a shift in the cointegrating parameter vector and/or a shift in the errors from being stationary to being integrated. The breakdown period may occur at the end of the sample, the beginning of the sample, or somewhere in between. For example, one might be interested in whether a recent event, such as the Asian currency crisis or a possible productivity slowdown, has caused an end-of-sample breakdown in a cointegrating relationship. Alternatively, one might be interested in whether some short policy regime shift or a war caused a middle-of-sample cointegration breakdown.

Tests in the literature for cointegration breakdown assume that the post-breakdown period is relatively long. These tests rely on asymptotics in which its length goes to infinity with the sample size. Examples include the tests of Hansen (1992), Quintos and Phillips (1993), Quintos (1997), and Kim (1999, 2000). Such tests are not appropriate for the case considered here in which the post-breakdown period is relatively short.

In this paper, we introduce tests for cointegration breakdown that are asymptotically valid when the length, m, of the post-breakdown period is fixed as the total sample size, $T+m$, goes to infinity. The tests rely on a subsampling-like method of computing critical values that is introduced in Andrews (2002) and described below. The critical values are easy to compute.

For simplicity, in the remainder of this section and in the bulk of the paper, we discuss tests for end-of-sample cointegration breakdown. Adjustment of the end-of-sample tests for breakdown occurring at the beginning or in the middle of the sample is straightforward.

The first test statistic that we consider, P_a, is the sum of squared post-break residuals evaluated at a pre-break estimator, such as the least squares (LS) estimator. This test statistic is motivated by the F statistic for parameter change over a short period in a regression model with iid normal errors and strictly exogenous regressors.

Simulations indicate that the P_a test over-rejects the null hypothesis somewhat when the null is true in many cases. In consequence, we consider finite sample adjustments to the test statistic and critical values that yield better finite sample size. This leads to two tests called the P_b and P_c tests.

Next, we consider the locally best invariant (LBI) test for a shift in the error distribution from being iid normal for all observations to being iid normal for the first T observations and then a normal unit root process for the last m observations. The resulting test statistic is given by the sum of squared reverse partial sums of the post-break residuals. That is, the statistic is of the form

$$\sum_{t=T+1}^{T+m}\left(\sum_{s=t}^{T+m}\widehat{u}_s\right)^2, \tag{1.1}$$

where $\widehat{u}_s$ is a residual. The form of this statistic is similar to tests considered in Nyblom and Makelainen (1983), King and Hillier (1985), Nyblom (1986, 1989), Nabeya and Tanaka (1988), Leybourne and McCabe (1989), Kwiatkowski, Phillips, Schmidt, and Shin (1992), Tanaka (1993), and Shin (1994). We introduce three tests, R_a, R_b, and R_c, that are analogous to the tests P_a, P_b, and P_c, but rely on a different quadratic function of the post-break residuals, viz., that given in (1.1).

Critical values for all the tests considered are obtained by a subsampling-like method that we call *parametric sub-sampling.* One computes the $T - m + 1$ test statistics that are analogous to the test statistic of interest but are for testing for cointegration breakdown over the m observations that start at the j-th observation, rather than for breakdown starting at the $(T+1)$-th observation, for $j = 1, ..., T - m + 1$. The $1 - \alpha$ sample quantile of these statistics is the significance level α critical value for the end-of-sample breakdown test statistic. Computation of the critical value is relatively easy. It just requires calculation of $T - m + 1$ versions of the original statistic. p-values are also obtained easily using this method.

The parametric subsampling critical values use subsamples of length m, the number of post-breakdown observations. There is no arbitrary smoothing parameter or block length parameter to select. No heteroskedasticity and autocorrelation consistent covariance matrix estimator is required. These critical values are not pure subsampling critical values because the test statistic for a given value of j depends on observations other than those indexed by $j, ..., j+m-1$ through the parameter estimator that is used to compute the residuals. See Politis, Romano, and Wolf (1999) regarding pure subsampling methods.

Parametric subsampling is used in Andrews (2002) to obtain critical values for tests of parameter instability over short time periods in models with stationary observations. Both linear and nonlinear models are considered. In contrast, this paper considers linear models only, but allows for nonstationary regressors and, hence, cointegrating regression models. The test statistics considered in the two papers also differ.

The tests considered here are not consistent tests because m is fixed as $T \to \infty$. Typically, however, they are asymptotically unbiased. The power of the tests depends on the magnitude of the breakdown, such as the magnitude of the parameter shift and/or the magnitude of the unit root error variance, relative to the pre-breakdown error variance. Power also depends on m. The larger is m, the greater is the power everything else being equal. Power may be low if m is small or the magnitude of the breakdown is not large. In consequence, failure to reject the null hypothesis should not be interpreted as strong evidence in favor of stable cointegration.

The paper presents some Monte Carlo simulations that are designed to assess the finite sample size and size-corrected power properties of the tests P_a-P_c and R_a-R_c. We consider models with a constant, time trend or no time trend, two or four unit root regressors, and zero or two stationary regressors. The errors, unit root regressor differences, and stationary regressors are first-order autoregressive (AR) with the same AR parameter. The AR parameters considered are $\rho = 0$, .4, and .8. The AR innovations considered are normal, chi-square with two degrees of freedom, t_3, and uniform. The unit root regressor differences are correlated with the errors in some cases considered. The pre-breakdown sample sizes are $T = 100$ and 250 and the post-breakdown sample sizes are $m = 10$, 5, and 1. We consider power against shifts in the cointegrating regressor vector, as well as in shifts in the error from being $I(0)$ to being $I(1)$. These are referred to as parameter shift alternatives and unit root alternatives, respectively.

The simulation results show that the P_a and R_a tests tend to over-reject in finite samples. The sizes of the P_b, P_c, R_b, and R_c tests are noticeably better and are quite good, especially considering that the range of cases considered is wide. For example, for the nominal 5% P_c test, the null rejection rate varies between .040 and .064 over 72 different model/parameter combinations. For $T = 100$, it varies between .028 and .081 with an average of .052.

The power results indicate that the P and R tests have power against both unit root alternatives and parameter shift alternatives. In fact, paradoxically, the P tests are somewhat better than the R tests for unit root alternatives and vice versa for parameter shift alternatives. The differences between the powers of the P_b, P_c, R_b, and R_c tests typically are not large. The best of these tests in terms of power is the R_c test because it is slightly more powerful than the R_b tests across most cases and has less variable power across different distributions than the P_b and P_c tests.

Combining the simulation results for size and power, we find the best two tests are the P_c and R_c tests. We have a slight preference for the P_c test because its size properties are somewhat better than those of the R_c test. The P_c test has pretty good size and power properties across the wide range of models and parameter combinations that are considered in the simulations.

The use of the P and R tests is illustrated by testing for breakdown of interest rate parity during the Asian financial crisis of 1997. Separate results are given for Thailand, the Philippines, Indonesia, and Singapore. The results indicate that the end-of-sample tests would have detected cointegration breakdown when the sample period considered is such that breakdown occurred at the end of the sample. (In this illustration, the "true" time of breakdown is taken to be the time estimated using a long time series, which includes the time series upon which the tests are constructed, and employing the method of Kim (2000).)

The remainder of this paper is organized as follows. All sections of the paper except Section 5 discuss *end-of-sample* cointegration breakdown tests. Section 2 introduces the model and hypotheses of interest. Section 3 presents the tests that are considered. Section 4 states high-level assumptions, provides sufficient conditions for these assumptions for the case of estimation by LS, and states the main asymptotic results. Section 5 discusses tests for cointegration breakdown that occurs at the beginning or in the middle of the sample. Section 6 provides some Monte Carlo results. Section 7 provides the empirical example. An Appendix contains proofs.

2 Model and Hypotheses

The model is

$$y_t = \begin{cases} x_t'\beta_0 + u_t & \text{for } t = 1, ..., T \\ x_t'\beta_t + u_t & \text{for } t = T+1, ..., T+m, \end{cases} \tag{2.1}$$

where $y_t, u_t \in R$ and $x_t, \beta_0, \beta_t \in R^k$. Under the maintained hypothesis, the errors for the first T time periods, $\{u_t : t = 1, ..., T\}$, are mean zero, stationary, and ergodic. In addition, under the maintained hypothesis, the regressors for all time periods, $\{x_t : t = 1, ..., T+m\}$, are linear combinations of unit root (I(1)) random variables, stationary random variables, and deterministic variables, such as a constant and a linear time trend. Precise assumptions are given in Section 4.

The null and alternative hypotheses are

$$H_0 : \begin{cases} \beta_t = \beta_0 \text{ for all } t = T+1, ..., T+m \text{ and} \\ \{u_t : t = 1, ..., T+m\} \text{ are stationary and ergodic} \end{cases}$$
$$H_1 : \begin{cases} \beta_t \neq \beta_0 \text{ for some } t = T+1, ..., T+m \text{ and/or} \\ \text{the distribution of } \{u_{T+1}, ..., u_{T+m}\} \text{ differs from} \\ \text{the distribution of } \{u_1, ..., u_m\}. \end{cases} \quad (2.2)$$

Under the null hypothesis, the model is a well-specified cointegrating regression model for all $t = 1, ..., T+m$. Under the alternative hypothesis, the model is a well-specified cointegrating regression model for all $t = 1, ..., T$, but for $t = T+1, ..., T+m$ the cointegrating relationship breaks down.

The breakdown may be due to (i) a shift in the cointegrating vector from β_0 to β_t, (ii) a shift in the distribution of u_t from being stationary to being a unit root random variable, (iii) some other shift in the distribution of $\{u_{T+1}, ..., u_{T+m}\}$ from that of $\{u_1, ..., u_m\}$, or (iv) some combination of the previous shifts. In the next few sections, we introduce tests that are designed especially for cases (i) and (ii).

3 Cointegration Breakdown Tests

3.1 $\mathbf{P}_a$ Test

First , we consider a test statistic that is a quadratic form in the "post-breakdown" residuals $\{\widehat{u}_t : t = T+1, ..., T+m\}$. The test rejects the null hypothesis if the test statistic exceeds a critical value that is determined using a *parametric subsampling* method.

For any $1 \leq r \leq s \leq T+m$, let

$$\begin{aligned} \mathbf{Y}_{r\text{-}s} &= (y_r, ..., y_s)', \\ \mathbf{X}_{r\text{-}s} &= (x_r, ..., x_s)', \text{ and} \\ \mathbf{U}_{r\text{-}s} &= (u_r, ..., u_s)'. \end{aligned} \quad (3.3)$$

Let

$$\begin{aligned} P_j(\beta, \Omega) &= (\mathbf{Y}_{j\text{-}(j+m-1)} - \mathbf{X}_{j\text{-}(j+m-1)}\beta)'\Omega(\mathbf{Y}_{j\text{-}(j+m-1)} - \mathbf{X}_{j\text{-}(j+m-1)}\beta) \text{ and} \\ P_j(\beta) &= P_j(\beta, I_m). \end{aligned} \quad (3.4)$$

for $j = 1, ..., T+1$, where Ω is some nonsingular $m \times m$ matrix and I_m denotes the m dimensional identity matrix.

Let $\widehat{\beta}_{r\text{-}s}$denote an estimator of β_0 based on the observations $t = r, ..., s$ for $1 \leq r < s \leq T + m$. For example, for the LS estimator,

$$\widehat{\beta}_{r\text{-}s} = (\mathbf{X}'_{r\text{-}s}\mathbf{X}_{r\text{-}s})^{-1}\mathbf{X}'_{r\text{-}s}\mathbf{Y}_{r\text{-}s} \tag{3.5}$$

(provided $\mathbf{X}'_{r\text{-}s}\mathbf{X}_{r\text{-}s}$ is nonsingular). Other estimators can also be considered. Such estimators include the fully modified estimator of Phillips and Hansen (1990), the ML estimator, see Johansen (1988, 1991), Ahn and Reinsel (1990), and Phillips (1991), and the asymptotically efficient estimators of Phillips and Loretan (1991), Saikonen (1991), Park (1992), and Stock and Watson (1993).

The first test statistic, P_a, that we consider is defined by

$$\begin{aligned} P_a &= P_{T+1}(\widehat{\beta}_{1\text{-}T}) \\ &= \sum_{t=T+1}^{T+m} (y_t - x_t'\widehat{\beta}_{1\text{-}T})^2. \end{aligned} \tag{3.6}$$

As defined, P_a is the post-breakdown sum of squared residuals. The statistic P_a is often referred to as a *predictive* statistic. The motivation for considering this statistic is that in a linear regression model with known error variance it is (proportional to) the F statistic for testing for a one time change in the regression parameter occurring at time $T+1$ when $m \leq k$, e.g., see Chow (1960). The F test has well-known optimal power properties in the (restricted) context in which the errors are iid normal and the regressors are strictly exogenous, e.g., see Scheffé (1959, Ch. 2).[2] Predictive statistics have been used by Dufour, Ghysels, and Hall (1994) and Andrews (2002) to test for end-of-sample instability in models with stationary observations.

Under the null hypothesis, the distribution of $P_{T+1}(\beta_0)$ is the same as that of $P_j(\beta_0)$ for all $j \geq 1$, because $P_j(\beta_0) = \sum_{t=j}^{j+m-1} u_t^2$ and $\{u_t : t \geq 1\}$ are stationary. The estimator $\widehat{\beta}_{1\text{-}T}$, which appears in the statistic P_a, converges in probability to the true parameter, β_0, under suitable assumptions. Hence, the asymptotic null distribution of P_a is the distribution of $P_1(\beta_0)$. This is established rigorously below.

The random variables $\{P_j(\beta_0) : j = 1, ..., T - m + 1\}$ are stationary and ergodic under H_0 and H_1. In consequence, the empirical distribution function (df) of $\{P_j(\beta_0) : j = 1, ..., T - m + 1\}$ is a consistent estimator of the df of $P_1(\beta_0)$. Hence, we can consistently estimate the df of $P_1(\beta_0)$ by using the empirical df of $\{P_j(\beta) : j = 1, ..., T - m + 1\}$ evaluated at a consistent estimator of β_0 (see Theorem 1 below).

The estimator $\widehat{\beta}_{1\text{-}T}$, which appears in the statistic P_a, does not depend on the observations indexed by $t = T + 1, ..., T + m$ that appear in $P_{T+1}(\beta)$. To mirror this property in the subsample statistics, we evaluate $P_j(\beta)$ at a "leave-m-out" estimator, $\widehat{\beta}_{(j)}$, that is analogous to $\widehat{\beta}_{1\text{-}T}$ but does not depend on the observations that appear in $P_j(\beta)$. By definition, for $j = 1, ..., T - m + 1$,

$$\begin{aligned} \widehat{\beta}_{(j)} = {} & \text{estimator of } \beta \text{ using observations indexed by } t = 1, ..., T \text{ with} \\ & t \neq j, ..., j + m - 1. \end{aligned} \tag{3.7}$$

For the types of estimators mentioned above, the estimator $\widehat{\beta}_{(j)}$ is consistent for β_0 (uniformly over j) under suitable assumptions.

Define

$$P_{a,j} = P_j(\widehat{\beta}_{(j)}) \text{ for } j = 1, ..., T - m + 1. \tag{3.8}$$

The empirical df of $\{P_{a,j} : j = 1, ..., T - m + 1\}$ is

$$\widehat{F}_{P_a,T}(x) = \frac{1}{T - m + 1} \sum_{t=1}^{T-m+1} 1(P_{a,j} \leq x). \tag{3.9}$$

This empirical distribution converges in probability (and almost surely) to the df of $P_1(\beta_0)$ (under suitable assumptions). In consequence, to obtain a test with asymptotic significance level α, we take the critical value for the test statistic P_a to be the $1 - \alpha$ sample quantile, $\widehat{q}_{P_a,1-\alpha}$, of $\{P_{a,j} : j = 1, ..., T - m + 1\}$. By definition,

$$\widehat{q}_{P_a,1-\alpha} = \inf\{x \in R : \widehat{F}_{P_a,T}(x) \geq 1 - \alpha\}. \tag{3.10}$$

One rejects H_0 if $P_a > \widehat{q}_{P_a,1-\alpha}$. Equivalently, one rejects H_0 if P_a exceeds $100(1-\alpha)\%$ of the values $\{P_{a,j} : j = 1, ..., T - m + 1\}$—that is, if

$$(T - m + 1)^{-1} \sum_{j=1}^{T-m+1} 1(P_a > P_{a,j}) \geq 1 - \alpha. \tag{3.11}$$

The p-value for the P_a test is

$$pv_{P_a} = (T - m + 1)^{-1} \sum_{j=1}^{T-m+1} 1(P_a \leq P_{a,j}). \tag{3.12}$$

3.2 $\mathbf{P}_b$ and $\mathbf{P}_c$ Tests

Simulations indicate that the P_a test over-rejects the null hypothesis in many scenarios. In consequence, we consider two variants of the P_a test that are designed to have better finite-sample properties.

We define the P_b and $P_{b,j}$ statistics as follows:

$$\begin{aligned} P_b &= P_{T+1}(\widehat{\beta}_{1\text{-}(T+\lceil m/2 \rceil)}) \text{ and} \\ P_{b,j} &= P_j(\widehat{\beta}_{(j)}) \text{ for } j = 1, ..., T - m + 1, \end{aligned} \tag{3.13}$$

where $\lceil m/2 \rceil$ denotes the smallest integer that is greater than or equal to $m/2$. The estimator $\widehat{\beta}_{1\text{-}(T+\lceil m/2 \rceil)}$ uses the observations $t = 1, ..., T + \lceil m/2 \rceil$. Critical values and p-values for P_b are obtained using $\{P_{b,j} : j = 1, ..., T - m + 1\}$ as in (3.10)-(3.12) with a replaced by b.

The motivation for the P_b test is as follows. The P_b statistic is somewhat less variable than the P_a statistic because the estimator $\widehat{\beta}_{1\text{-}(T+\lceil m/2 \rceil)}$ depends on the observations indexed by $t = T + 1, ..., T + \lceil m/2 \rceil$ and, hence, the residuals indexed by $t = T+1, ..., T + \lceil m/2 \rceil$ (upon which $P_{T+1}(\cdot)$ depends) are less variable when computed using $\widehat{\beta}_{1\text{-}(T+\lceil m/2 \rceil)}$ than when computed using $\widehat{\beta}_{1\text{-}T}$. We base the critical values for P_b on the same statistics $P_j(\widehat{\beta}_{(j)})$ as for the P_a statistic. In consequence, the test P_b rejects less frequently under H_0 than the P_a test does.

