Tests for Cointegration Breakdown Over a Short Time Period

Donald W. K. ANDREWS
Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520-8281
(*donald.andrews@yale.edu*)

Jae-Young KIM
School of Economics, Seoul National University, Seoul, Korea (*jykim017@snu.ac.kr*)

This article 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 postbreakdown sum of squared residuals and sum of squared reverse partial sums of residuals. Critical values are provided using a subsampling method. Asymptotic results take the number of observations in the breakdown period, m, to be fixed while the total sample size, $T + m$, goes to infinity.

KEY WORDS: Cointegration; Least squares estimator; Model breakdown; Parameter change test; Structural change.

1. INTRODUCTION

This article addresses the problem of cointegration breakdown over a short time period. 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 a change in the stock market 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 postbreakdown 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), Hansen and Johansen (1999), and Kim (1999). For the case of cointegration with known cointegrating vector, the tests of Kim (2000) and Kim, Belaire-Franch, and Amador (2002) also apply. These tests are not appropriate for the case considered here in which the postbreakdown period is relatively short.

In this article we introduce tests for cointegration breakdown that are asymptotically valid when the length, m, of the postbreakdown 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 introduced by Andrews (2003) and described later. The critical values are easy to compute.

For simplicity, in the remainder of this section and in the bulk of the article, 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, is the sum of squared postbreak residuals evaluated at a full-sample estimator, such as the least squares (LS) estimator,

$$P = \sum_{t=T+1}^{T+m} \widehat{u}_t^2, \tag{1}$$

where $\widehat{u}_t$ is a residual. 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.

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, R, is given by the sum of squared reverse partial sums of the postbreak residuals,

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

where $\widehat{u}_s$ is a residual. The form of this statistic is similar to tests considered by Gardner (1969), MacNeill (1978), Nyblom and Makelainen (1983), King and Hillier (1985), Nyblom (1986, 1989), Nabeya and Tanaka (1988), Leybourne and McCabe (1989), Perron (1991), Kwiatkowski, Phillips, Schmidt, and Shin (1992), Tanaka (1993), and Shin (1994).

Critical values for the two test statistics considered are obtained by a subsampling method. 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 jth observation, rather than for breakdown starting at the $(T + 1)$th observation, for $j = 1, \ldots, T - m + 1$. 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, just requiring calculation of $T - m + 1$ versions of the original statistic. p values are also easily obtained using this method.

The subsampling critical values use subsamples of length m, the number of postbreakdown observations. There is no arbitrary smoothing parameter or block length parameter to select; no heteroscedasticity and autocorrelation consistent covariance

Journal of Business & Economic Statistics
October 2006, Vol. 24, No. 4
DOI 10.1198/073500106000000297

matrix estimator is required. (For the use of subsampling in other contexts, see Politis, Romano, and Wolf 1999.)

Andrews (2003) used subsampling 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 article considers linear models only, but allows for nonstationary regressors and hence cointegrating regression models. The test statistics considered in the two articles 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 prebreakdown error variance. Power also depends on m. The larger the m, the greater the power, everything else being equal. Power may be low if m is small or if the magnitude of the breakdown is not large. Consequently, failure to reject the null hypothesis should not be interpreted as strong evidence in favor of stable cointegration.

The article presents some Monte Carlo simulations designed to assess the finite-sample size and size-corrected power properties of the P and R tests. 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-squared with 2 df, t_3, and uniform. The unit root regressor differences are correlated with the errors in some of the cases considered. The prebreakdown sample sizes are $T = 100$ and 250, and the postbreakdown 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.

The simulation results demonstrate that the null rejection probabilities of the P and R tests are quite good, especially considering that the range of cases considered is wide. For example, for the nominal 5% P test and $T = 250$, 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 test is somewhat better than the R test for unit root alternatives and vice versa for parameter shift alternatives. The differences between the powers of the P and R tests typically are not large. The best test in terms of power is the R test, because it has less variable power across different distributions than the P test.