Next, we consider a statistic P_c that depends on the complete sample estimator $\widehat{\beta}_{1\text{-}(T+m)}$:

$$P_c = P_{T+1}(\widehat{\beta}_{1\text{-}(T+m)}). \tag{3.14}$$

This statistic is less variable than either P_a or P_b because $\widehat{\beta}_{1\text{-}(T+m)}$ makes use of the observations $t = T+1, ..., T+m$ and, hence, the residuals for these time periods are less variable when computed using $\widehat{\beta}_{1\text{-}(T+m)}$ than when computed using $\widehat{\beta}_{1\text{-}T}$ or $\widehat{\beta}_{1\text{-}(T+\lceil m/2 \rceil)}$. Simulations show that the test based on the statistic P_c and the subsample statistics $\{P_j(\widehat{\beta}_{(j)}) : j = 1, ..., T-m+1\}$ tends to under-reject the null hypothesis in a broad array of cases. Hence, we introduce subsample statistics for use with P_c that are somewhat more variable than $\{P_j(\widehat{\beta}_{(j)}) : j = 1, ..., T-m+1\}$.

We define the "leave-$m/2$-out" estimator, $\widehat{\beta}_{2(j)}$, by

$$\begin{aligned} \widehat{\beta}_{2(j)} = & \text{ estimator of } \beta \text{ using observations indexed by } t = 1, ..., T \text{ with} \\ & t \neq j, ..., j + \lceil m/2 \rceil - 1 \end{aligned} \tag{3.15}$$

for $j = 1, ..., T-m+1$. By definition, $\widehat{\beta}_{2(j)}$ is analogous to $\widehat{\beta}_{r\text{-}s}$, but is based on the observations indicated above.

We define the $P_{c,j}$ statistics as follows:

$$P_{c,j} = P_j(\widehat{\beta}_{2(j)}) \text{ for } j = 1, ..., T-m+1. \tag{3.16}$$

Critical values and p-values for P_c are obtained using $\{P_{c,j} : j = 1, ..., T-m+1\}$ as in (3.10)-(3.12) with a replaced by c. The test based on P_c and $P_{c,j}$ rejects noticeably less frequently under the null hypothesis than the P_a test and somewhat less frequently than the P_b test, see the simulation results of Section 6.

3.3 Tests with Estimated Weight Matrix

The P_a, P_b, and P_c tests are designed for the case where the errors in the regression model are uncorrelated—although the tests have correct size asymptotically whether or not the errors are correlated. If the errors are correlated, it might be advantageous in terms of power to include weights in the statistics based on an estimator of the error covariance matrix. We considered some tests that do so, but found that they were somewhat inferior to the P_a-P_c tests in terms of closeness of nominal and true size and in terms of size-corrected power across the range of models considered in Section 6.

We state the definition of these tests here but do not discuss them further. These tests are the same as P_a-P_c except that $\Omega = I_m$ is replaced by

$$\widehat{\Omega}_{1\text{-}(T+m)} = \left(\frac{1}{T+1} \sum_{j=1}^{T+1} \widehat{\mathbf{U}}_{j,j+m-1} \widehat{\mathbf{U}}'_{j,j+m-1} \right)^{-1}, \text{ where}$$

$$\widehat{\mathbf{U}}_{j,j+m-1} = \mathbf{Y}_{j,j+m-1} - \mathbf{X}_{j,j+m-1}\widehat{\beta}_{1\text{-}(T+m)}, \tag{3.17}$$

The estimator $\widehat{\Omega}_{1\text{-}(T+m)}$ is an estimator of the inverse of the $m \times m$ covariance matrix of the errors $\Omega_0^{-1} = (E\mathbf{U}_{1\text{-}m}\mathbf{U}_{1\text{-}m}')^{-1}$.

3.4 Locally Best Invariant Test for Unit Root Alternatives

The P_v tests for $v = a$, b, and c are motivated by the F test for a one time change in the parameter vector β. We now consider the locally best invariant (LBI) test statistic for the presence of unit root errors from $t = T+1$ to $t = T+m$ in a linear regression model with iid normal errors, known error variance (under the null), and exogenous regressors. We use the form of this statistic to construct tests that are asymptotically valid under more general conditions on the errors and regressors.

The model and LBI statistic that we consider is similar to those considered in the papers listed in the Introduction.

For the purposes of generating the LBI test statistic, the model we consider is

$$\begin{aligned} y_t &= x_t'\beta_0 + u_t \text{ for } t = 1, ..., T+m, \\ u_t &= \psi_t + \lambda^{1/2}\widetilde{\psi}_t, \\ \psi_t &\sim \text{iid } N(0,1) \text{ for } t = 1, ..., T+m, \\ \widetilde{\psi}_t &= \begin{cases} 0 & \text{for } t = 1, ..., T \\ \widetilde{\psi}_{t-1} + \varepsilon_t & \text{for } t = T+1, ..., T+m, \text{ and} \end{cases} \\ \varepsilon_t &\sim \text{iid } N(0,1) \text{ for } t = T+1, ..., T+m, \end{aligned} \tag{3.18}$$

where ε_{t_1}, ψ_{t_2}, and x_{t_3} are independent of each other for all t_1, t_2, and t_3. The null and alternative hypotheses of interest are

$$H_0 : \lambda = 0 \text{ and } H_1 : \lambda > 0. \tag{3.19}$$

When the regressors are integrated, the null hypothesis consists of cointegration for the whole sample, whereas the alternative hypothesis consists of cointegration for the observations $t = 1, ..., T$ and lack of cointegration (i.e., spurious regression) for the observations $t = T+1, ..., T+m$.

Conditional on $\{x_t : t = 1, ..., T+m\}$, we have

$$\begin{aligned} \mathbf{Y}_{1\text{-}(T+m)} &\sim N(\mathbf{X}_{1\text{-}(T+m)}\beta_0,\ I_{T+m} + \lambda V), \text{ where} \\ V &= \text{Diag}\{\mathbf{0}_T, A_m\}, \\ [A_m]_{k,\ell} &= \min\{k, \ell\} \text{ for } k, \ell = 1, ..., m, \end{aligned} \tag{3.20}$$

and $\mathbf{0}_T$ is a $T \times T$ matrix of zeros. That is, V is a $(T+m) \times (T+m)$ matrix consisting of zeros except in the lower diagonal $m \times m$ block, which is given by the $m \times m$ matrix A_m.

We consider invariance with respect to the following standard transformations in a linear model:

$$\begin{aligned} \mathbf{Y}_{1\text{-}(T+m)} &\rightarrow \mathbf{Y}_{1\text{-}(T+m)} + \mathbf{X}_{1\text{-}(T+m)}\gamma, \\ \beta_0 &\rightarrow \beta_0 + \gamma. \end{aligned} \tag{3.21}$$

The maximal invariant statistic S for these transformations is defined as follows. Let J be a $(T+m) \times (T+m-k)$ matrix that satisfies $J'J = I_{T+m-k}$ and $JJ' = I_{T+m} - X(X'X)^{-1}X'$, where $X = \mathbf{X}_{1\text{-}(T+m)}$. We have

$$S = J'\mathbf{Y}_{1\text{-}(T+m)} \sim N(0,\ I_{T+m-k} + \lambda J'VJ). \tag{3.22}$$

By Ferguson (1967, p. 235), the rejection region of the LBI test is

$$\left.\frac{d}{d\lambda}\log f_{T+m}(S|\lambda)\right|_{\lambda=0} >\ K, \tag{3.23}$$

where $f_{T+m}(S|\lambda)$ is the density of S evaluated at S and K is a constant. In the present case,

$$\begin{aligned}
2\frac{d}{d\lambda}\log f_{T+m}(S|\lambda) &= -S'\frac{d}{d\lambda}(I_{T+m-k} + \lambda J'VJ)^{-1}S \\
&= S'(I_{T+m-k} + \lambda J'VJ)^{-1}J'VJ(I_{T+m-k} + \lambda J'VJ)^{-1}S \text{ and} \\
\left.2\frac{d}{d\lambda}\log f_{T+m}(S|\lambda)\right|_{\lambda=0} &= S'J'VJS \\
&= \widehat{\mathbf{U}}'_{1\text{-}(T+m)}V\widehat{\mathbf{U}}_{1\text{-}(T+m)} \\
&= \widehat{\mathbf{U}}'_{(T+1)\text{-}(T+m)}A_m\widehat{\mathbf{U}}_{(T+1)\text{-}(T+m)},
\end{aligned} \tag{3.24}$$

where

$$\widehat{\mathbf{U}}_{1\text{-}(T+m)} = JS = JJ'\mathbf{Y}_{1\text{-}(T+m)} = \mathbf{Y}_{1\text{-}(T+m)} - \mathbf{X}_{1\text{-}(T+m)}\widehat{\beta}_{LS,1\text{-}(T+m)} \tag{3.25}$$

and $\widehat{\beta}_{LS,1\text{-}(T+m)}$ is the LS estimator from the regression of $\mathbf{Y}_{1\text{-}(T+m)}$ on $\mathbf{X}_{1\text{-}(T+m)}$. Hence, the LBI test statistic is a quadratic form in the post-change residual vector with weight matrix A_m.

3.5 $\mathbf{R}_a$-$\mathbf{R}_c$ Tests

The LBI test statistic of (3.24) can be written using (3.4) as

$$P_{T+1}(\widehat{\beta}_{LS,1\text{-}(T+m)}, A_m). \tag{3.26}$$

That is, the LBI test statistic is just like the P_a-P_c statistics except that it uses the weight matrix A_m instead of the identity matrix. In this section, we define three tests R_a-R_c that are analogous to the P_a-P_c tests defined above, but use the weight matrix A_m instead of I_m.

Define

$$\begin{aligned}
R_a &= P_{T+1}(\widehat{\beta}_{1\text{-}T}, A_m),\ R_{a,j} = P_{T+1}(\widehat{\beta}_{(j)}, A_m), \\
R_b &= P_{T+1}(\widehat{\beta}_{1\text{-}(T+\lceil m/2\rceil)}, A_m),\ R_{b,j} = P_{T+1}(\widehat{\beta}_{(j)}, A_m), \text{ and} \\
R_c &= P_{T+1}(\widehat{\beta}_{1\text{-}(T+m)}, A_m),\ R_{c,j} = P_{T+1}(\widehat{\beta}_{2(j)}, A_m).
\end{aligned} \tag{3.27}$$

The estimators $\widehat{\beta}_{(j)}$ and $\widehat{\beta}_{2(j)}$ used in the sub-sample statistics $R_{a,j}$, $R_{b,j}$, and $R_{c,j}$ are chosen for the same reasons as for the P_a-P_c tests. Critical values and p-values for the R_a-R_c tests are obtained as in (3.10)-(3.12) with P replaced by R and with a equal to a, b, or c. The estimator $\widehat{\beta}_{r\text{-}s}$ used with the R_a-R_c tests could be the LS estimator or some other estimator.

It turns out that the R_v test statistic for $v = a$, b, or c is a sum of squares of reverse partial sums of residuals. To see this, let Q be the $m \times m$ matrix that has ones on and above the main diagonal and zeros below the main diagonal. Then, $A_m = Q'Q$ and R_a can be written as

$$R_a = P_{T+1}(\widehat{\beta}_{1\text{-}T}, A_m) = (Q\widehat{\mathbf{U}}(\widehat{\beta}_{1\text{-}T}))'Q\widehat{\mathbf{U}}(\widehat{\beta}_{1\text{-}T}) = \sum_{t=T+1}^{T+m} \left(\sum_{s=t}^{T+m} (y_s - x_s'\widehat{\beta}_{1\text{-}T}) \right)^2. \tag{3.28}$$

The statistics $R_{a,j}$, R_b, $R_{b,j}$, etc. can be written in the same way with $\widehat{\beta}_{1\text{-}T}$ replaced by the appropriate estimator.

As shown below, the R_a-R_c tests are asymptotically valid in a much broader class of models than the model of (3.18) that generates the LBI test.

4 Asymptotic Results

4.1 Assumptions

To simplify the theoretical analysis of the tests introduced above, we consider a transformation of the regressor vector, x_t, that separates the unit root and deterministic components of x_t from its stationary components. This transformation need not be known by the user of the tests. It is employed only in the theoretical analysis of the tests. Let

$$\begin{aligned}
z_t &= H'x_t = \begin{pmatrix} z_{1,t} \\ z_{2,t} \end{pmatrix}, \ \gamma_0 = H^{-1}\beta_0 = \begin{pmatrix} \gamma_{1,0} \\ \gamma_{2,0} \end{pmatrix}, \\
\gamma_t &= H^{-1}\beta_t = \begin{pmatrix} \gamma_{1,t} \\ \gamma_{2,t} \end{pmatrix}, \ \widehat{\gamma}_{r\text{-}s} = H^{-1}\widehat{\beta}_{r\text{-}s} = \begin{pmatrix} \widehat{\gamma}_{1,r\text{-}s} \\ \widehat{\gamma}_{2,r\text{-}s} \end{pmatrix}, \\
\widehat{\gamma}_{(j)} &= H^{-1}\widehat{\beta}_{(j)} = \begin{pmatrix} \widehat{\gamma}_{1,(j)} \\ \widehat{\gamma}_{2,(j)} \end{pmatrix}, \text{ and } \widehat{\gamma}_{2(j)} = H^{-1}\widehat{\beta}_{2(j)} = \begin{pmatrix} \widehat{\gamma}_{1,2(j)} \\ \widehat{\gamma}_{2,2(j)} \end{pmatrix},
\end{aligned} \tag{4.1}$$

where H is a non-random nonsingular $k \times k$ matrix, $z_{\ell,t}, \gamma_{\ell,0}, \gamma_{\ell,t}, \widehat{\gamma}_{\ell,r\text{-}s}, \widehat{\gamma}_{\ell,(j)}, \widehat{\gamma}_{\ell,2(j)} \in R^{k_\ell}$ for $\ell = 1, 2$, and $k = k_1 + k_2$. We assume that H is chosen such that the transformed regressor vector $z_{1,t}$ contains only unit root and/or deterministic variables and the transformed regressor vector $z_{2,t}$ contains only stationary mean zero random variables.

The model can be rewritten as

$$y_t = \begin{cases} z_t'\gamma_0 + u_t & \text{for } t = 1, ..., T \\ z_t'\gamma_t + u_t & \text{for } t = T+1, ..., T+m. \end{cases} \tag{4.2}$$

Let w_t denote the vector of errors and stationary regressors:

$$w_t = \left(\begin{array}{c} u_t \\ z_{2,t} \end{array} \right). \tag{4.3}$$

In order to determine the behavior of the random critical values defined above under both H_0 and H_1, it is convenient to consider a sequence of random variables $\{w_{0,t} : t \geq 1\}$ that are stationary and ergodic under both H_0 and H_1. Under H_0, w_t equals $w_{0,t}$ for $t = 1, ..., T+m$. Under H_1, $w_t = w_{0,t}$ for $t = 1, ..., T$ and $w_t = w_{T,t}$ for $t = T+1, ..., T+m$, where $\{w_{T,t} : t = T+1, ..., T+m\}$ are some random variables whose joint distribution may differ from that of $\{w_{0,t} : t = T+1, ..., T+m\}$. We assume that the distribution under H_1 of $\{w_{T,t} : t = T+1, ..., T+m\}$ is independent of T. That is, we consider fixed, not local, alternatives. Note that the variables $\{(y_t, w_t, z_{1,t}) : t = 1, ..., T+m\}$ are from a triangular array under H_1, rather than a sequence, because the breakdown point T changes as $T \rightarrow \infty$.

We make the following assumptions.

Assumption 1. $\{w_{0,t} : t \geq 1\}$ are mean zero, stationary, and ergodic random vectors under H_0 and H_1. The distribution of $\{z_{1,t} : t = 1, ..., T\}$ is the same under H_0 and H_1. Under H_1, the distribution of $\{w_{T,t} : t = T+1, ..., T+m\}$ does not depend on T.

Assumption 2. $E|u_t| < \infty$, $E||u_t z_{2,t}|| < \infty$, and $E||z_{2,t}||^2 < \infty$ for $t \leq T$.

Assumption 3. $\max_{t \leq T+m} ||B_T^{-1} z_{1,t}|| = O_p(1)$ for some non-random positive-definite diagonal $k_1 \times k_1$ matrices $\{B_T : T \geq 1\}$ under H_0 and H_1.

Assumption 4. When $v = a$, $||B_T(\widehat{\gamma}_{1,1\text{-}T} - \gamma_{1,0})|| \rightarrow_p 0$, $||\widehat{\gamma}_{2,1\text{-}T} - \gamma_{2,0}|| \rightarrow_p 0$, $\max_{j=1,...,T-m+1} ||B_T(\widehat{\gamma}_{1,(j)} - \gamma_{1,0})|| \rightarrow_p 0$, and $\max_{j=1,...,T-m+1} ||\widehat{\gamma}_{2,(j)} - \gamma_{2,0}|| \rightarrow_p 0$ with m fixed, under H_0 and H_1, where B_T is as in Assumption 3. When $v = b$, the same conditions hold but with $\widehat{\gamma}_{\ell,1\text{-}T}$ replaced by $\widehat{\gamma}_{\ell,1\text{-}(T+\lceil m/2 \rceil)}$ for $\ell = 1, 2$. When $v = c$, the same conditions hold but with $\widehat{\gamma}_{\ell,1\text{-}T}$ and $\widehat{\gamma}_{\ell,(j)}$ replaced by $\widehat{\gamma}_{\ell,1\text{-}(T+m)}$ and $\widehat{\gamma}_{\ell,2(j)}$, respectively, for $\ell = 1, 2$.

Assumption 5. The distribution function of $R_1(\beta_0)$ or $P_1(\beta_0)$ is continuous and increasing at its $1 - \alpha$ quantile.

By the definition of w_t given above and the second condition of Assumption 1, the joint distribution of all the variables for time periods $t = 1, ..., T$ is the same under H_0 and H_1. This implies that Assumption 5 and the first set of moment conditions in Assumption 2 hold under both H_0 and H_1.

Assumption 1 is relatively weak in terms of the restriction it puts on the temporal dependence of the errors and stationary regressors. For example, ergodicity allows for long-memory. Assumption 2 imposes mild moment conditions on the errors and stationary regressors. For example, the errors do not need to have a finite variance.

Assumption 3 requires that the (transformed) unit root and deterministic regressors, $z_{1,t}$, can be properly normalized. The diagonal element of B_T that corresponds to a unit root variable in $z_{1,t}$ with mean zero "asymptotically weakly dependent" innovations is $T^{1/2}$.[3] Examples of "asymptotically weakly dependent" random variables include strong mixing random variables, linear processes with absolutely summable covariances, and near-epoch dependent (NED) processes. The diagonal element is $T^{1/2}$ in this case because $T^{-1/2}$ times a partial sum of mean zero asymptotically weakly dependent random variables converges weakly to a scaled Brownian motion by a functional central limit theorem (FCLT) under suitable moment conditions. There are numerous results in the literature that provide primitive sufficient conditions for this to hold. See Section 4.2 below. Given weak convergence of the partial sum of the innovations, the continuous mapping theorem (CMT) implies that the condition in Assumption 3 holds for a unit root element of $z_{1,t}$.

The diagonal element of B_T that corresponds to a constant in $z_{1,t}$ is just one. Thus, the condition of Assumption 3 holds trivially for a constant term in $z_{1,t}$. The diagonal element of B_T that corresponds to a linear time trend, t, in $z_{1,t}$ is T. Because $\max_{t \leq T+m}(t/T) = 1 + m/T$, the condition in Assumption 3 also holds trivially for a linear time trend.

As an example of a typical B_T matrix, suppose $z_{1,t} = (1, t, r_t')'$, where r_t is a p vector of unit root variables with mean zero asymptotically weakly dependent innovations. Then, we have

$$B_T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{1/2} I_p \end{pmatrix}. \tag{4.4}$$

Assumption 3 also allows for unit root processes with stationary long-memory or fractional difference innovations. The diagonal element of B_T that corresponds to a unit root process with stationary innovations that have long-memory or fractional difference parameter $d \in (-1/2, 1/2)$ is $(1/2) + d$. This follows by results for the weak convergence of the partial sums of such processes, e.g., see Sowell (1990, Thms. 1 and 2).

Assumption 4 concerns the behavior of the transformed estimators $\widehat{\gamma}_{1\text{-}T}$, $\widehat{\gamma}_{1\text{-}(T+\lceil m/2 \rceil)}$, $\widehat{\gamma}_{1\text{-}(T+m)}$, $\widehat{\gamma}_{(j)}$, and $\widehat{\gamma}_{2(j)}$. The assumptions are not very restrictive. For example, the estimator of the parameters on stationary regressors just needs to be consistent and most such estimators are actually $T^{1/2}$-consistent. The estimator of the parameters on unit root regressors (based on mean zero asymptotically weakly dependent innovations) just need to be $T^{1/2}$-consistent, and most such estimators are actually T-consistent. Similarly, the conditions on deterministic regressors are weaker than what most estimators satisfy.

In the following section, we provide sufficient conditions for Assumption 4 when the LS estimator is used.

Assumption 5 holds if the errors have an absolutely continuous component, which is not very restrictive.

4.2 Least Squares Estimation

In this section, we give sufficient conditions for Assumption 4 for the case where the estimator employed is the LS estimator defined in (3.5). The conditions given are also sufficient for Assumptions 2 and 3.

We consider weak convergence (denoted $\Rightarrow$) of a stochastic process, $\nu_T(\cdot)$, defined on $[0,1]$ to a limit process that has bounded continuous sample paths a.s. The precise definition of "weak convergence" requires specification of a pseudo-metric on the space of functions on $[0,1]$. We use the uniform metric, as in Pollard (1984).

Let $[a]$ denote the integer part of a.

The following assumption, combined with Assumption 1, is sufficient for Assumptions 2-4 when the estimator used in the test statistics is the LS estimator.

Assumption LS. **(a)** $Ez_{2,t}u_t = 0$, $E|u_t|^{1+\delta} < \infty$, $E||u_t z_{2,t}||^{1+\delta} < \infty$, and $E||z_{2,t}||^{2+\delta} < \infty$ for some $\delta > 0$ for $t \leq T$.
(b) $\nu_T(\cdot) \Rightarrow \nu(\cdot)$ as $T \to \infty$, where $\nu_T(r) = B_T^{-1} z_{1,[Tr]}$ for $r \in [0,1]$, $\{B_T : T \geq 1\}$ are non-random positive-definite diagonal $k_1 \times k_1$ matrices, and $\nu(\cdot)$ is some stochastic process that has bounded and continuous sample paths a.s.
(c) $T^{-1} \sum_{t=1}^{T} B_T^{-1} z_{1,t}(u_t, z_{2,t}') = o_p(1)$.
(d) $\max_{t=T+1,\ldots,T+m} ||B_T^{-1} z_{1,t}|| = O_p(1)$ under H_0 and H_1.
(e) $\int_0^1 \nu(r)\nu(r)' dr$ and $\Sigma_{2,0} = Ez_{2,t}z_{2,t}'$ are positive definite a.s.

All parts of Assumption LS except part (d) involve variables indexed by $t \leq T$. These variables have the same distribution under H_0 and H_1 by the definition of $\{w_t : t \leq T\}$ given above and Assumption 1. In consequence, the conditions in Assumption LS hold under both H_0 and H_1.

The first condition of Assumption LS(a) specifies that the stationary regressors, $z_{2,t}$, are not endogenous. This is needed for the estimators of $\gamma_{2,0}$ to be consistent. The remaining conditions of Assumption LS(a) are a slight strengthening of the moment conditions of Assumption 2 (that is used to obtain uniformity of $\widehat{\gamma}_{(j)} - \gamma_0 \to_p 0$ over $j = 1, \ldots, T+m-1$).

Assumption LS(b) holds under a variety of different conditions stated in the literature. We give two examples below—one using strong mixing and the other using linear process conditions.

Assumption LS(c) is a weaker condition than is often satisfied under common conditions in the literature. Typically, the random variable in Assumption LS(c) multiplied by $T^{1/2}$ converges in distribution to some random variable and, hence, Assumption LS(c) holds with $O_p(T^{-1/2})$ in place of $o_p(1)$. Two examples of sufficient conditions for Assumption LS(c) are given below.

Assumption LS(d) is not very restrictive because the maximum is over a finite number, m, of terms. Assumption LS(d) is automatically satisfied if (i) the unit root and deterministic regressors, $z_{1,t}$, come from a sequence, rather than a triangular array, and (ii) $B_T^{-1} B_{T+m} = O(1)$.[4] Condition (i) is innocuous under H_0. Under H_1, it might be restrictive because one might want to allow the behavior of the unit root regressors to change after the breakdown point. If so, then Assumption LS(d) specifies the extent to which the unit root regressors can exhibit different behavior under H_1 after the breakpoint. Condition (ii) on B_T is satisfied in all cases of interest.

The following conditions plus Assumption 1 are sufficient for Assumptions LS(b)-

(d): Under H_0 and H_1,

(i) $z_{1,t}$ contains a vector of polynomials in t with non-negative exponents and/or a unit root random vector $z^*_{1,t}$ that satisfies $z^*_{1,t} = z^*_{1,t-1} + v_t$ for $t = 1, 2, ...,$ where $z^*_{1,0} = O_p(1)$,

(ii) $\sup_{t \geq 1} E||(w'_{0,t}, v'_t)||^{\beta+\varepsilon} < \infty$ for some $\beta > 2$ and $\varepsilon > 0$, and

(iii) $\{(w'_{0,t}, v'_t)' : t \geq 1\}$ is a weakly stationary strong mixing sequence of mean zero random variables with strong mixing numbers that satisfy

$$\sum_{r=1}^{\infty} \alpha^{1-2/\beta}(r) < \infty. \tag{4.5}$$

In this case, any element of $\nu(r)$ (defined in Assumption LS(b)) that corresponds to a polynomial in t, say t^a for $a \geq 0$, is r^a for $r \in [0, 1]$. In addition, the sub-vector of $\nu(r)$ that corresponds to unit root elements of $\nu_T(r)$ is a vector Brownian motion, $\{B(r) : r \in [0, 1]\}$ with covariance matrix

$$\Omega^* = E\zeta_1\zeta'_1 + \sum_{k=2}^{\infty} E\zeta_1\zeta'_k + \sum_{k=2}^{\infty} E\zeta_k\zeta'_1, \text{ where } \zeta_t = (w'_{0,t}, v'_t)'. \tag{4.6}$$

Sufficiency of (4.5) for Assumptions LS(b)-(d) follows from Lemma 2.2 and Theorem 2.6 of Phillips (1988b) when $z_{1,t}$ contains just a unit root random vector. (The diagonal elements of B_T are all $T^{1/2}$ in this case.) When $z_{1,t}$ contains a polynomial, say t^a for $a \geq 0$, we take the corresponding element of B_T to be T^a and the polynomial element of $\nu_T(r)$ converges to the non-random polynomial r^a uniformly over $r \in [0, 1]$. Hence, Assumptions LS(b) and (d) hold when polynomials are present. The elements of $T^{-1}\sum_{t=1}^{T} B_T^{-1} z_{1,t}(u_t, z'_{2,t})$ that correspond to polynomials in $z_{1,t}$ are $o_p(1)$ because after normalization by B_T^{-1} the polynomials are bounded by one and, hence, a WLLN for triangular arrays of mean zero L^2-bounded strong mixing random variables gives the desired result (see Andrews (1988, Thm. 2 and Remark 4 of Sec. 3)).

The following conditions plus Assumption 1 are an alternative set of sufficient conditions for Assumptions LS(b)-(d): Under H_0 and H_1,

(i) condition (i) of (4.5) holds,

(ii) $(w'_{0,t}, v'_t)' = \sum_{j=-\infty}^{\infty} C_j \varepsilon_{t-j}$ and $\{\varepsilon_t : t \geq 1\}$ are iid with mean zero and variance $\Delta > 0$,

(iii) $\sum_{j=-\infty}^{\infty} ||C_j||_* < \infty$ and $\sum_{k=1}^{\infty} \left(||\sum_{j=k}^{\infty} C_j||_* + ||\sum_{j=k}^{\infty} C_{-j}||_* \right) < \infty$,

where $||C_j||_* = \max_k |\sum_{\ell} C_{j,k,\ell}|$ and $C_{j,k,\ell} = [C_j]_{k,l}$. (4.7)

The limit random vector $\nu(r)$ that arises in Assumption LS(b) is the same in this case as defined in the paragraph containing (4.5). Sufficiency of the conditions in (4.7) follows from the Theorem and its proof in Phillips (1988a) when $z_{1,t}$ contains just a unit root vector. The extension to the case where $z_{1,t}$ may also contain polynomials is as above.

Assumption LS(e) is standard in the literature. It rules out the case where one or more regressor is redundant. This assumption is not critical because the test statistics depend on residuals, which depend on the column space spanned by the regressors, not on the regressors themselves. We utilize this condition because it is not very restrictive and its elimination would complicate the results and the proofs.

Lemma 1 *Assumptions* 1 *and* LS *imply that Assumptions* 2-4 *hold for* $v = a$, b, *and* c.

Comment. Analogues of Lemma 1 could be established for other estimators, such as fully modified, ML, and various other asymptotically efficient estimators mentioned above. For brevity, we do not do so.

4.3 Asymptotic Results

We now state the asymptotic results that justify the use of the parametric subsample critical values that are introduced above.

Let $\widehat{F}_{P_v,T}(x)$ denote the empirical df based on $\{P_{v,j} : j = 1, ..., T - m + 1\}$ for $v = a$, b, and c. That is,

$$\widehat{F}_{P_v,T}(x) = \frac{1}{T - m + 1} \sum_{j=1}^{T-m+1} 1(P_{v,j} \leq x). \tag{4.8}$$

Let $F_P(x)$ denote the df of $P_1(\beta_0)$ at x. Let $q_{P,1-\alpha}$ denote the $1-\alpha$ quantile of $P_1(\beta_0)$. Let $\widehat{q}_{P_v,1-\alpha}$ denote the $1-\alpha$ sample quantile of $\{P_{v,j} : j = 1, ..., T-m+1\}$ for $v = a$, b, and c, as defined in (3.10).

Let P_∞ be a random variable with the same distribution as $P_{T+1}(\beta_0)$. Under Assumptions 1-5 and H_0, the distribution of $P_{T+1}(\beta_0)$ equals that of $P_1(\beta_0)$. Also, the distribution of $P_{T+1}(\beta_0)$ does not depend on T under either H_0 or H_1. Under H_0, this holds by stationarity. Under H_1, this holds because the distribution of $\{w_{T,t} : t = T+1, ..., T+m\}$ is assumed to be independent of T, which is appropriate for fixed alternatives.

Define $\widehat{F}_{R_v,T}(x)$, $F_R(x)$, $q_{R,1-\alpha}$, $\widehat{q}_{R_v,1-\alpha}$, and R_∞ analogously with R in place of P.

The main result of the paper is the following.

Theorem 1 *Suppose Assumptions* 1-5 *hold for* $v = a$, b, *or* c. *Then, as* $T \to \infty$,
(a) $P_v \to_d P_\infty$ *under* H_0 *and* H_1,
(b) $\widehat{F}_{P_v,T}(x) \to_p F_P(x)$ *for all* x *in a neighborhood of* $q_{P,1-\alpha}$ *under* H_0 *and* H_1,
(c) $\widehat{q}_{P_v,1-\alpha} \to_p q_{P,1-\alpha}$ *under* H_0 *and* H_1, *and*
(d) $\Pr(P_v > \widehat{q}_{P_v,1-\alpha}) \to \alpha$ *under* H_0.
(e) *Parts* (a)-(d) *hold with* R *in place of* P.

Comments: 1. The asymptotic distribution of P_v under H_0 and H_1 is given in part (a) of the Theorem.

2. Part (c) of the Theorem shows that the random critical value $\widehat{q}_{P_v,1-\alpha}$ has the same asymptotic behavior under H_1 as under H_0. This is desirable for the power of the test.

3. Part (d) of the Theorem shows that the asymptotic size of the test is α, as desired.

4. Part (a) shows that P_v does not diverge to infinity as $T \to \infty$ under H_1. Hence, P_v is not a consistent test. This is due to the assumption that the number, m, of post-breakdown observations is fixed and does not go to infinity in the asymptotics. However, if $P_{T+1}(\beta_0)$ is stochastically greater than $P_1(\beta_0)$ under H_1, then P_v is an asymptotically unbiased test.

5. Parts (c) and (d) of the Theorem follow easily from part (b). The idea of the proof of part (b) is to show that (i) the difference between $\widehat{F}_{P_v,T}(x)$ and a smoothed version of it, say $\widehat{F}_{P_v,T}(x, h_T)$, converges in probability to zero, where h_T indexes the amount of smoothing and $h_T \to 0$ as $T \to \infty$, (ii) the difference between $\widehat{F}_{P_v,T}(x, h_T)$ and an analogous df with $\widehat{\beta}_{(j)}$ or $\widehat{\beta}_{2(j)}$ replaced by β_0 converges in probability to zero, (iii) the difference between the latter and the empirical df of $\{P_j(\beta_0) : j = 1, ..., T - m + 1)\}$ converges in probability to zero as $T \to \infty$, and (iv) the difference between the latter and its expectation, $F_P(x)$, is asymptotically negligible. The reason for considering a smoothed version of $\widehat{F}_{P_v,T}(x)$ is that it is a smooth function of $P_{v,j}$ and, hence, result (ii) can be established by taking a mean-value expansion about $P_j(\beta_0)$. Result (iv) holds by the ergodic theorem because $\{P_j(\beta_0) : j = 1, ..., T - m + 1)\}$ is a finite subset of stationary and ergodic random variables using Assumption 1.

5 Tests for Breakdown at the Beginning, or in the Middle, of the Sample

The tests introduced above for detecting cointegration breakdown at the end of the sample can be altered to detect breakdown occurring at the beginning or in the middle of the sample. For example, one might be interested in determining the most suitable starting date for a model. Or, one might be interested in whether a model behaves differently during a policy regime shift or during war years than in other years in the sample. Such periods of potential breakdown are often of relatively short duration, so that asymptotic tests that are based on their length going to infinity are not appropriate. In such cases, the testing method introduced above is useful because the length, m, of the time period of potential breakdown is taken to be fixed and finite in the asymptotics.

We consider testing for cointegration breakdown for the m observations indexed by $t = t_0, ..., t_0 + m - 1$ when the total number of observations is $T + m$. The null and alternative hypotheses are given by

$$H_0 : \begin{cases} y_t = x_t'\beta_0 + u_t \text{ for all } t = 1, ..., T+m \text{ and} \\ \{u_t : t \geq 1\} \text{ are stationary and ergodic} \end{cases}$$

$$H_1 : \begin{cases} y_t = x_t'\beta_0 + u_t \text{ for all } t = 1, ..., t_0 - 1,\ t_0 + m, ..., T+m \text{ and} \\ y_t = x_t'\beta_t + u_t \text{ with } \beta_t \neq \beta_0 \text{ for some } t = t_0, ..., t_0 + m - 1 \text{ and/or} \\ \text{the distribution of } \{u_{t_0}, ..., u_{t_0+m-1}\} \text{ differs from that of} \\ \text{error sequences } \{u_t, ..., u_{t+m-s}\} \text{ that do not overlap with it.} \end{cases} \quad (5.1)$$

One can construct tests for these hypotheses by moving the observations $\{(y_t, x_t) : t = t_0, ..., t_0 + m - 1\}$ to the end of the sample and moving the observations after $t = T+m-1$ up to fill the gap. The observations originally indexed by $t = t_0, ..., t_0+m-1$ are subsequently indexed by $t = T, ..., T+m$ and the tests defined above can be used to test the hypotheses in (5.1).

6 Monte Carlo Experiment

In this section, we describe some Monte Carlo results that are designed to assess and compare the size and power properties of the tests P_a-P_c and R_a-R_c.