Combining the simulation results for size and power, we have a slight preference for the P test, because its size properties are somewhat better than those of the R test. The P test has pretty good size and power properties across the wide range of models and parameter combinations considered in the simulations.

Use of the P and R tests is illustrated by the results of Carstensen (2006) (which follows this article in this issue of JBES) on testing for breakdown of a cointegrating relation in a conventional money demand equation for the European Monetary Union for the end-of-sample periods 1999Q1–2003Q2 and 2001Q4–2003Q2 based on quarterly data starting in 1980. The baseline model has dependent variables given by real M3 and regressors given by real GDP and the difference between a short-term interest rate and the rate on M3. Carstensen found that the P and R tests strongly reject the null hypothesis of stability over these periods. On the other hand, the full-sample fluctuation and Nyblom tests of Hansen and Johansen (1999) for the period 1980Q1–2003Q2 fail to reject the null of stability at conventional significance levels. Hence in this empirical application, the P and R tests provide useful tools for detecting model misspecification.

The remainder of the article is organized as follows. All sections except Section 5 discuss end-of-sample cointegration breakdown tests. Section 2 introduces the model and hypotheses of interest. Section 3 presents the tests considered. Section 4 provides the asymptotic justification for the subsample critical values for a standard cointegration model estimated by least squares. Results that apply to more general cointegration models and estimators other than least squares are given in Appendix A, and proofs of the asymptotic results are given in Appendix B. Section 5 discusses tests for cointegration breakdown occurring at the beginning or in the middle of the sample. Section 6 provides some Monte Carlo results.

2. MODEL AND HYPOTHESES

The model is

$$y_t = \begin{cases} x_t'\beta_0 + u_t & \text{for } t = 1, \ldots, T \\ x_t'\beta_t + u_t & \text{for } t = T+1, \ldots, T+m, \end{cases} \tag{3}$$

where $y_t, u_t \in \mathbb{R}$ and $x_t, \beta_0, \beta_t \in \mathbb{R}^k$. Under the maintained hypothesis, the errors for the first T time periods, $\{u_t : t = 1, \ldots, T\}$, are mean 0, stationary, and ergodic. In addition, under the maintained hypothesis, the regressors for all time periods, $\{x_t : t = 1, \ldots, T+m\}$, 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. The regressors are not assumed to be strictly exogenous. The precise assumptions are given in Section 4 and Appendix A.

The null and alternative hypotheses are

$$\begin{aligned} H_0 &: \begin{cases} \beta_t = \beta_0 \text{ for all } t = T+1, \ldots, T+m \text{ and} \\ \{u_t : t = 1, \ldots, T+m\} \text{ are stationary and ergodic,} \end{cases} \\ H_1 &: \begin{cases} \beta_t \neq \beta_0 \text{ for some } t = T+1, \ldots, T+m \text{ and/or} \\ \text{the distribution of } \{u_{T+1}, \ldots, u_{T+m}\} \text{ differs from} \\ \text{the distribution of } \{u_1, \ldots, u_m\}. \end{cases} \end{aligned} \tag{4}$$

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

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

3. COINTEGRATION BREAKDOWN TESTS

3.1 *P* Test

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

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

$$\begin{aligned} \mathbf{Y}_{r:s} &= (y_r, \ldots, y_s)', \\ \mathbf{X}_{r:s} &= (x_r, \ldots, x_s)', \qquad \text{and} \\ \mathbf{U}_{r:s} &= (u_r, \ldots, u_s)'. \end{aligned} \tag{5}$$

For $j = 1, \ldots, T+1$, let

$$\begin{aligned} P_j(\beta, \Omega) = \big(\mathbf{Y}_{j:(j+m-1)} - \mathbf{X}_{j:(j+m-1)}\beta\big)' \\ \times \Omega\big(\mathbf{Y}_{j:(j+m-1)} - \mathbf{X}_{j:(j+m-1)}\beta\big), \end{aligned} \tag{6}$$

where Ω is some nonsingular $m \times m$ matrix. The P and R tests are both defined using $P_j(\beta, \Omega)$, but with different choices of Ω.

Let $\widehat{\beta}_{1:(T+m)}$ denote an estimator of β_0 based on the observations $t = 1, \ldots, T+m$. For example, for the LS estimator,

$$\widehat{\beta}_{1:(T+m)} = \big(\mathbf{X}'_{1:(T+m)}\mathbf{X}_{1:(T+m)}\big)^{-1}\mathbf{X}'_{1:(T+m)}\mathbf{Y}_{1:(T+m)} \tag{7}$$

(provided that $\mathbf{X}'_{1:(T+m)}\mathbf{X}_{1:(T+m)}$ is nonsingular). Other estimators can also be considered, including the fully modified estimator of Phillips and Hansen (1990); the maximum likelihood (ML) estimator (see Johansen 1988, 1991; Ahn and Reinsel 1990; 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, that we consider is defined by

$$\begin{aligned} P = P_{T+1}\big(\widehat{\beta}_{1:(T+m)}\big) = \sum_{t=T+1}^{T+m} \big(y_t - x_t'\widehat{\beta}_{1:(T+m)}\big)^2, \\ \text{where } P_j(\beta) = P_j(\beta, I_m) \end{aligned} \tag{8}$$

and I_m denotes the m-dimensional identity matrix. As defined, P is the postbreakdown sum of squared residuals. The statistic P is often referred to as a *predictive* statistic. The form of this statistic is motivated by the F statistic for testing for a one-time change in the regression parameter occurring at time $T+1$ when $m \le k$ in a linear regression model with known error variance (see, e.g., 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 (see, e.g., Scheffé 1959, chap. 2). Predictive statistics have been used by Dufour, Ghysels, and Hall (1994) and Andrews (2003) to test for end-of-sample instability in models with stationary observations.

We note that when $m > k$, the F statistic is based on the projection of the postbreakdown residual vector on the postbreakdown regressor matrix. One can define a test statistic, call it S, that corresponds to this. Andrews (2003) did this for linear and nonlinear models with stationary observations. In the present context, however, the subsampling critical values that we use for the tests do not deliver an asymptotically valid critical value, because the regressors are not stationary and the statistic S depends on the regressors. Hence we do not consider the S statistic any further in this article.

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

The random variables $\{P_j(\beta_0) : j = 1, \ldots, T-m+1\}$ are stationary and ergodic under H_0 and H_1. Consequently, the empirical distribution function (df) of $\{P_j(\beta_0) : j = 1, \ldots, T-m+1\}$ is a consistent estimator of the df of $P_1(\beta_0)$. Hence we can consistently estimate the df of $P_1(\beta_0)$ using the empirical df of $\{P_j(\beta) : j = 1, \ldots, T-m+1\}$ evaluated at a consistent estimator of β_0 (see Thm. 2).

Simulations show that when the subsample statistics are evaluated at $\widehat{\beta}_{1:(T+m)}$ (or the analogous estimator that uses only the observations indexed by $t = 1, \ldots, T$), the test tends to overreject the null hypothesis somewhat. A simple finite-sample adjustment to the subsample statistics to make them more variable is to evaluate them at the leave-m-out estimators defined as follows: For $j = 1, \ldots, T-m+1$, let

$$\begin{aligned} \widehat{\beta}_{(j)} = \text{ estimator of } \beta \text{ using observations indexed by} \\ t = 1, \ldots, T \text{ with } t \ne j, \ldots, j+m-1. \end{aligned} \tag{9}$$

But simulations also show when the subsample statistics are evaluated at $\widehat{\beta}_{(j)}$, the test tends to underreject the null hypothesis somewhat in a broad array of cases (see Andrews and Kim 2003 for details).