6.1 Experimental Design

We consider linear regression models estimated by LS. For results under the null hypothesis, the model we consider is

$$y_t = x_t'\beta_0 + u_t \text{ for } t = 1, ..., T+m \quad (6.1)$$

with $\beta_0 = 0$. We consider two values of T: 100 and 250. We consider three values of m: 10, 5, and 1. In the base model that we consider, we take

$$y_t = \beta_{1,0} + t\beta_{2,0} + x_{1,t}'\beta_{3,0} + x_{2,t}'\beta_{4,0} + u_t, \quad (6.2)$$

where $x_{1,t}$ is a vector of unit root regressors and $x_{2,t}$ is a vector of stationary mean zero regressors. The errors, u_t, the difference of the unit root regressors, $Dx_{1,t} = x_{1,t} - x_{1,t-1}$, and the stationary regressors, $x_{2,t}$, are all AR(1) processes with the same AR(1) parameter ρ and the same innovation distribution G. We consider three values of ρ: 0, .4, and .8. We consider four innovation distributions: (i) standard normal ($N(0,1)$), (ii) chi-squared with two degrees of freedom (χ_2^2) recentered and rescaled to have mean zero and variance one, (iii) t with three degrees of freedom (t_3) rescaled to have variance one, and (iv) uniform (U) on $[-\sqrt{12}/2, \sqrt{12}/2]$, which has mean zero and variance one. The different innovation distributions display standard behavior ($N(0,1)$), skewness (χ_2^2), excess kurtosis (t_3), and thin tails (U).

The stationary regressors $x_{2,t}$ are independent of the errors, the unit root regressors, and each other. The unit root regressors and errors may be correlated with the correlation between $Dx_{1,t}$ and u_t being $\rho_{Dx,u}$. This correlation is achieved by taking each element of Dx_t and u_t to have a common component.

The errors, the differences of the unit root regressors, and the stationary regressors are generated as follows. The innovations to the various AR(1) processes that are utilized are

$$\{(\psi_t^*, \xi_t^{*\prime}, \eta_t^*, x_{2,t}^{*\prime})' : t = 1, ..., T+m\}, \tag{6.3}$$

where $\psi_t^*, \eta_t^* \in R$, $\xi_t^* \in R^{d_{x_1}}$, $x_{2,t}^* \in R^{d_{x_2}}$, and d_{x_1} and d_{x_2} denote the dimensions of $x_{1,t}$ and $x_{2,t}$, respectively. The innovations are iid across the elements of $(\psi_t^*, \xi_t^{*\prime}, \eta_t^*, x_{2,t}^{*\prime})'$ and across t. Each element of $(\psi_t^*, \xi_t^{*\prime}, \eta_t^*, x_{2,t}^{*\prime})'$ has distribution G for G as above. The AR(1) processes based on these innovations are

$$\begin{aligned} \psi_t &= \rho\psi_{t-1} + \psi_t^*, \\ \xi_t &= \rho\xi_{t-1} + \xi_{,t}^*, \\ \eta_t &= \rho\eta_{t-1} + \eta_{,t}^*, \text{ and} \\ x_{2,t} &= \rho x_{2,t-1} + x_{2,t}^* \end{aligned} \tag{6.4}$$

for $t = 1, ..., T+m$. The elements of the initial conditions $(\psi_0, \xi_0', \eta_0, x_{2,0}')'$ are iid each with distribution G but rescaled to yield a variance stationary AR(1) sequence. (For example, $(1-\rho^2)^{1/2}\psi_0$ has distribution G.)

The errors and differences of the unit root regressors are

$$\begin{aligned} u_t &= (1-\rho_{Dx,u})^{1/2}\psi_t + \rho_{Dx,u}^{1/2}\eta_t \text{ and} \\ Dx_{1,t} &= (1-\rho_{Dx,u})^{1/2}\xi_t + \rho_{Dx,u}^{1/2}\eta_t 1_{d_{x_1}}. \end{aligned} \tag{6.5}$$

where $1_{d_{x_1}}$ denotes a d_{x_1}-vector of ones. As defined, the errors and regressors have correlation $\rho_{Dx,u}$.

The base model that we consider has an intercept, time trend, two unit root regressors, and two stationary regressors all with standard normal innovations and no correlation between the unit root regressors and the error:

$$\begin{aligned} \text{Base Model:} \quad & \text{BC(i). } x_t = (1, t, x_{1,t}', x_{2,t}')' \text{ and } x_{1,t}, x_{2,t} \in R^2. \\ & \text{BC(ii). } \rho_{Dx,u} = 0. \\ & \text{BC(iii). } G = N(0,1). \end{aligned} \tag{6.6}$$

We consider seven variants of the base model. Models 2-4 differ from the base model in terms of the distribution of the innovations. Models 5 and 6 differ from the base model in that $\rho_{Dx,u} = .4$ and $\rho_{Dx,u} = .8$, respectively. Model 7 is the same as the base model except there are no stationary regressors and there are four unit root regressors. Model 8 is the same as the base model except there is no time trend.

Models 2-7 are summarized as follows:

$$\begin{array}{lll} \text{Model 2} & (\chi^2_2 \text{ Distn}): & \text{BC(i) and BC(ii) hold and } G = \chi^2_2. \\ \text{Model 3} & (t_3 \text{ Distn}): & \text{BC(i) and BC(ii) hold and } G = t_3. \\ \text{Model 4} & (U \text{ Distn}): & \text{BC(i) and BC(ii) hold and } G = U. \\ \text{Model 5} & (\rho_{Dx,u} = .4) & \text{BC(i) and BC(iii) hold and } \rho_{Dx,u} = .4. \\ \text{Model 6} & (\rho_{Dx,u} = .8) & \text{BC(i) and BC(iii) hold and } \rho_{Dx,u} = .8. \\ \text{Model 7} & (\text{No Stat. Regr.}) & \text{BC(ii) and BC(iii) hold, } x_t = (1, t, x_{1,t}')', \\ & & \text{and } x_{1,t} \in R^4. \\ \text{Model 8} & (\text{No Time Trend}) & \text{BC(ii) and BC(iii) hold, } x_t = (1, x_{1,t}', x_{2,t}')', \\ & & \text{and } x_{1,t}, x_{2,t} \in R^2. \end{array} \tag{6.7}$$

For each of the eight models, we consider three values of ρ, two values of T, and three values of m.

For each of the eight models, we report the actual rejection rates of the nominal 5% tests P_a, P_b, P_c, R_a, R_b, and R_c.

In addition, we report the size-corrected power of the tests for two types of alternatives to the null hypothesis. The first type of alternative is where cointegration breaks down at time $t = T$ because the errors are a unit root process for $t = T+1$ to $t = T + m$. These are referred to as *unit root* alternatives. In this case, the model is the same as under the null except that for $t = T + 1, ..., T + m$, the error is given by

$$\begin{aligned} u_t &= (1 - \rho_{Dx,u})^{1/2} \psi_t + \rho_{Dx,u} \eta_t \\ &\quad + \sqrt{2} \sum_{s=1}^{t-T} [(1 - \rho_{Dx,u})^{1/2} \widetilde{\psi}_s + \rho_{Dx,u} \widetilde{\eta}_s], \end{aligned} \tag{6.8}$$

where $\{(\widetilde{\psi}_s, \widetilde{\eta}_s) : s = 1, ..., m\}$ has the same distribution $\{(\psi_s, \eta_s) : s = 1, ..., m\}$ and is independent of all other random variables in the model. The multiplicative factor $\sqrt{2}$ is chosen so that the rejection rates of the tests are in an informative range. One can increase or decrease power to any desired level by altering the multiplicative factor.

The second type of alternative considered is a *parameter shift* alternative. In this case, the model is a cointegrating model for all $t = 1, ..., T + m$, but the cointegrating vector is different before and after $t = T$. For this alternative, the true distribution of the data is the same as under the null except that for $t = T + 1, ..., T + m$ the true parameter β_0 is proportional to a vector of ones with $||\beta_0|| = .25$. The value .25 is chosen so that the rejection rates of the tests are in an informative range. One can increase or decrease power to any desired level by altering $||\beta_0||$.

The power results that we report are for size-corrected tests because we do not want to confound power differences with size distortions. Size-correction is not as straightforward with the tests considered here as it is in some situations because the tests' critical values are sample quantiles, not constants. We determine by simulation the significance levels that yield the finite sample null rejection rates to be as close to the desired test size, .05, as possible for each innovation distribution and each T, m, ρ, and $\rho_{Dx,u}$ value when the observations are generated under the null. (The rejection rates cannot be made exactly equal to .05 because the sample quantile functions are not continuous. But, the differences are fairly small.) These significance levels are employed when computing the size-corrected power of the nominal .05 tests. Note that this method of size correction is equivalent to the standard method of adjusting a test's critical value for any test that has a non-random critical value.

All the results reported are based on $40,000$ simulation repetitions. This yields simulation standard errors of (approximately) .001 for the simulated null rejection rates of nominal .05 tests and simulated standard errors in the interval (.0020, .0025) for the simulated alternative hypothesis rejection rates when these rejection rates are in the interval (.20, .80).

6.2 Monte Carlo Results

6.2.1 Size

Table I presents the test size results for nominal .05 tests. The first six rows of Table I give the average rejection rate and the range of the rejection rates over all eight models and nine (m, ρ) values for each of the six tests. The remaining rows in the table give the average and range of the rejection rates over the eight models for each (m, ρ) value and each test.

Tables A-I and A-II give the rejection rates for each of the 72 model/(m, ρ) combinations. Some of the results stated below are based on these more detailed tables.

When $m = 1$, separate results are not given for R_a, R_b, R_c, and P_c because $P_a = R_a$ and $P_b = P_c = R_b = R_c$ when $m = 1$.

The main results are as follows:

1. The P_a and R_a tests have the highest rejection rates. They over-reject the null by a noticeable margin, especially when $T = 100$. The remaining four tests have rejection rates that are much closer to the desired level .05. The P_b and R_b tests have higher rejection rates than the P_c and R_c tests and tend to over-reject the null by a small margin . The P_c and R_c tests have the lowest rejection rates. The R_c test, however, under-rejects in many cases. For example, when $\rho = 0$ or .4 and $m = 10$ or 5, its average rejection rates over the eight models are between .032 and .038 when $T = 100$. It appears that the P_c test has the best size performance overall. Its average rejection rate is close to .05 when $T = 100$ or 250. Its deviation from this value is at most .014 when $T = 250$ over the wide range of models considered.

2. Not surprisingly, all tests perform noticeably better in terms of size when $T = 250$ than when $T = 100$. In particular, the range of rejection rates for each test shrinks considerably when T is increased.

3. For the tests P_a and R_a, the rejection rates increase as ρ or m increases. But, the other tests are not very sensitive to the values of ρ and m.

4. For the P_b and P_c tests, the rejection rates are higher for the χ^2_2 and t_3 distributions than for the normal and lower for the uniform. For the R_b and R_c

tests, there is no clear pattern of variation of the rejection rates with the type of distribution.

5. For the P_b, P_c, R_b, and R_c tests, the rejection rates for $\rho_{Dx,u} = 0$ and .4 are quite similar. The rejection rates tend to be somewhat higher for $\rho_{Dx,u} = .8$. But, overall, the sensitivity to $\rho_{Dx,u}$ is fairly low.

6. For the P_b, P_c, R_b, and R_c tests, sensitivity also is low with respect to the number of stationary regressors versus the number of unit root regressors. That is, their rejection rates do not change much between the Base Model and the No Stationary Regressors Model.

7. The results for the No Time Trend Model are somewhat unique. The P_b and P_c tests react differently to the elimination of the time trend. The rejection rate of P_b goes down, while that of P_c goes up. This is true for both $T = 100$ and $T = 250$. In consequence, the No Time Trend Model is the only model for which the rejection rates of P_c are higher than those of P_b. The same pattern is observed for the R_b and R_c tests.

To conclude, we find that the rejection rates of the P_a and R_a tests are too high, and these tests are clearly inferior in terms of size compared to the other tests. The P_b, P_c, R_b, and R_c tests have size performances that are similar. But, the P_b and R_b tests tend to reject the null hypothesis somewhat too often compared to the P_c and R_c tests. The R_c test tends to under-reject the null too often compared to the P_c test. Hence, the P_c test has the best overall size properties. Considering the very wide range of models and (m, ρ) values considered, which range from t_3 distributions to $\rho_{Dx,u} = .8$, the size performance of the P_c test seems quite good. This is especially true for $T = 250$.

6.2.2 Power

Table II provides the size-corrected power results for the unit root and parameter shift alternatives. Averages of rejection rates are reported for the same models and (m, ρ) values as in Table I. Tables A-III to A-VI give the rejection rates for each of the 72 model/(m, ρ) combinations.

The principle findings are as follows:

1. The power of the P_a test is almost always greater than or equal to that of the P_b and P_c tests. In many cases, the differences are small, but in some cases the differences are noticeable. The same pattern holds for the R_a test in comparison with the R_b and R_c tests. Nevertheless, because the size properties of the P_a and R_a tests are poor compared to the other tests, we focus on the power properties of the other tests.

2. The simulation results indicate that the P tests have considerable power against unit root alternatives, even though they are designed for parameter shift alternatives. Likewise, the R tests have considerable power against parameter shift

alternatives even though they are designed for unit root alternatives. In fact, paradoxically, the P tests tend to outperform the R tests for unit root alternatives and vice versa with parameter shift alternatives. But, the differences are not large.

3. The average power across all cases of the P_b and P_c tests is the same for unit root alternatives with $T = 100$ and 250 and for parameter shift alternatives with $T = 250$. It differs by only .02 for parameter shift alternatives with $T = 100$, in which case P_b has higher power.

4. The R_c test has slightly higher power than the R_b test across most cases. The difference typically is .02 when $T = 100$ and .01 when $T = 250$.

5. For all tests, power increases sharply with m. This occurs because m determines the amount of information that is available regarding the post-break time period.

6. For all tests, power increases by a small amount (roughly .03) as T increases from 100 to 250 for parameter shift alternatives. Power increases by a substantial amount (roughly .15) as T increases from 100 to 250 for unit root alternatives.

7. Power of the tests is not very sensitive to changes in $\rho_{Dx,u}$ or to shifts from the Base Model to the No Time Trend Model. The latter result is somewhat surprising. The P tests have lower power for the No Stationary Regressors Model than the Base Model for parameter shift alternatives.

8. Power for the P tests is more sensitive to the distribution than it is for the R tests. For the P tests, power is lower for the t_3 and χ^2_2 distributions than for the normal and higher for the uniform than the normal.

Overall, the power of the P_b, P_c, R_b, and R_c tests is fairly comparable. The R_b test is dominated by that of the R_c test, but only by a small margin. The P_b and P_c tests have similar power. The P_b and P_c tests have somewhat higher power than the R_c test against unit root alternatives, but vice versa for parameter shift alternatives. The R_c test tends to have power that is less variable across changes in the model, such as changes in the distribution, than the P tests. In consequence, the R_c test is deemed to have the best overall power properties among the tests P_b, P_c, R_b, and R_c, but by a small margin.

Combining the size and power results, we find that the choice of the best test among the six tests considered is not clear-cut. The tests P_a and R_a are clearly inferior to the other in terms of size. However, the remaining four tests are not as easy to distinguish. The P_b and R_b tests tend to reject too often under the null compared to the P_c and R_c tests. In addition, the R_b test has slightly lower power than the R_c test. Hence, the P_c and R_c tests appear to be the best two tests. Of these two tests, the P_c test has somewhat better size properties because the R_c test is somewhat under-sized in a number of cases. On the other hand, the R_c test has power that is less variable across different distributions than the P_c test. On balance, the P_c test seems to be preferable because of its size properties. In consequence, we recommend using the P_c test.

7 An Empirical Example

In this section, we provide an empirical illustration of the P and R tests. We consider interest rate parity for several Asian countries in a recent period of financial crisis. We investigate the hypothesis that turbulence in financial markets in the region led to a breakdown in interest rate parity.

Interest rate parity is written as

$$i_t = i_t^* + \frac{E_{t+1}^e - E_t}{E_t}, \tag{7.1}$$

where i_t is the domestic interest rate, i_t^* is the foreign interest rate, E_t is the spot exchange rate, and E_{t+1}^e is the forward exchange rate. Thus, interest rate parity says that the domestic interest rate, i_t, equals the foreign interest rate, i_t^*, plus the expected rate of depreciation of the domestic currency, $(E_{t+1}^e - E_t)/E_t$. The above relation is an equilibrium condition in the currency market.

If we allow an equilibrium error, u_t, in the relation, we have

$$i_t = i_t^* + \frac{E_{t+1}^e - E_t}{E_t} + u_t. \tag{7.2}$$

It is plausible to think that turbulence in the financial market may cause an important change in the equilibrium error process, i.e., a breakdown in interest rate parity. In particular, the error process, u_t, may change from a stationary process to a unit root process if the financial crisis causes a breakdown in interest rate parity and the underlying variables i_t, i_t^*, E_t, and E_{t+1}^e are unit root processes.

The data we consider are daily observations on the domestic interest rate, spot exchange rate, and forward exchange rate of four East Asian countries: Thailand, the Philippines, Indonesia, and Singapore. The exchange rates are in US dollars. These data are taken from the college data bank of the College of Business Administration at Seoul National University. The foreign interest rate is the three month U.S. treasury bill rate. The observations are daily (excluding weekends) from December 31, 1996 to July 13, 1998, which yields a total of 400 observations.

Figures 1.1(a), ..., 1.4(a) graph the u_t process, as defined in (7.2), for each of the four countries. In each of these four figures a vertical line indicates an estimated breakpoint obtained using the method Kim (2000) applied to all 400 observations. For the four countries the estimated breakpoints occur at the 5/15/97, 7/01/97, 7/10/97, and 10/10/97 dates, respectively. Note that Kim's method looks for a breakpoint anywhere in the time period considered not just at the end of the sample.