Hence we introduce subsample statistics for use with P that are less variable than $\{P_j(\widehat{\beta}_{(j)}) : j = 1, \ldots, T-m+1\}$ but more variable than $\{P_j(\widehat{\beta}_{1:(T+m)}) : j = 1, \ldots, T-m+1\}$. We define the "leave-$m/2$-out" estimator, $\widehat{\beta}_{2(j)}$, as

$$\begin{aligned} \widehat{\beta}_{2(j)} = \text{ estimator of } \beta \text{ using observations indexed by} \\ t = 1, \ldots, T \text{ with } t \ne j, \ldots, j + \lceil m/2 \rceil - 1 \end{aligned} \tag{10}$$

for $j = 1, \ldots, T-m+1$, where $\lceil m/2 \rceil$ denotes the smallest integer that is greater than or equal to $m/2$. For the types of estimators mentioned earlier, the estimator $\widehat{\beta}_{2(j)}$ is consistent for β_0 (uniformly over j) under suitable assumptions, see below.

We define the P_j subsample statistics as

$$P_j = P_j\big(\widehat{\beta}_{2(j)}\big) \quad \text{for } j = 1, \ldots, T-m+1. \tag{11}$$

The empirical df of $\{P_j : j = 1, \ldots, T-m+1\}$ is

$$\widehat{F}_{P,T}(x) = \frac{1}{T-m+1} \sum_{j=1}^{T-m+1} \mathbb{1}(P_j \le x). \tag{12}$$

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

$$\widehat{q}_{P,1-\alpha} = \inf\{x \in \mathbb{R} : \widehat{F}_{P,T}(x) \ge 1-\alpha\}. \tag{13}$$

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

$$(T-m+1)^{-1} \sum_{j=1}^{T-m+1} \mathbb{1}(P > P_j) \geq 1-\alpha. \tag{14}$$

The p value for the P test is

$$pv_P = (T-m+1)^{-1} \sum_{j=1}^{T-m+1} \mathbb{1}(P \leq P_j). \tag{15}$$

3.2 The P Test With Estimated Weight Matrix

The P test is designed for the case in which 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 this, but found that they were somewhat inferior to the P test in terms of closeness of nominal and true sizes and in terms of size-corrected power across the range of models considered in Section 6.

3.3 Locally Best Invariant Test for Unit Root Alternatives

The P test is motivated by the F test for a one-time change in the parameter vector β. We now consider the 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 are similar to those considered in the articles represented in Section 1.

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

$$\begin{aligned} &y_t = x_t'\beta_0 + u_t \quad \text{for } t = 1, \ldots, T+m, \\ &u_t = \psi_t + \lambda^{1/2}\widetilde{\psi}_t, \\ &\psi_t \sim \text{iid}\,\mathrm{N}(0,1) \quad \text{for } t = 1, \ldots, T+m, \\ &\widetilde{\psi}_t = \begin{cases} 0 & \text{for } t = 1, \ldots, T \\ \widetilde{\psi}_{t-1} + \varepsilon_t & \text{for } t = T+1, \ldots, T+m, \end{cases} \\ &\varepsilon_t \sim \text{iid}\,\mathrm{N}(0,1) \quad \text{for } t = T+1, \ldots, T+m, \end{aligned} \tag{16}$$

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

$$H_0 : \lambda = 0 \qquad \text{and} \qquad H_1 : \lambda > 0. \tag{17}$$

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 $t = 1, \ldots, T$ and lack of cointegration (i.e., spurious regression) for the observations $t = T+1, \ldots, T+m$.

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

$$\begin{aligned} &\mathbf{Y}_{1:(T+m)} \sim \mathrm{N}\big(\mathbf{X}_{1:(T+m)}\beta_0,\ I_{T+m} + \lambda V\big), \\ &\quad \text{where} \quad V = \text{diag}\{\mathbf{0}_T, A_m\}, \\ &\qquad\qquad [A_m]_{k,\ell} = \min\{k,\ell\} \text{ for } k,\ell = 1, \ldots, m, \end{aligned} \tag{18}$$

and $\mathbf{0}_T$ is a $T \times T$ matrix of 0's. That is, V is a $(T+m) \times (T+m)$ matrix consisting of 0's 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:

$$\begin{aligned} \mathbf{Y}_{1:(T+m)} &\to \mathbf{Y}_{1:(T+m)} + \mathbf{X}_{1:(T+m)}\gamma, \\ \beta_0 &\to \beta_0 + \gamma. \end{aligned} \tag{19}$$