We use the data to show what would happen if one tested for cointegration breakdown at the end of the sample (with $m = 10$) for a variety of different sample periods. To do so, we consider the sample periods that start in December 31, 1996 and include

91 observations, 92 observations, ..., 400 observations. Because $m = 10$, the values of the hypothetical breakpoint, T, as defined in (2.1), (2.2), and (3.6), are 81, 82,..., 390. For each value of T, we compute the p-value for the P and R tests. Note that there is no difference between the P_a, P_b, and P_c tests because the coefficients in the cointegrating relation (7.2) are known, and likewise for the R_a, R_b, and R_c tests. The p-values for the P tests are graphed in Figures 1.1(b), ..., 1.4(b) with the horizontal axis indexed by T. For example, the p-value at the point 231 is the p-value for the P-test with $T = 231$ and $m = 10$, which depends on 241 observations. The p-values for the R tests are graphed in Figures 1.1(c),..., 1.4(c). In each of the (b) and (c) graphs a horizontal line is drawn at .05. A significance level .05 test rejects the null hypothesis for samples corresponding to those values of T for which the p-value graph lies below the horizontal .05 line.

The purpose of the (b) and (c) figures is not to do some sort of rolling analysis of the data set of 400 observations. If one had all 400 observations and one did not know if and when cointegration breakdown has occurred, then the appropriate test to use would be Kim's (2000) test or some similar test. Rather, the purpose of these figures is to show succinctly, for a range of values of T, what would happen if one had a specific data set with $T + m$ observations and one carried out a P or R test for end-of-sample cointegration breakdown over the last $m = 10$ observations. The estimated breakpoints using Kim's (2000) method (that uses all 400 observations) can be used to evaluate the performance of the end-of-sample tests. An end-of-sample test is designed to have power for breakdowns that occur at time $T, ..., T + m - 1$. Hence, an end-of-sample test performs well if it rejects the null hypothesis when the sample considered is such that T, $T + 1, ...,$ and/or $T + 9$ equals the Kim estimate of the breakpoint.

Inspection of the p-value graphs shows that they are below the .05 line for T values around the Kim-estimated breakpoints in all eight cases. In particular, for Thailand, the Kim-estimated breakpoint occurs at 5/15/97 and the p-value graphs for the P and R tests are below .05 for the periods 5/2/97 - 6/18/97 and 5/7/97 - 6/16/97, respectively. For the Philippines, the corresponding dates are 7/01/97, 6/20/97 - 7/21/97, and 6/23/97 - 7/17/97. For Indonesia, the corresponding dates are 7/10/97, 6/30/97 - 7/21/97, and 7/2/97 - 7/17/97. For Singapore, the corresponding dates are 10/10/97, 10/20/97 - 10/31/97, and 10/10/97 - 10/30/97. This illustrates that the end-of-sample tests have sufficient power in these cases to reject the null hypothesis when the null is false (at least as indicated by the Kim estimate of the breakpoint).

The p-value graphs all rise above the .05 line for some values of T larger than the Kim-estimated breakpoints. This indicates that if one tests for a breakdown at the end of the sample, but the actual breakdown occurs some time earlier in the sample period (i.e., before time T), then the P and R tests do not have high power. This is as expected.

We also note that the p-value graphs of the P and R tests are quite similar, especially when one or other graph is near the .05 rejection line. For three of four countries, the P test detects a break a few days before the R test does. Hence, the P test slightly out-performs the R test in this example.

8 Appendix of Proofs

Proof of Theorem 1. The proof is carried out using the transformed parameter estimators $\widehat{\gamma}_{1\text{-}T}$, etc. and transformed regressors z_t, rather than the estimators $\widehat{\beta}_{1\text{-}T}$, etc. and regressors x_t using the fact that $z_t'\widehat{\gamma}_{1\text{-}T} = x_t'\widehat{\beta}_{1\text{-}T}$. Hence, for notational simplicity, but with some abuse of notation, in the proof we let

$$\begin{aligned} P_j(\gamma, \Sigma) &= (\mathbf{Y}_{j\text{-}(j+m-1)} - \mathbf{Z}_{j\text{-}(j+m-1)}\gamma)'\Sigma^{-1}(\mathbf{Y}_{j\text{-}(j+m-1)} - \mathbf{Z}_{j\text{-}(j+m-1)}\gamma) \text{ and} \\ P_j(\gamma) &= P_j(\gamma, I_m), \text{ where} \\ \mathbf{Z}_{r\text{-}s} &= (z_r, ..., z_s)'. \end{aligned} \tag{8.1}$$

Similarly, in the proof, we take $\widehat{F}_{P_a,T}(x)$ and $F_P(x)$ to be defined with $P_j(\widehat{\beta}_{(j)})$ and $P_j(\beta_0)$ replaced by $P_j(\widehat{\gamma}_{(j)})$ and $P_j(\gamma_0)$, respectively.

We start by proving parts (a)-(d) for $v = a$. First, we bound the difference $P_j(\widehat{\gamma}) - P_j(\gamma_0)$, where $\widehat{\gamma}$ denotes $\widehat{\gamma}_{1\text{-}T}$ or $\widehat{\gamma}_{(j)}$. For $\varepsilon > 0$, define the set $L_{1,T}(\varepsilon)$ by

$$\begin{aligned} L_{1,T}(\varepsilon) \;\; = \;\; & \{||B_T(\widehat{\gamma}_{1,1\text{-}T} - \gamma_{1,0})|| \le \varepsilon, \; ||\widehat{\gamma}_{2,1\text{-}T} - \gamma_{2,0}|| \le \varepsilon, \\ & ||B_T(\widehat{\gamma}_{1,(j)} - \gamma_{1,0})|| \le \varepsilon, \; ||\widehat{\gamma}_{2,(j)} - \gamma_{2,0}|| \le \varepsilon, \\ & \forall j = 1, ..., T - m + 1\}. \end{aligned} \tag{8.2}$$

For $c > 0$, define the set $L_{2,T}(c)$ by

$$L_{2,T}(c) = \left\{ \max_{t \le T+m} ||B_T^{-1} z_{1,t}|| \le c \right\}. \tag{8.3}$$

By Assumption 4, there exists a sequence of positive constants $\{\varepsilon_T : T \ge 1\}$ such that $\varepsilon_T \to 0$ and $\Pr(L_{1,T}(\varepsilon_T)) \to 1$ as $T \to \infty$. Let $\{c_T : T \ge 1\}$ be any sequence of constants such that $c_T \to \infty$ and $c_T\varepsilon_T \to 0$ as $T \to \infty$ (e.g., $c_T = \varepsilon_T^{-1/2}$). By Assumption 3, $\Pr(L_{2,T}(c_T)) \to 1$ as $T \to \infty$. Let

$$L_T = L_{1,T}(\varepsilon_T) \cap L_{2,T}(c_T). \tag{8.4}$$

We have

$$\Pr(L_T) \to 1 \text{ and } \Pr(\overline{L}_T) \to 0 \text{ as } T \to \infty, \tag{8.5}$$

where $\overline{L}_T$ denotes the complement of L_T.

Now, for $\widehat{\gamma} = (\widehat{\gamma}_1', \widehat{\gamma}_2')'$ equal to $\widehat{\gamma}_{1\text{-}T}$ or $\widehat{\gamma}_{(j)}$ and for $j = 1, ..., T+1$, we have: on the set L_T,

$$\begin{aligned} & |P_j(\widehat{\gamma}) - P_j(\gamma_0)| \\ = \; & |-2 \sum_{t=j}^{j+m-1} u_t z_t'(\widehat{\gamma} - \gamma_0) + \sum_{t=j}^{j+m-1} (z_t'(\widehat{\gamma} - \gamma_0))^2| \\ \le \; & 2 \sum_{t=j}^{j+m-1} |u_t| \max_{s \le T+m} ||B_T^{-1} z_{1,s}|| \cdot ||B_T(\widehat{\gamma}_1 - \gamma_0)|| + 2 \sum_{t=j}^{j+m-1} ||u_t z_{2,t}|| \cdot ||\widehat{\gamma}_2 - \gamma_0|| \end{aligned}$$

$$
\begin{aligned}
&+\sum_{t=j}^{j+m-1}\left(\max_{s\leq T+m}||B_T^{-1}z_{1,s}||\cdot||B_T(\widehat{\gamma}_1-\gamma_0)||+||z_{2,t}||\cdot||\widehat{\gamma}_2-\gamma_0||\right)^2\\
&\leq 2\sum_{t=j}^{j+m-1}|u_t|c_T\varepsilon_T+2\sum_{t=j}^{j+m-1}||u_tz_{2,t}||\varepsilon_T+\sum_{t=j}^{j+m-1}(c_T\varepsilon_T+||z_{2,t}||\varepsilon_T)^2\\
&=g_j(\varepsilon_T,c_T),
\end{aligned}
\tag{8.6}
$$

where the last equality defines $g_j(\varepsilon_T, c_T)$. Note that $g_j(\varepsilon_T, c_T)$ is identically distributed for $j = 1, ..., T-m+1$ under H_0 and H_1, because $w_t = w_{0,t}$ for $t = 1, ..., T$ is stationary. Also note that $g_{T+1}(\varepsilon, c)$ has distribution that is independent of T if ε and c do not depend on T, because the distribution of $\{w_t : t = T+1, ..., T+m\}$ does not depend on T by Assumption 1.

We prove part (a) first. Let $x \in R$ be a continuity point of the df of $P_{T+1}(\gamma_0)$. We have

$$
\begin{aligned}
&\Pr(P_{T+1}(\widehat{\gamma}_{1\text{-}T})\leq x)\\
&=\Pr(\{P_{T+1}(\widehat{\gamma}_{1\text{-}T})\leq x\}\cap L_T)+\Pr(\{P_{T+1}(\widehat{\gamma}_{1\text{-}T})\leq x\}\cap \overline{L}_T)\\
&\leq \Pr(\{P_{T+1}(\gamma_0)\leq x+g_j(\varepsilon_T,c_T)\}\cap L_T)+o(1)\\
&=\Pr(P_{T+1}(\gamma_0)\leq x)+o(1)\\
&=\Pr(P_\infty\leq x)+o(1),
\end{aligned}
\tag{8.7}
$$

where the inequality holds by (8.5) and (8.6), the second equality holds because $g_j(\varepsilon, c) \to 0$ a.s. as $(\varepsilon, c\varepsilon) \to (0, 0)$, $P_{T+1}(\gamma_0)$ and $g_j(\varepsilon, c)$ have distributions that do not depend on T, and x is a continuity point of $P_{T+1}(\gamma_0)$, and the last equality holds by the definition of P_∞. Equation (8.7) also holds with $\geq$ in place of $\leq$ and $-g_j(\varepsilon_T, c_T)$ in place of $+g_j(\varepsilon_T, c_T)$. Hence, part (a) is proved.

Next, we prove part (b). We introduce the following notation. For some random or non-random vectors $\{\gamma_j : j = 1, ..., T-m+1\}$, let $\widehat{F}_T(x, \{\gamma_j\})$ denote the empirical df based on $\{P_j(\gamma_j) : j = 1, ..., T-m+1\}$. That is,

$$
\widehat{F}_T(x,\{\gamma_j\})=\frac{1}{T-m+1}\sum_{j=1}^{T-m+1}1(P_j(\gamma_j)\leq x)
\tag{8.8}
$$

for $x \in R$. Note that $\widehat{F}_{P_a,T}(x) = \widehat{F}_T(x, \{\widehat{\gamma}_{(j)}\})$.

We define a smoothed version of the df $\widehat{F}_T(x, \{\gamma_j\})$ as follows. Let $k(\cdot)$ be a monotone decreasing, everywhere differentiable, real function on R with bounded derivative and such that $k(x) = 1$ for $x \in (-\infty, 0]$, $k(x) \in [0, 1]$ for $x \in (0, 1)$, and $k(x) = 0$ for $x \in [1, \infty)$. For example, one could take $k(x) = \cos(\pi x)/2 + 1/2$ for $x \in (0, 1)$. For $\{\gamma_j\}$ as above, we define the smoothed df

$$
\widehat{F}_T(x,\{\gamma_j\},h_T)=\frac{1}{T-m+1}\sum_{j=1}^{T-m+1}k((P_j(\gamma_j)-x)/h_T),
\tag{8.9}
$$

where $\{h_T : T \geq 1\}$ is a sequence of positive constants that satisfies $h_T \to 0$ and $c_T \varepsilon_T / h_T \to 0$. For example, if $c_T = \varepsilon_T^{-1/2}$, then one can take $h_T = \varepsilon_T^{1/4}$.

We have

$$\begin{aligned}
&|\widehat{F}_{P_a,T}(x) - F_P(x)| \leq \sum_{i=1}^{4} D_{i,T}, \text{ where} \\
&D_{1,T} = |\widehat{F}_{P_a,T}(x) - \widehat{F}_T(x, \{\widehat{\gamma}_{(j)}\}, h_T)|, \\
&D_{2,T} = |\widehat{F}_T(x, \{\widehat{\gamma}_{(j)}\}, h_T) - \widehat{F}_T(x, \{\gamma_0\}, h_T)|, \\
&D_{3,T} = |\widehat{F}_T(x, \{\gamma_0\}, h_T) - \widehat{F}_T(x, \{\gamma_0\})|, \text{ and} \\
&D_{4,T} = |\widehat{F}_T(x, \{\gamma_0\}) - F_P(x)|. \qquad (8.10)
\end{aligned}$$

We have $D_{4,T} \to_p 0$ under H_0 and H_1 by the ergodic theorem. This holds because $\{P_1(\gamma_0), ..., P_{T-m+1}(\gamma_0)\}$ only depend upon the errors $\{u_1, ..., u_T\}$, which come from the stationary and ergodic sequence $\{w_{0,t} : t \geq 1\}$, and not on the post-breakdown errors $\{u_{T+1}, ..., u_{T+m}\}$. Each random variable $P_j(\gamma_0)$ is the same measurable function of m observations $\{w_{0,j}, ..., w_{0,j+m-1}\}$ for $j = 1, ..., T-m+1$, where m is fixed and finite. Hence, $\{P_1(\gamma_0), ..., P_{T-m+1}(\gamma_0)\}$ is a finite subsequence of a stationary and ergodic sequence of random variables that depend on $\{w_{0,t} : t \geq 1\}$ and the ergodic theorem applies by Assumption 1.

We have

$$D_{1,T} \leq \frac{1}{T-m+1} \sum_{j=1}^{T-m+1} 1(P_j(\widehat{\gamma}_{(j)}) - x \in (0, h_T)), \qquad (8.11)$$

because $\widehat{F}_{P_a,T}(x)$ and $\widehat{F}_T(x, \{\widehat{\gamma}_{(j)}\}, h_T)$ only differ when $(P_j(\widehat{\gamma}_{(j)}) - x)/h_T \in (0, 1)$.

Now, for all $\delta > 0$,

$$\begin{aligned}
&\Pr(D_{1,T} > \delta) \\
&\leq \Pr(\{D_{1,T} > \delta\} \cap L_T) + \Pr(\overline{L}_T) \\
&\leq \Pr\left(\frac{1}{T-m+1} \sum_{j=1}^{T-m+1} 1\left(P_j(\gamma_0) - x \in (-g_j(\varepsilon_T, c_T),\ h_T + g_j(\varepsilon_T, c_T))\right) > \delta\right) \\
&\quad + o(1) \\
&\leq E1\left(P_1(\gamma_0) - x \in (-g_1(\varepsilon_T, c_T),\ h_T + g_1(\varepsilon_T, c_T))\right)/\delta + o(1) \\
&= o(1), \qquad (8.12)
\end{aligned}$$

where the second inequality holds using (8.5), (8.6), and (8.11), the third inequality uses Markov's inequality and the identical distributions of $P_j(\gamma_0)$ for $j = 1, ..., T-m+1$, and the equality holds by the bounded convergence theorem because $g_1(\varepsilon_T, c_T) \to 0$ a.s. and $h_T \to 0$ as $T \to \infty$, and $\Pr(P_1(\beta_0) \neq x) = 1$ by Assumption 5. Hence, $D_{1,T} \to_p 0$.

An analogous, but simpler, argument shows that $D_{3,T} \to_p 0$.

For the proof of part (b) when $v = a$, it remains to show that $D_{2,T} \to_p 0$. By mean-value expansions about $P_j(\gamma_0)$, we have: on the set L_T,

$$\begin{aligned} D_{2,T} &= \left| \frac{1}{T-m+1} \sum_{j=1}^{T-m+1} k'((\widetilde{P}_j - x)/h_T)(P_j(\widehat{\gamma}_{(j)}) - P_j(\gamma_0))/h_T \right| \\ &\leq \frac{B}{T-m+1} \sum_{j=1}^{T-m+1} g_j(\varepsilon_T, c_T)/h_T, \end{aligned} \tag{8.13}$$

where $k'(\cdot)$ denotes the derivative of $k(\cdot)$, $\widetilde{P}_j$ lies between $P_j(\widehat{\gamma}_{(j)})$ and $P_j(\gamma_0)$, $B < \infty$ denotes the bound on the derivative of $k(\cdot)$, and the inequality holds by (8.6).

By the dominated convergence theorem,

$$Eg_1(\varepsilon_T, c_T)/h_T \to 0 \text{ as } T \to \infty, \tag{8.14}$$

using the moment conditions in Assumption 2 and the fact that $c_T\varepsilon_T/h_T \to 0$ and $\varepsilon_T/h_T \to 0$ by the definitions of h_T, c_T, and ε_T.

We now have

$$\begin{aligned} &\Pr(D_{2,T} > \delta) \\ &\leq \Pr(\{D_{2,T} > \delta\} \cap L_T) + \Pr(\overline{L}_T) \\ &\leq \Pr\left(\frac{B}{T-m+1} \sum_{j=1}^{T-m+1} g_j(\varepsilon_T, c_T)/h_T > \delta \right) + o(1) \\ &\leq \delta^{-1} B E g_1(\varepsilon_T, c_T)/h_T + o(1) \\ &= o(1), \end{aligned} \tag{8.15}$$

where the second inequality holds by (8.5) and (8.13), the third inequality holds by Markov's inequality and the identical distributions of $\{g_j(\varepsilon_T, c_T) : j = 1, ..., T-m+1\}$, and the equality holds by (8.14). This completes the proof of part (b).

Part (c) is implied by part (b) using Assumption 5. This is a standard result. It follows from the fact that for all small $\varepsilon > 0$, $\widehat{F}_{P_a,T}(q_{P,1-\alpha} - \varepsilon) \to_p F_P(q_{P,1-\alpha} - \varepsilon) < 1 - \alpha$ and $\widehat{F}_{P_a,T}(q_{P,1-\alpha} + \varepsilon) \to_p F_P(q_{P,1-\alpha} + \varepsilon) > 1 - \alpha$.

Part (d) is implied by parts (a) and (c) using Assumption 5.

This completes the proof for the case where $v = a$.

The proofs of parts (a)-(d) of the Theorem for $v = b$ and c are essentially the same as that for $v = a$ because Assumption 4 implies that the estimators $(\widehat{\gamma}_{1\text{-}(T+\lceil m/2 \rceil-1)}, \widehat{\gamma}_{(j)})$ and $(\widehat{\gamma}_{1\text{-}(T+m)}, \widehat{\gamma}_{2(j)})$ behave like $(\widehat{\gamma}_{1\text{-}T}, \widehat{\gamma}_{(j)})$ asymptotically.

Part (e) holds by altering the proofs of parts (a)-(d) given above. Let C be an $m \times m$ such that $C'C = A_m$. Then,

$$P_j(\gamma, A_m) = (C\mathbf{Y}_{j\text{-}(j+m-1)} - C\mathbf{Z}_{j\text{-}(j+m-1)}\gamma)'(C\mathbf{Y}_{j\text{-}(j+m-1)} - \mathbf{C}\mathbf{Z}_{j\text{-}(j+m-1)}\gamma). \tag{8.16}$$

Define

$$(\widetilde{u}_j, ..., \widetilde{u}_{j+m-1})' = \widetilde{\mathbf{U}}_{j\text{-}(j+m-1)} = C\mathbf{U}_{j\text{-}(j+m-1)},$$

$$
\begin{aligned}
(\widetilde{z}_j, ..., \widetilde{z}_{j+m-1})' &= \widetilde{\mathbf{Z}}_{j\text{-}(j+m-1)} = C\mathbf{Z}_{j\text{-}(j+m-1)}, \\
\widetilde{z}_t &= (\widetilde{z}_{1,t}', \widetilde{z}_{2,t}')', \qquad (8.17)
\end{aligned}
$$

where $\widetilde{z}_{1,t} \in R^{k_1}$. By construction, for $t = j, ..., j + m - 1$,

$$
\begin{aligned}
\widetilde{z}_{1,t} &= \widetilde{\mathbf{Z}}_{1,j\text{-}(j+m-1)}' c_{t-j+1}, \text{ where} \\
C &= (c_1, ..., c_m)', \qquad (8.18)
\end{aligned}
$$

$c_j \in R^m$ for $j = 1, ..., m$, and $\widetilde{\mathbf{Z}}_{1,j\text{-}(j+m-1)}$ denotes the first k_1 columns of $\widetilde{\mathbf{Z}}_{j\text{-}(j+m-1)}$.

An analogue of (8.6) holds with $P_j(\gamma)$ replaced by $P_j(\gamma, A_m)$ by replacing u_t, z_t, $z_{2,t}$, and $\max_{s \leq T+m} ||B_T^{-1} z_{1,s}||$ by $\widetilde{u}_t$, $\widetilde{z}_t$, $\widetilde{z}_{2,t}$, and $m^2 \max_{s \leq T+m} ||B_T^{-1} z_{1,s}||$, respectively, provided $L_{2T}(c)$ is defined with c replaced by c/m^2 on the right-hand side of (8.3). This holds because

$$
\begin{aligned}
& \max_{j=1,...T-m+1} \max_{t=j,...,j+m-1} ||B_T^{-1} \widetilde{z}_{1,t}|| \\
=\; & \max_{j=1,...T-m+1} \max_{t=j,...,j+m-1} ||B_T^{-1} (z_{1,j}, ..., z_{1,j+m-1}) c_{t-j+1}|| \\
\leq\; & m^2 \max_{s \leq T+m} ||B_T^{-1} z_{1,s}||, \qquad (8.19)
\end{aligned}
$$

where the inequality uses the fact that the elements of C are all less than or equal to m in absolute value.

In the present case, $g_j(\varepsilon_T, c_T)$ is defined as in the last equality of (8.6) but with u_t and $z_{2,t}$ replaced by $\widetilde{u}_t$ and $\widetilde{z}_{2,t}$, respectively. Given this definition of $g_j(\varepsilon_T, c_T)$, the rest of the proofs of parts (a)-(d) hold without change when $P_j(\gamma)$ is replaced by $P_j(\gamma, A_m)$. This completes the proof of part (e). $\square$

Proof of Lemma 1. We start by showing $B_T(\widehat{\gamma}_{1,1\text{-}T} - \gamma_{1,0}) \rightarrow_p 0$. First, note that

$$
T^{-1} \sum_{t=1}^{T} B_T^{-1} z_{1,t} z_{1,t}' B_T^{-1} = \int_0^1 \nu_{1,T}(r) \nu_{1,T}(r)' dr \qquad (8.20)
$$

by definition of $\nu_T(r)$. Let $\nu_{1,2,T} = T^{-1} \sum_{t=1}^{T} B_T^{-1} z_{1,t} z_{2,t}'$. By the partitioned regression formula,

$$
\begin{aligned}
& B_T(\widehat{\gamma}_{1,1\text{-}T} - \gamma_{1,0}) \\
=\; & \left(\int_0^1 \nu_{1,T}(r) \nu_{1,T}(r)' dr - \nu_{1,2,T} (T^{-1} \sum_{t=1}^{T} z_{2,t} z_{2,t}')^{-1} \nu_{1,2,T}' \right)^{-1} \\
& \times \left(T^{-1} \sum_{t=1}^{T} B_T^{-1} z_{1,t} u_t - \nu_{1,2,T} (T^{-1} \sum_{t=1}^{T} z_{2,t} z_{2,t}')^{-1} T^{-1} \sum_{t=1}^{T} z_{2,t} u_t) \right) \\
=\; & o_p(1), \qquad (8.21)
\end{aligned}
$$

where the second equality holds because (i) $\nu_{1,2,T} \rightarrow_p 0$ by Assumption LS(c); (ii) $T^{-1} \sum_{t=1}^{T} z_{2,t} z_{2,t}' \rightarrow_p \Sigma_{2,0} > 0$ by the ergodic theorem and Assumptions 1, LS(a), and LS(e); (iii) the integral converges in distribution to $\int_0^1 \nu_1(r)\nu_1(r)'dr$ (which is positive definite a.s. by Assumption LS(e)) by Assumption LS(b) and the continuous mapping theorem; (iv) $T^{-1}\sum_{t=1}^T B_T^{-1}z_{1,t}u_t \to_p 0$ by Assumption LS(c); and (v) $T^{-1}\sum_{t=1}^T z_{2,t}u_t \to_p 0$ by the ergodic theorem and Assumptions 1 and LS(a).

Similarly, we have

$$\begin{aligned}
&\widehat{\gamma}_{2,1\text{-}T} - \gamma_{2,0} \\
&= \left(T^{-1}\sum_{t=1}^T z_{2,t}z_{2,t}' - \nu_{1,2,T}'\left(\int_0^1 \nu_{1,T}(r)\nu_{1,T}(r)'dr\right)^{-1}\nu_{1,2,T}\right)^{-1} \\
&\quad \times \left(T^{-1}\sum_{t=1}^T z_{2,t}u_t - \nu_{1,2,T}'\left(\int_0^1 \nu_{1,T}(r)\nu_{1,T}(r)'dr\right)^{-1} T^{-1}\sum_{t=1}^T B_T^{-1}z_{1,t}u_t)\right) \\
&= o_p(1). \qquad (8.22)
\end{aligned}$$

Next, to obtain the properties specified in Assumption 4 for $\widehat{\gamma}_{\ell,(j)}$ for $\ell = 1, 2$, it suffices to show that

$$\begin{aligned}
K_{1,T} &= \max_{j=1,\ldots,T+m-1} \|T^{-1}\sum_{t=j}^{j+m-1} B_T^{-1}z_{1,t}z_{1,t}'B_T^{-1}\| = o_p(1), \\
K_{2,T} &= \max_{j=1,\ldots,T+m-1} \|T^{-1}\sum_{t=j}^{j+m-1} z_{2,t}z_{2,t}'\| = o_p(1), \\
K_{3,T} &= \max_{j=1,\ldots,T+m-1} \|T^{-1}\sum_{t=j}^{j+m-1} z_{2,t}u_t\| = o_p(1), \\
K_{4,T} &= \max_{j=1,\ldots,T+m-1} \|T^{-1}\sum_{t=j}^{j+m-1} B_T^{-1}z_{1,t}z_{2,t}'\| = o_p(1), \\
K_{5,T} &= \max_{j=1,\ldots,T+m-1} \|T^{-1}\sum_{t=j}^{j+m-1} B_T^{-1}z_{1,t}u_t\| = o_p(1), \qquad (8.23)
\end{aligned}$$

These conditions are sufficient because (8.21) and (8.22) show that the differences between $B_T(\widehat{\gamma}_{\ell,1\text{-}T} - \gamma_{\ell,0})$ and $B_T(\widehat{\gamma}_{\ell,(j)} - \gamma_{\ell,0})$ are captured by the terms in (8.23).

We have

$$K_{1,T} \leq T^{-1}m \sup_{r\in[0,1]} \|\nu_T(r)\|^2 = T^{-1}O_p(1) = o_p(1), \qquad (8.24)$$

where the first equality holds by Assumption LS(b) and the continuous mapping theorem.

To establish the conditions of (8.23) for $K_{2,T}$-$K_{5,T}$, we use the following result. Suppose that $\{\xi_t : t \geq 1\}$ is a sequence of mean zero random variables and $\sup_{t\geq 1} E\|\xi_t\|^{1+\delta} < \infty$ for some $\delta > 0$. Let $\tau_j = \sum_{t=j}^{j+m-1}\xi_t$. Then, for all $\varepsilon > 0$,

$$\Pr(T^{-1}\max_{j\leq T-m+1}\|\tau_j\| > \varepsilon) = \Pr(\cup_{j=1}^{T-m+1}\{\|\tau_j\| > T\varepsilon\})$$

$$
\begin{aligned}
&\leq \sum_{j=1}^{T-m+1} \Pr(||\tau_j|| > T\varepsilon) \\
&\leq (T-m+1)E||\tau_j||^{1+\delta}T^{-(1+\delta)}\varepsilon^{-(1+\delta)} \\
&= o(1),
\end{aligned} \tag{8.25}
$$

where the second inequality uses Markov's inequality. Hence,

$$
\max_{j \leq T-m+1} ||T^{-1} \sum_{t=j}^{j+m-1} \xi_t|| \rightarrow_p 0. \tag{8.26}
$$

Applying (8.26) with $\xi_t = z_{2,t}z_{2,t}' - Ez_{2,t}z_{2,t}'$ gives $K_{2,T} \rightarrow_p 0$ using the facts that $E||z_{2,t}||^{2+\delta} < \infty$ by Assumption LS(a) and $Ez_{2,t}z_{2,t}'$ does not depend on t or T for $t \leq T$ by Assumption 1. Applying (8.26) with $\xi_t = z_{2,t}u_t$ gives $K_{3,T} \rightarrow_p 0$ using the fact that $E||z_{2,t}u_t||^{1+\delta} < \infty$ by Assumption LS(a).

For $K_{4,T}$, we have

$$
K_{4,T} \leq \max_{j=1,\ldots,T+m-1} T^{-1} \sum_{t=j}^{j+m-1} ||z_{2,t}|| \cdot \max_{r \in [0,1]} ||\nu_T(r)|| = o_p(1), \tag{8.27}
$$

where the equality holds by Assumption LS(b) and by applying (8.26) with $\xi_t = ||z_{2,t}||$ using Assumptions 1 and LS(a). An analogous argument with u_t in place of $z_{2,t}$ gives $K_{5,T} \rightarrow_p 0$ because $E|u_t|^{1+\delta} < \infty$ by Assumption LS(a). This completes the proof for $\widehat{\gamma}_{\ell,(j)}$ for $\ell = 1, 2$.

The properties specified in Assumption 4 for $\widehat{\gamma}_{\ell,2(j)}$ for $\ell = 1, 2$ hold by essentially the same argument as for $\widehat{\gamma}_{\ell,(j)}$.

To obtain the properties specified in Assumption 4 for $\widehat{\gamma}_{\ell,1\text{-}(T+m)}$ for $\ell = 1, 2$, it suffices to show that the conditions in (8.23) hold with the sums being over $t = T+1, ..., T+m$, rather than $t = j, ..., j+m-1$, and with the max over j deleted. These conditions are sufficient because (8.21) and (8.22) show that the differences between $B_T(\widehat{\gamma}_{\ell,1\text{-}T} - \gamma_{\ell,0})$ and $B_T(\widehat{\gamma}_{\ell,1\text{-}(T+m)} - \gamma_{\ell,0})$ are captured by the terms in (8.23) with the adjustments just described.

Let $K_{i,T}^* = o_p(1)$ for $i = 1, ..., 5$ denote the conditions in (8.23) with these changes. We have

$$
K_{1,T}^* \leq T^{-1}m \max_{t=T+1,\ldots,T+m} ||B_T^{-1}z_{1,t}||^2 = o_p(1), \tag{8.28}
$$

where the equality holds by Assumption LS(d). Next, we have, for all $\varepsilon > 0$,

$$
\Pr(||T^{-1} \sum_{i=1}^{m} z_{2,T+i}z_{2,T+i}'|| > \varepsilon) = \Pr(|| \sum_{i=1}^{m} z_{2,T+i}z_{2,T+i}'|| > T\varepsilon) = o(1), \tag{8.29}
$$

where the equality holds because the distribution of $\sum_{i=1}^{m} ||z_{2,T+i}z_{2,T+i}'||$ does not depend on T by Assumption 1. Hence, $K_{2,T}^* = o_p(1)$. An analogous argument with $z_{2,t}z_{2,t}'$ replaced by $z_{2,t}u_t$ gives $K_{3,T}^* = o_p(1)$.

Next, we have

$$K_{4,T}^{*} \leq T^{-1} \sum_{t=T+1}^{T+m} ||z_{2,t}|| \cdot \max_{r \in [0,1]} ||\nu_T(r)|| = o_p(1), \tag{8.30}$$

where the equality holds by Assumption LS(b) and the argument in (8.29). An analogous argument with $||z_{2,t}||$ replaced by $|u_t|$ gives $K_{5,T}^{*} = o_p(1)$. This completes the proof for $\widehat{\gamma}_{\ell,1\text{-}(T+m)}$.

The properties specified in Assumption 4 for $\widehat{\gamma}_{\ell,1\text{-}(T+\lceil m/2 \rceil)}$ for $\ell = 1, 2$ are established by the same argument as just given for $\widehat{\gamma}_{\ell,1\text{-}(T+m)}$ but with m replaced by $\lceil m/2 \rceil$. Hence, Assumption 4 holds.

Assumption LS(a) obviously implies Assumption 2.

Finally, Assumption LS(b) and the continuous mapping theorem imply that

$$\max_{t \leq T} ||B_T^{-1} z_{1,t}|| = \sup_{r \in [0,1]} ||\nu_T(r)|| \rightarrow_d \sup_{r \in [0,1]} ||\nu(r)|| < \infty \text{ a.s.} \tag{8.31}$$

This result, combined with Assumption LS(d), establishes Assumption 3. □

Footnotes

[1] The first author gratefully acknowledges the research support of the National Science Foundation via grant number SES-0001706.

[2] When $m > k$, the F statistic is based on the projection of the post-breakdown residual vector on the post-breakdown regressor matrix. One can define a test statistic S_a that corresponds to this. Andrews (2002) does so for linear and nonlinear models with stationary observations. In the present context, however, the parametric subsampling critical values we use for the tests does not deliver an asymptotically valid critical value because the regressors are not stationary and the statistic S_a depends on the regressors.

[3] Alternatively, one could take this element of B_T to be $(T+m)^{1/2}$. The choices $T^{1/2}$ and $(T+m)^{1/2}$ are equivalent because m does not depend on T. We choose $T^{1/2}$ for notational simplicity.

[4] The reason that conditions (i) and (ii) are sufficient for Assumption LS(d) is that Assumption LS(b) and the continuous mapping theorem imply that $\sup_{r \in [0,1]} ||\nu_T(r)||$ $\rightarrow_d \sup_{r \in [0,1]} ||\nu(r)|| < \infty$ a.s. The left-hand side equals $\max_{t \leq T} ||B_T^{-1} z_{1,t}||$. If $z_{1,t}$ comes from a sequence, this implies that $\max_{t \leq T+m} ||B_{T+m}^{-1} z_{1,t}|| = O_p(1)$. Combined with condition (ii), this yields Assumption LS(d).

References

Ahn, S. K. and G. C. Reinsel (1990): "Estimation for Partially Nonstationary Autoregressive Models," *Journal of the American Statistical Association,* 85, 813-823.

Andrews, D. W. K. (1988): "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables," *Econometric Theory,* 4, 458-467.

——— (2002): "End-of-sample Instability Tests," Cowles Foundation Discussion Paper No. 1369, Yale University. Available at http://cowles.econ.yale.edu.

Chow, G. C. (1960): "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica,* 28, 591-605.

Dufour, J.-M., E. Ghysels, and A. Hall (1994): "Generalized Predictive Tests and Structural Change Analysis in Econometrics," *International Economic Review,* 35, 199-229.

Ferguson, T. S. (1967): *Mathematical Statistics: A Decision Theoretic Approach.* New York: Academic Press.

Hansen, B. E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics,* 10, 321-336.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control,* 12, 231-255.

——— (1991): "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregression Models," *Econometrica,* 59, 1551-1580.

Kim, J.-Y. (1999): "Testing Structural Change in a Cointegration Relation," unpublished manuscript, SUNY Albany.

——— (2000): "Detection of Change in Persistence of Linear Time Series," *Journal of Econometrics,* 95, 97-116.

King, M. L. and G. H. Hillier (1985): "Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression Model," *Journal of the Royal Statistical Society B,* 47, 98-102.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics,* 54, 159-178.

Leybourne, S. J. and B. P. M. McCabe (1989): "On the Distribution of Some Test Statistics for Coefficient Constancy," *Biometrika,* 76, 169-177.