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 $X = \mathbf{X}_{1:(T+m)}$. We have

$$S = J'\mathbf{Y}_{1:(T+m)} \sim \mathrm{N}(0, I_{T+m-k} + \lambda J'VJ). \tag{20}$$

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

$$\frac{d}{d\lambda} \log f_{T+m}(S|\lambda)\bigg|_{\lambda=0} > K, \tag{21}$$

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

$$\begin{aligned} 2\frac{d}{d\lambda} \log f_{T+m}(S|\lambda) &= -S' \frac{d}{d\lambda}(I_{T+m-k} + \lambda J'VJ)^{-1} S \\ &= S'(I_{T+m-k} + \lambda J'VJ)^{-1} \\ &\quad \times J'VJ(I_{T+m-k} + \lambda J'VJ)^{-1} S, \\ 2\frac{d}{d\lambda} \log f_{T+m}(S|\lambda)\bigg|_{\lambda=0} &= S'J'VJS \\ &= \widehat{\mathbf{U}}'_{1:(T+m)} V \widehat{\mathbf{U}}_{1:(T+m)} \\ &= \widehat{\mathbf{U}}'_{(T+1):(T+m)} A_m \widehat{\mathbf{U}}_{(T+1):(T+m)}, \end{aligned} \tag{22}$$

where

$$\begin{aligned} \widehat{\mathbf{U}}_{1:(T+m)} &= JS = JJ'\mathbf{Y}_{1:(T+m)} \\ &= \mathbf{Y}_{1:(T+m)} - \mathbf{X}_{1:(T+m)}\widehat{\beta}_{LS,1:(T+m)} \end{aligned} \tag{23}$$

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

3.4 R Tests

The LBI test statistic of (22) can be written using (6) as

$$P_{T+1}\big(\widehat{\beta}_{LS,1:(T+m)}, A_m\big). \tag{24}$$

That is, the LBI test statistic is just like the P statistic except that it uses the weight matrix A_m instead of the identity matrix. In this section we define a test R that is analogous to the P test defined earlier, but uses the weight matrix A_m instead of I_m.

Define

$$R = P_{T+1}\big(\widehat{\beta}_{1:(T+m)}, A_m\big) \qquad \text{and}$$
$$R_j = P_{T+1}\big(\widehat{\beta}_{2(j)}, A_m\big). \tag{25}$$

The estimator $\widehat{\beta}_{2(j)}$ used in the subsample statistic R_j is chosen for the same reasons as cited above for the P tests. Critical values and p values for the R test are obtained as in (13)–(15), with (P, P_j) replaced by (R, R_j). The estimator $\widehat{\beta}_{1:(T+m)}$ used with the R test can be the LS estimator or some other estimator.

It turns out that the R test statistic is a sum of squares of reverse partial sums of residuals. To see this, let Q be the $m \times m$ matrix that has 1's on and above the main diagonal and 0's below the main diagonal. Then $A_m = Q'Q$, and R can be written as

$$R = P_{T+1}\big(\widehat{\beta}_{1:(T+m)}, A_m\big) = \big(Q\widehat{\mathbf{U}}\big(\widehat{\beta}_{1:(T+m)}\big)\big)'Q\widehat{\mathbf{U}}\big(\widehat{\beta}_{1:(T+m)}\big)$$
$$= \sum_{t=T+1}^{T+m}\left(\sum_{s=t}^{T+m}\big(y_s - x_s'\widehat{\beta}_{1:(T+m)}\big)\right)^2. \tag{26}$$

The statistic R_j can be written in the same way with $\widehat{\beta}_{1:(T+m)}$ replaced by $\widehat{\beta}_{2(j)}$.

As shown next, the R test is asymptotically valid in a much broader class of models than the model of (16) that generates the LBI test.