Nabeya, S. and K. Tanaka (1988): "Asymptotic Theory of a Test for the Constancy of Regression Coefficients Against the Random Walk Alternative," *Annals of Statistics,* 16, 218-235.

Nyblom, J. (1986): "Testing for Deterministic Linear Trend in Time Series," *Journal of the American Statistical Association,* 81, 545-549.

——— (1989): "Testing for the Consistency of Parameters Over Time," *Journal of the American Statistical Association,* 84, 223-230.

Nyblom, J. and T. Makelainen (1983): "Comparisons of tests for the Presence Random Walk Coefficients in a Simple Linear Model," *Journal of the American Statistical Association,* 78, 856-864.

Park, J. Y. (1992): "Canonical Cointegrating Regressions," *Econometrica,* 60, 119-144.

Phillips, P. C. B. (1988a): "Weak Convergence of Sample Covariance Matrices to Stochastic Integrals Via Martingale Approximations," *Econometric Theory,* 4, 528-533.

——— (1988b): "Weak Convergence to the Matrix Stochastic Integral $\int_0^1 BdB'$," *Journal of Multivariate Analysis,* 24, 252-264.

——— (1991): "Optimal Inference in Cointegrated Systems," *Econometrica,* 59, 283-306.

Phillips, P. C. B. and B. E. Hansen (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies,* 57, 99-125.

Phillips, P. C. B. and M. Loretan (1991): "Estimating Long Run Economic Equilibria," *Review of Economic Studies,* 58, 407-436.

Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling.* Springer Series in Statistics. New York: Springer.

Pollard, D. (1984): *Convergence of Stochastic Processes.* New York: Springer.

Quintos, C. E. (1997): "Stability Tests in Error Correction Models," *Journal of Econometrics,* 82, 289-315.

Quintos, C. E. and P. C. B. Phillips (1993): "Parameter Constancy in Cointegrating Regressions," *Empirical Economics,* 18, 675-706.

Saikonen, P. (1991): "Asymptotically Efficient Estimation of Cointegrating Regression," *Econometric Theory,* 7, 1-21.

Scheffé, H. (1959): *The Analysis of Variance.* New York: Wiley.

Shin, Y. (1994): "A Residual-based Test of the Null of Cointegration Against the Alternative of No Cointegration," *Econometric Theory,* 10, 91-115.

Sowell, F. (1990): "The Fractional Unit Root Distribution," *Econometrica,* 58, 495-505.

Stock, J. H. and M. Watson (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica,* 61, 783-820.

Tanaka, K. (1993): "An Alternative Approach to the Asymptotic Theory of Spurious Regression, Cointegration, and Near Cointegration," *Econometric Theory,* 9, 36-61.

Table I

Average and Range of Null Rejection Rates Over Eight Models for Nominal .05 Tests

m	ρ	Test	$T = 100$ Avg	$T = 100$ Range	$T = 250$ Avg	$T = 250$ Range
Avg		P_a	.113	[.060, .251]	.074	[.052, .127]
Over		P_b	.059	[.029, .090]	.054	[.040, .067]
All		P_c	.052	[.028, .081]	.051	[.040, .064]
Nine						
(m, ρ)		R_a	.113	[.017, .257]	.074	[.033, .134]
Values		R_b	.058	[.017, .097]	.053	[.033, .073]
		R_c	.046	[.020, .077]	.047	[.032, .065]
10	0	P_a	.113	[.094, .139]	.074	[.069, .081]
		P_b	.067	[.050, .084]	.059	[.053, .065]
		P_c	.064	[.046, .081]	.057	[.049, .064]
		R_a	.087	[.017, .123]	.065	[.033, .079]
		R_b	.058	[.017, .072]	.054	[.033, .063]
		R_c	.034	[.020, .045]	.041	[.032, .047]
10	4	P_a	.136	[.114, .166]	.082	[.074, .091]
		P_b	.066	[.040, .082]	.056	[.043, .065]
		P_c	.056	[.041, .075]	.051	[.042, .063]
		R_a	.139	[.109, .168]	.086	[.075, .097]
		R_b	.062	[.019, .071]	.056	[.033, .061]
		R_c	.038	[.026, .051]	.042	[.033, .050]
10	8	P_a	.202	[.160, .251]	.106	[.093, .127]
		P_b	.078	[.029, .090]	.060	[.040, .067]
		P_c	.054	[.044, .066]	.048	[.040, .057]
		R_a	.207	[.159, .257]	.110	[.093, .134]
		R_b	.084	[.024, .097]	.056	[.040, .073]
		R_c	.053	[.041, .067]	.050	[.044, .059]

Table I (cont.)

m	ρ	Test	$T = 100$ Avg	$T = 100$ Range	$T = 250$ Avg	$T = 250$ Range
5	0	P_a	.076	[.060, .090]	.062	[.058, .067]
		P_b	.047	[.039, .052]	.051	[.046, .055]
		P_c	.045	[.034, .051]	.049	[.043, .054]
		R_a	.085	[.071, .103]	.065	[.059, .074]
		R_b	.042	[.025, .051]	.048	[.038, .056]
		R_c	.032	[.024, .038]	.042	[.036, .047]
5	4	P_a	.090	[.071, .110]	.068	[.060, .076]
		P_b	.043	[.032, .049]	.049	[.044, .054]
		P_c	.037	[.028, .046]	.046	[.040, .052]
		R_a	.097	[.081, .115]	.070	[.064, .079]
		R_b	.044	[.027, .048]	.049	[.040, .053]
		R_c	.034	[.026, .042]	.044	[.039, .049]
5	8	P_a	.147	[.125, .180]	.088	[.079, .103]
		P_b	.060	[.039, .067]	.057	[.047, .060]
		P_c	.047	[.039, .057]	.052	[.047, .060]
		R_a	.155	[.128, .190]	.090	[.081, .107]
		R_b	.066	[.041, .072]	.060	[.049, .066]
		R_c	.051	[.043, .061]	.053	[.051, .057]
1	0	P_a $(= R_a)$	.067	[.060, .075]	.054	[.052, .062]
		P_b $(= P_c = R_b = R_c)$	.044	[.035, .048]	.045	[.042, .049]
1	4	P_a $(= R_a)$	.074	[.069, .081]	.058	[.061, .061]
		P_b $(= P_c = R_b = R_c)$	.050	[.044, .053]	.048	[.047, .049]
1	8	P_a $(= R_a)$	.110	[.101, .122]	.072	[.068, .078]
		P_b $(= P_c = R_b = R_c)$	.074	[.072, .077]	.062	[.060, .065]

Table II

Average of Rejection Rates Over Eight Models for Size-corrected .05 Tests

			Unit Root Alternative		Parameter Shift Alternative	
m	ρ	Test	$T = 100$	$T = 250$	$T = 100$	$T = 250$
Avg		P_a	.59	.61	.69	.81
Over		P_b	.58	.61	.65	.80
All		P_c	.58	.61	.63	.80
Nine						
(m, ρ)		R_a	.55	.58	.70	.82
Values		R_b	.54	.58	.65	.81
		R_c	.56	.59	.67	.82
10	0	P_a	.88	.88	.68	.80
		P_b	.87	.88	.57	.77
		P_c	.82	.87	.50	.75
		R_a	.76	.83	.71	.85
		R_b	.76	.83	.63	.84
		R_c	.80	.85	.65	.84
10	4	P_a	.89	.90	.75	.82
		P_b	.86	.90	.65	.84
		P_c	.87	.91	.61	.83
		R_a	.78	.82	.74	.85
		R_b	.77	.82	.64	.85
		R_c	.78	.84	.68	.85
10	8	P_a	.87	.86	.84	.89
		P_b	.82	.88	.73	.89
		P_c	.85	.89	.74	.89
		R_a	.80	.80	.82	.88
		R_b	.74	.81	.68	.87
		R_c	.75	.81	.69	.88

Table II (cont.)

m	ρ	Test	Unit Root Alternative $T=100$	Unit Root Alternative $T=250$	Parameter Shift Alternative $T=100$	Parameter Shift Alternative $T=250$
5	0	P_a	.66	.70	.61	.77
		P_b	.65	.70	.55	.75
		P_c	.62	.69	.54	.75
		R_a	.62	.68	.66	.81
		R_b	.63	.68	.63	.80
		R_c	.64	.69	.65	.80
5	4	P_a	.66	.72	.68	.82
		P_b	.68	.72	.67	.82
		P_c	.67	.73	.64	.81
		R_a	.62	.67	.69	.83
		R_b	.62	.67	.66	.82
		R_c	.63	.67	.68	.83
5	8	P_a	.68	.68	.81	.88
		P_b	.66	.70	.79	.88
		P_c	.66	.68	.77	.88
		R_a	.65	.66	.81	.87
		R_b	.62	.67	.75	.87
		R_c	.62	.67	.76	.87
1	0	P_a	.23	.25	.46	.65
		P_b	.23	.25	.46	.65
1	4	P_a	.23	.24	.61	.76
		P_b	.23	.24	.62	.76
1	8	P_a	.25	.24	.80	.87
		P_b	.25	.24	.80	.87

Table A-I

True Size of Nominal .05 Tests for Models with a Constant, Time Trend, Two Unit Root Regressors, Two Stationary Regressors and Innovation Distributions Given by (i) Normal, (ii) χ^2_2, (iii) t_3, and (iv) Uniform

			Normal		χ^2_2		t_3		Uniform	
			T		T		T		T	
m	ρ	Test	100	250	100	250	100	250	100	250
10	0	P_a	.107	.073	.099	.069	.104	.071	.118	.078
		P_b	.066	.059	.081	.064	.084	.065	.054	.054
		P_c	.061	.056	.078	.062	.081	.064	.046	.049
		R_a	.120	.078	.119	.078	.121	.076	.123	.079
		R_b	.059	.055	.064	.057	.065	.057	.061	.055
		R_c	.031	.039	.037	.042	.041	.045	.029	.038
10	.4	P_a	.137	.082	.116	.074	.123	.078	.142	.089
		P_b	.067	.056	.077	.063	.082	.065	.061	.054
		P_c	.052	.048	.071	.060	.075	.062	.041	.042
		R_a	.140	.085	.134	.084	.139	.082	.137	.088
		R_b	.067	.058	.066	.059	.071	.058	.065	.060
		R_c	.034	.041	.038	.043	.043	.045	.035	.043
10	.8	P_a	.202	.109	.195	.097	.191	.100	.207	.110
		P_b	.082	.063	.088	.060	.089	.067	.084	.063
		P_c	.048	.047	.062	.050	.066	.057	.048	.046
		R_a	.208	.111	.207	.108	.205	.110	.207	.112
		R_b	.090	.069	.093	.065	.096	.071	.092	.069
		R_c	.051	.050	.056	.048	.057	.055	.052	.050

Table A-I (cont.)

			Normal		χ^2_2		t_3		Uniform	
			T		T		T		T	
m	ρ	Test	100	250	100	250	100	250	100	250
5	0	P_a	.075	.061	.063	.058	.060	.060	.087	.065
		P_b	.047	.051	.052	.053	.048	.055	.039	.046
		P_c	.044	.049	.051	.052	.048	.054	.034	.043
		R_a	.082	.065	.081	.061	.078	.062	.084	.065
		R_b	.043	.048	.048	.050	.048	.050	.040	.048
		R_c	.029	.040	.037	.045	.037	.045	.027	.039
5	.4	P_a	.092	.069	.071	.060	.074	.064	.102	.073
		P_b	.044	.050	.047	.052	.049	.054	.041	.045
		P_c	.036	.044	.045	.050	.046	.052	.028	.040
		R_a	.097	.071	.091	.068	.092	.068	.100	.070
		R_b	.047	.051	.048	.052	.048	.053	.045	.047
		R_c	.033	.044	.037	.047	.037	.049	.031	.041
5	.8	P_a	.150	.091	.138	.082	.134	.079	.154	.091
		P_b	.063	.060	.064	.058	.067	.058	.062	.057
		P_c	.045	.052	.051	.052	.053	.052	.041	.048
		R_a	.156	.093	.150	.087	.147	.085	.158	.092
		R_b	.070	.063	.069	.060	.072	.060	.070	.061
		R_c	.050	.055	.052	.053	.055	.053	.049	.053
1	0	P_a $(= R_a)$	.067	.053	.060	.052	.061	.054	.075	.062
		P_b $(= P_c = R_b = R_c)$	.045	.044	.048	.049	.047	.049	.035	.042
1	.4	P_a $(= R_a)$	.073	.057	.069	.054	.070	.055	.081	.061
		P_b $(= P_c = R_b = R_c)$	.048	.048	.053	.049	.052	.049	.044	.047
1	.8	P_a $(= R_a)$	.112	.072	.106	.069	.103	.069	.117	.077
		P_b $(= P_c = R_b = R_c)$	.075	.061	.073	.060	.072	.060	.077	.064

Table A-II

True Size of Nominal .05 Tests for Models with Normal Innovation Distributions and (i) $\rho_{Dx,u} = .4$, (ii) $\rho_{Dx,u} = .8$, (iii) No Stationary Regressors, and (iv) No Time Trend

			$\rho_{Dx,u} = .4$		$\rho_{Dx,u} = .8$		No Stat. Regr.		No Time Trend	
			T		T		T		T	
m	ρ	Test	100	250	100	250	100	250	100	250
10	0	P_a	.119	.076	.139	.081	.125	.078	.094	.069
		P_b	.068	.060	.071	.061	.065	.060	.050	.053
		P_c	.063	.057	.061	.057	.059	.056	.066	.059
		R_a	.132	.084	.154	.092	.144	.089	.093	.068
		R_b	.063	.058	.072	.063	.059	.054	.017	.033
		R_c	.031	.040	.034	.044	.020	.032	.045	.047
10	.4	P_a	.141	.084	.148	.084	.166	.091	.114	.075
		P_b	.068	.056	.069	.057	.067	.055	.040	.043
		P_c	.051	.046	.051	.048	.049	.045	.061	.054
		R_a	.141	.087	.146	.089	.168	.097	.109	.075
		R_b	.069	.059	.070	.061	.066	.058	.019	.033
		R_c	.037	.041	.036	.043	.026	.033	.051	.050
10	.8	P_a	.204	.106	.204	.106	.251	.127	.160	.093
		P_b	.082	.061	.083	.060	.090	.064	.029	.040
		P_c	.049	.045	.049	.045	.044	.040	.063	.056
		R_a	.207	.108	.207	.107	.257	.134	.159	.093
		R_b	.091	.066	.090	.067	.097	.073	.024	.040
		R_c	.052	.048	.050	.048	.041	.044	.067	.059

Table A-II (cont.)

			$\rho_{Dx,u}$ = .4		$\rho_{Dx,u}$ = .8		No Stat. Regr.		No Time Trend	
			T		T		T		T	
m	ρ	Test	100	250	100	250	100	250	100	250
5	0	P_a	.080	.063	.090	.067	.084	.064	.069	.059
		P_b	.048	.051	.050	.054	.048	.051	.040	.048
		P_c	.044	.049	.044	.050	.043	.047	.049	.052
		R_a	.090	.068	.103	.074	.094	.068	.071	.059
		R_b	.045	.051	.051	.056	.039	.046	.025	.038
		R_c	.030	.043	.035	.047	.024	.036	.038	.045
5	.4	P_a	.092	.067	.096	.069	.110	.076	.080	.065
		P_b	.042	.049	.043	.048	.043	.048	.032	.044
		P_c	.035	.045	.035	.044	.032	.042	.043	.049
		R_a	.097	.070	.100	.071	.115	.079	.081	.064
		R_b	.046	.051	.047	.050	.043	.050	.027	.040
		R_c	.033	.043	.034	.042	.026	.039	.042	.048
5	.8	P_a	.150	.088	.147	.087	.180	.103	.125	.081
		P_b	.062	.057	.061	.056	.065	.060	.039	.047
		P_c	.045	.049	.044	.049	.039	.049	.057	.055
		R_a	.157	.091	.153	.088	.190	.107	.128	.081
		R_b	.069	.060	.067	.059	.072	.066	.041	.049
		R_c	.050	.052	.048	.051	.043	.052	.061	.057
1	0	P_a $(= R_a)$	.067	.055	.070	.053	.071	.054	.065	.052
		P_b $(= P_c = R_b = R_c)$	.044	.045	.046	.045	.043	.044	.048	.045
1	.4	P_a $(= R_a)$	.077	.059	.078	.059	.078	.060	.069	.056
		P_b $(= P_c = R_b = R_c)$	.051	.049	.050	.049	.048	.048	.051	.049
1	.8	P_a $(= R_a)$	.109	.074	.109	.072	.122	.078	.101	.068
		P_b $(= P_c = R_b = R_c)$	.073	.062	.072	.061	.075	.065	.076	.061

Table A-III

Power of Significance Level .05 Size-corrected Tests Against **Unit Root Alternatives** for Models with Innovation Distributions Given by (i) Normal, (ii) χ^2_2, (iii) t_3, and (iv) Uniform

			Normal		χ^2_2		t_3		Uniform	
			T		T		T		T	
m	ρ	Test	100	250	100	250	100	250	100	250
10	0	P_a	.91	.93	.80	.79	.72	.64	.95	.97
		P_b	.90	.93	.74	.76	.65	.61	.95	.97
		P_c	.87	.92	.66	.78	.57	.58	.95	.97
		R_a	.78	.84	.77	.83	.72	.78	.78	.85
		R_b	.79	.84	.76	.83	.71	.79	.79	.85
		R_c	.82	.86	.80	.85	.75	.80	.83	.86
10	.4	P_a	.91	.93	.86	.86	.79	.73	.93	.95
		P_b	.89	.93	.81	.85	.73	.73	.93	.96
		P_c	.90	.93	.76	.86	.66	.77	.94	.96
		R_a	.79	.82	.78	.82	.74	.77	.79	.83
		R_b	.81	.83	.76	.83	.74	.79	.81	.83
		R_c	.77	.84	.79	.84	.73	.80	.78	.85
10	.8	P_a	.88	.88	.86	.85	.81	.76	.90	.90
		P_b	.83	.89	.81	.87	.75	.79	.84	.91
		P_c	.87	.90	.82	.88	.73	.80	.89	.92
		R_a	.81	.80	.81	.80	.76	.73	.82	.82
		R_b	.76	.81	.75	.81	.71	.76	.76	.83
		R_c	.76	.82	.74	.82	.69	.76	.77	.83

Table A-III (cont.)

m	ρ	Test	Normal T 100	Normal T 250	χ^2_2 T 100	χ^2_2 T 250	t_3 T 100	t_3 T 250	Uniform T 100	Uniform T 250
5	0	P_a	.71	.74	.51	.57	.43	.51	.81	.85
		P_b	.69	.74	.47	.55	.41	.48	.81	.85
		P_c	.68	.73	.42	.53	.37	.47	.81	.85
		R_a	.65	.69	.62	.67	.55	.60	.65	.71
		R_b	.65	.69	.64	.68	.55	.61	.67	.71
		R_c	.66	.70	.64	.68	.56	.62	.69	.72
5	.4	P_a	.68	.75	.60	.66	.50	.55	.74	.80
		P_b	.72	.75	.58	.65	.50	.57	.78	.82
		P_c	.71	.76	.54	.64	.46	.55	.79	.82
		R_a	.63	.68	.60	.66	.52	.58	.65	.70
		R_b	.64	.68	.62	.68	.55	.60	.67	.70
		R_c	.65	.68	.63	.68	.56	.59	.66	.70
5	.8	P_a	.69	.69	.65	.66	.56	.56	.74	.73
		P_b	.69	.71	.62	.69	.54	.59	.73	.76
		P_c	.69	.71	.63	.68	.54	.58	.73	.76
		R_a	.66	.66	.63	.64	.55	.57	.70	.71
		R_b	.63	.67	.61	.67	.51	.58	.68	.72
		R_c	.64	.68	.61	.66	.54	.59	.67	.72
1	0	P_a $(= R_a)$	.22	.25	.17	.16	.18	.19	.34	.36
		P_b $(= P_c = R_b = R_c)$	.23	.25	.17	.16	.17	.18	.33	.36
1	.4	P_a $(= R_a)$	.24	.25	.16	.19	.16	.16	.29	.31
		P_b $(= P_c = R_b = R_c)$	.23	.25	.19	.19	.18	.18	.30	.32
1	.8	P_a $(= R_a)$	.25	.25	.22	.21	.20	.18	.29	.28
		P_b $(= P_c = R_b = R_c)$	.26	.25	.21	.22	.19	.19	.29	.28

Table A-IV

Power of Significance Level .05 Size-corrected Tests Against **Unit Root Alternatives** for Models with Normal Innovation Distributions and (i) $\rho_{DX,U} = .4$, (ii) $\rho_{DX,U} = .8$, (iii) No Stationary Regressors, and (iv) No Time Trend

			$\rho_{Dx,u}$ = .4		$\rho_{Dx,u}$ = .8		No Stat. Regr.		No Time Trend	
			T		T		T		T	
m	ρ	Test	100	250	100	250	100	250	100	250
10	0	P_a	.91	.93	.90	.93	.90	.93	.91	.93
		P_b	.90	.93	.88	.93	.88	.92	.90	.93
		P_c	.87	.92	.88	.92	.84	.91	.90	.93
		R_a	.76	.82	.74	.81	.75	.82	.81	.85
		R_b	.76	.83	.73	.82	.74	.83	.76	.84
		R_c	.81	.85	.77	.84	.80	.85	.84	.86
10	.4	P_a	.91	.93	.90	.93	.89	.92	.92	.93
		P_b	.88	.93	.89	.93	.87	.93	.91	.94
		P_c	.90	.94	.90	.94	.87	.93	.91	.94
		R_a	.79	.83	.78	.82	.77	.81	.80	.84
		R_b	.77	.83	.76	.82	.74	.82	.73	.82
		R_c	.80	.84	.79	.84	.78	.84	.82	.85
10	.8	P_a	.87	.87	.87	.87	.86	.87	.89	.88
		P_b	.83	.89	.82	.89	.79	.88	.88	.91
		P_c	.87	.91	.86	.91	.85	.90	.87	.90
		R_a	.81	.80	.80	.80	.80	.80	.82	.82
		R_b	.75	.82	.75	.82	.72	.81	.70	.80
		R_c	.76	.82	.75	.82	.73	.81	.78	.83

Table A-IV (cont.)

m	ρ	Test	$\rho_{Dx,u}$ = .4 T=100	$\rho_{Dx,u}$ = .4 T=250	$\rho_{Dx,u}$ = .8 T=100	$\rho_{Dx,u}$ = .8 T=250	No Stat. Regr. T=100	No Stat. Regr. T=250	No Time Trend T=100	No Time Trend T=250
5	0	P_a	.69	.74	.69	.73	.68	.74	.72	.74
		P_b	.69	.74	.69	.73	.67	.73	.70	.73
		P_c	.68	.72	.68	.73	.62	.71	.70	.74
		R_a	.62	.68	.61	.69	.62	.68	.66	.70
		R_b	.64	.69	.62	.67	.62	.68	.62	.68
		R_c	.65	.69	.62	.68	.64	.69	.68	.70
5	.4	P_a	.68	.75	.68	.75	.67	.74	.71	.75
		P_b	.72	.72	.72	.75	.70	.75	.73	.76
		P_c	.71	.75	.71	.75	.68	.75	.73	.76
		R_a	.63	.70	.63	.68	.60	.66	.66	.69
		R_b	.64	.65	.63	.68	.62	.67	.61	.68
		R_c	.65	.67	.63	.68	.62	.67	.66	.69
5	.8	P_a	.69	.69	.69	.69	.69	.69	.69	.70
		P_b	.68	.71	.68	.71	.65	.71	.70	.71
		P_c	.69	.71	.68	.71	.66	.71	.68	.71
		R_a	.66	.67	.66	.67	.67	.65	.66	.68
		R_b	.64	.68	.64	.68	.63	.68	.59	.66
		R_c	.62	.67	.62	.68	.60	.66	.65	.68
1	0	P_a $(= R_a)$	.23	.25	.23	.25	.22	.25	.24	.26
		P_b $(= P_c = R_b = R_c)$	.24	.25	.25	.25	.23	.25	.23	.25
1	.4	P_a $(= R_a)$	.24	.25	.24	.25	.24	.25	.23	.25
		P_b $(= P_c = R_b = R_c)$	.23	.24	.23	.25	.23	.25	.24	.25
1	.8	P_a $(= R_a)$	.25	.25	.25	.25	.27	.26	.27	.25
		P_b $(= P_c = R_b = R_c)$	.26	.25	.26	.25	.26	.25	.26	.25

Table A-V

Power of Significance Level .05 Size-corrected Tests Against **Parameter Shift Alternatives** for Models with Innovation Distributions Given by (i) Normal, (ii) χ^2_2, (iii) t_3, and (iv) Uniform

			Normal		χ^2_2		t_3		Uniform	
			T		T		T		T	
m	ρ	Test	100	250	100	250	100	250	100	250
10	0	P_a	.69	.81	.60	.72	.58	.66	.72	.85
		P_b	.59	.79	.45	.68	.44	.61	.66	.84
		P_c	.50	.76	.35	.67	.34	.57	.61	.82
		R_a	.70	.84	.69	.84	.68	.83	.70	.85
		R_b	.64	.83	.61	.83	.61	.82	.64	.84
		R_c	.67	.84	.65	.84	.65	.83	.67	.84
10	.4	P_a	.75	.85	.71	.80	.69	.77	.76	.86
		P_b	.66	.84	.57	.79	.56	.74	.69	.86
		P_c	.62	.83	.48	.77	.47	.76	.67	.86
		R_a	.73	.84	.72	.85	.72	.84	.73	.85
		R_b	.63	.83	.63	.84	.65	.84	.64	.84
		R_c	.67	.84	.66	.84	.64	.84	.68	.84
10	.8	P_a	.83	.89	.82	.88	.81	.85	.83	.89
		P_b	.72	.89	.71	.88	.70	.85	.72	.89
		P_c	.74	.89	.70	.88	.67	.85	.75	.89
		R_a	.81	.87	.81	.87	.80	.85	.81	.88
		R_b	.70	.87	.67	.87	.67	.85	.68	.87
		R_c	.67	.87	.69	.87	.68	.86	.71	.87

Table A-V (cont.)

			Normal		χ_2^2		t_3		Uniform	
			T		T		T		T	
m	ρ	Test	100	250	100	250	100	250	100	250
5	0	P_a	.62	.77	.49	.69	.47	.70	.67	.81
		P_b	.56	.75	.42	.67	.40	.67	.64	.80
		P_c	.64	.74	.38	.65	.37	.66	.62	.80
		R_a	.66	.80	.64	.79	.62	.79	.65	.80
		R_b	.63	.80	.61	.79	.59	.78	.65	.80
		R_c	.64	.80	.63	.79	.62	.79	.64	.80
5	.4	P_a	.67	.82	.63	.79	.60	.77	.69	.83
		P_b	.66	.81	.58	.77	.56	.77	.69	.83
		P_c	.65	.81	.54	.77	.52	.76	.68	.83
		R_a	.68	.82	.66	.81	.64	.81	.69	.82
		R_b	.65	.81	.65	.82	.65	.81	.67	.82
		R_c	.67	.82	.66	.82	.63	.81	.67	.82
5	.8	P_a	.81	.87	.79	.87	.77	.86	.81	.88
		P_b	.77	.87	.74	.87	.73	.86	.77	.88
		P_c	.77	.87	.75	.87	.73	.86	.78	.88
		R_a	.80	.86	.79	.87	.78	.86	.80	.87
		R_b	.75	.86	.74	.87	.73	.86	.75	.87
		R_c	.75	.86	.75	.87	.73	.86	.75	.87
1	0	P_a $(= R_a)$	.42	.62	.43	.61	.44	.63	.50	.67
		P_b $(= P_c = R_b = R_c)$	.44	.63	.43	.61	.43	.63	.50	.67
1	.4	P_a $(= R_a)$	.59	.74	.57	.75	.58	.75	.62	.76
		P_b $(= P_c = R_b = R_c)$	.59	.74	.60	.75	.60	.75	.63	.76
1	.8	P_a $(= R_a)$	.79	.87	.79	.87	.78	.86	.79	.87
		P_b $(= P_c = R_b = R_c)$	.79	.87	.79	.87	.78	.86	.79	.87

Table A-VI

Power of Significance Level .05 Size-corrected Tests Against **Parameter Shift Alternatives** for Models with Normal Innovation Distributions and (i) $\rho_{Dx,u} = .4$, (ii) $\rho_{Dx,u} = .8$, (iii) No Stationary Regressors, and (iv) No Time Trend

			$\rho_{Dx,u} = .4$		$\rho_{Dx,u} = .8$		No Stat. Regr.		No Time Trend	
			T		T		T		T	
m	ρ	Test	100	250	100	250	100	250	100	250
10	0	P_a	.74	.85	.77	.88	.67	.81	.70	.81
		P_b	.66	.84	.69	.87	.51	.77	.58	.79
		P_c	.59	.82	.65	.86	.40	.74	.58	.79
		R_a	.73	.86	.75	.87	.67	.84	.74	.85
		R_b	.70	.85	.64	.87	.55	.82	.63	.83
		R_c	.66	.86	.55	.86	.61	.83	.72	.85
10	.4	P_a	.80	.88	.82	.90	.73	.85	.76	.85
		P_b	.71	.88	.76	.90	.59	.83	.68	.84
		P_c	.69	.87	.74	.90	.54	.81	.68	.84
		R_a	.77	.87	.78	.89	.71	.84	.74	.85
		R_b	.69	.87	.71	.88	.55	.83	.62	.83
		R_c	.71	.87	.73	.88	.61	.83	.72	.85
10	.8	P_a	.86	.91	.88	.92	.82	.88	.84	.89
		P_b	.76	.91	.79	.92	.69	.87	.78	.89
		P_c	.79	.91	.82	.93	.64	.88	.78	.89
		R_a	.84	.89	.86	.90	.80	.87	.82	.88
		R_b	.72	.89	.74	.90	.57	.85	.67	.86
		R_c	.74	.89	.77	.90	.54	.86	.75	.88

Table A-VI (cont.)

			$\rho_{Dx,u}$ = .4		$\rho_{Dx,u}$ = .8		No Stat. Regr.		No Time Trend	
			T		*T*		*T*		*T*	
m	ρ	Test	100	250	100	250	100	250	100	250
5	0	P_a	.67	.81	.71	.83	.60	.76	.62	.77
		P_b	.63	.80	.68	.83	.52	.74	.57	.75
		P_c	.61	.79	.67	.82	.46	.72	.57	.76
		R_a	.68	.82	.71	.84	.63	.79	.66	.81
		R_b	.67	.82	.69	.84	.58	.78	.62	.79
		R_c	.68	.82	.70	.84	.61	.79	.67	.80
5	.4	P_a	.72	.85	.75	.87	.67	.81	.69	.82
		P_b	.71	.83	.75	.87	.63	.81	.68	.82
		P_c	.70	.84	.74	.87	.60	.80	.68	.82
		R_a	.72	.86	.75	.87	.65	.81	.71	.82
		R_b	.70	.82	.73	.86	.62	.80	.65	.82
		R_c	.71	.84	.73	.86	.63	.81	.70	.83
5	.8	P_a	.84	.90	.86	.91	.81	.87	.80	.87
		P_b	.81	.90	.83	.91	.74	.87	.79	.88
		P_c	.81	.90	.82	.91	.75	.87	.78	.88
		R_a	.83	.89	.84	.90	.80	.87	.80	.87
		R_b	.78	.89	.81	.90	.72	.87	.74	.86
		R_c	.79	.89	.81	.90	.73	.86	.78	.87
1	0	P_a $(= R_a)$	.50	.68	.55	.71	.43	.62	.43	.63
		P_b $(= P_c = R_b = R_c)$	.51	.67	.56	.71	.43	.62	.43	.63
1	.4	P_a $(= R_a)$	.65	.78	.69	.80	.60	.74	.59	.74
		P_b $(= P_c = R_b = R_c)$	.64	.78	.69	.80	.59	.74	.60	.74
1	.8	P_a $(= R_a)$	.82	.88	.83	.90	.79	.87	.79	.87
		P_b $(= P_c = R_b = R_c)$	.82	.88	.84	.90	.79	.86	.79	.87

Figure 1.1 Thailand

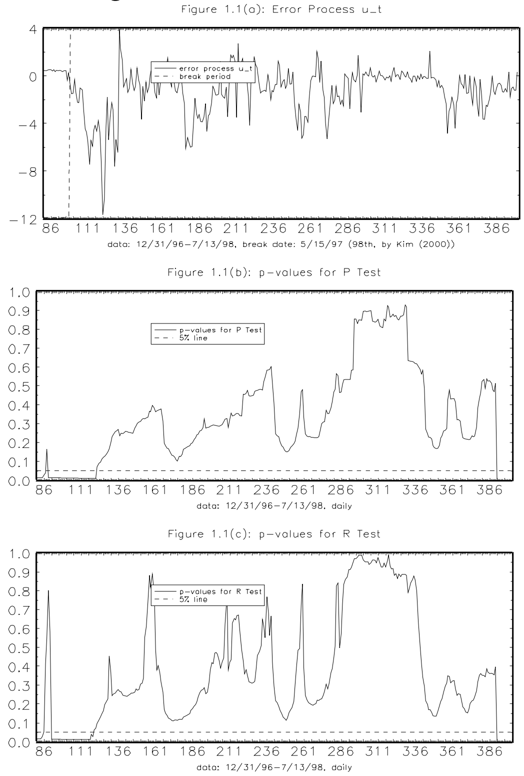

Figure 1.2 Philippines

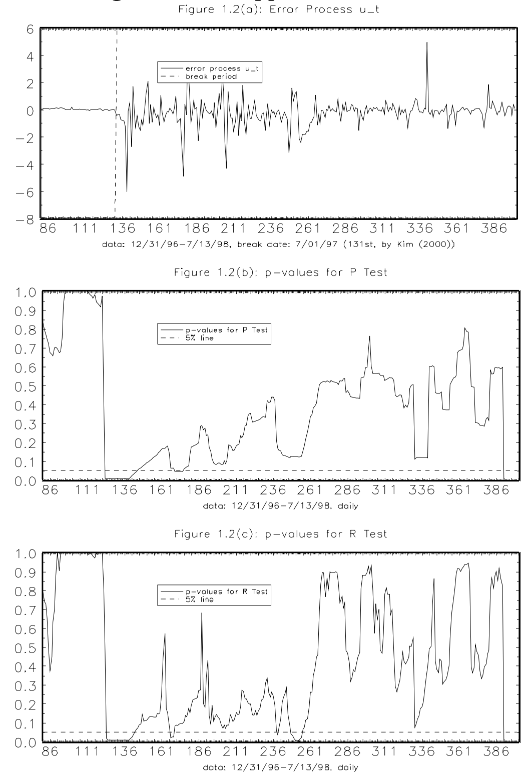

Figure 1.3 Indonesia

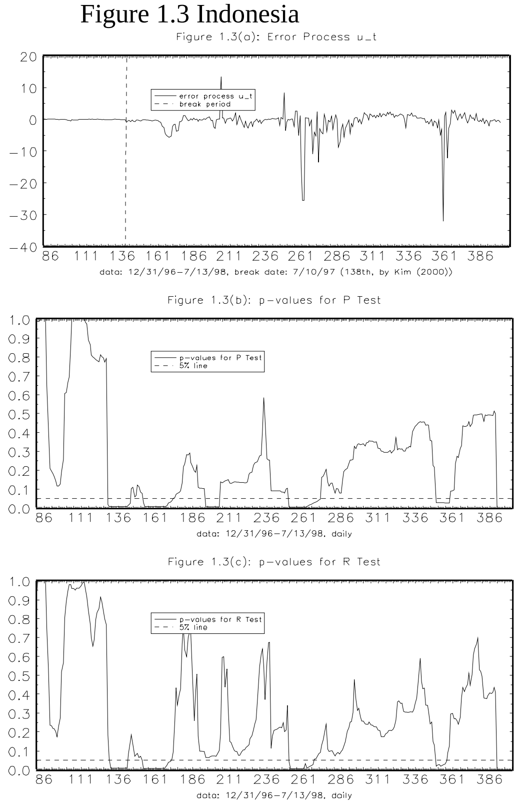

Figure 1.4 Singapore