4. ASYMPTOTIC RESULTS

4.1 Assumptions

For simplicity, in this section we consider the leading case in which the regressor vector x_t contains a constant, a linear time trend, a p_1-vector $x_{1,t}$ of integrated variables, and a p_2-vector $x_{2,t}$ of stationary variables for $p_j \geq 0$ for $j = 1, 2$,

$$y_t = \beta_{1,0} + t\beta_{2,0} + x_{1,t}'\beta_{3,0} + x_{2,t}'\beta_{4,0} + u_t. \tag{27}$$

(The results also cover the case in which any of these regressors does not appear in the model; for example, there need not be a time trend or stationary regressors.) We suppose that the parameters are estimated by LS.

In the Appendix we consider the more general case in which the elements of x_t are arbitrary linear combinations of deterministic trends, integrated variables, and stationary variables. In the Appendix we also allow for estimation methods other than LS. The asymptotic results given in this section hold for the more general model and other (consistent) estimation methods.

The following assumption is assumed to hold under the null hypothesis.

Assumption S. (a) $x_{1,t} = x_{1,t-1} + v_t$ for $t = 1, 2, \ldots$, where $x_{1,0} = O_p(1)$.

(b) $\{(u_t, x_{2,t}, v_t) : t \geq 1\}$ is a stationary strong mixing sequence of mean-0 random variables with strong mixing numbers that satisfy $\sum_{r=1}^{\infty} \alpha^{1-2/\beta}(r) < \infty$.

(c) $Ex_{2,t}u_t = 0$.

(d) $\sup_{t\geq 1} E\|(u_t, x_{2,t}', v_t')\|^{\beta+\varepsilon} < \infty$ for some $\beta > 2$ and $\varepsilon > 0$.

(e) $\Sigma_{2,0} = Ex_{2,t}x_{2,t}'$ and $\Omega^* = Ev_1v_1' + \sum_{k=2}^{\infty} Ev_1v_k' + \sum_{k=2}^{\infty} Ev_kv_1'$ are positive definite.

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

Assumptions S(a) and S(b) imply that the regressors $\{x_{1,t} : t \geq 1\}$ are integrated of order 1 and $\{(u_t, x_{2,t}, v_t) : t \geq 1\}$ is a stationary asymptotically weakly dependent time series. Assumption S(c) specifies that the stationary regressors, $x_{2,t}$, are not endogenous. Assumption S(d) is a relatively weak moment condition. Assumption S(e) guarantees that none of the stationary or nonstationary regressors is redundant. Assumption S(f) holds if the errors have an absolutely continuous component, which is not very restrictive.

Assumption S(b) assumes that $\{(x_{2,t}, v_t) : t \geq 1\}$ is a stationary sequence. This is not essential for the tests considered here to have the desired asymptotic null rejection rate. One could have structural breaks in the $\{(x_{2,t}, v_t) : t \geq 1\}$ sequence (or other types of nonstationarity) under the null hypothesis. (For brevity, we do not prove this here.) Hansen (2000) argued that it is desirable to allow for structural breaks in the regressor sequence under the null hypothesis.

4.2 Asymptotic Results

Next we state the asymptotic results that justify the use of the subsample critical values introduced earlier. Let $\widehat{F}_{P,T}(x)$ denote the empirical df based on $\{P_j : j = 1, \ldots, T-m+1\}$, that is,

$$\widehat{F}_{P,T}(x) = \frac{1}{T-m+1}\sum_{j=1}^{T-m+1} \mathbb{1}(P_j \leq x). \tag{28}$$

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 $P_1(\beta_0)$, and let $\widehat{q}_{P,1-\alpha}$ denote the $(1-\alpha)$th sample quantile of $\{P_j : j = 1, \ldots, T-m+1\}$, as defined in (13).

Let P_∞ be a random variable with the same distribution as $P_{T+1}(\beta_0)$. Under Assumption S, 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 H_0 by stationarity. Define $\widehat{F}_{R,T}(x)$, $F_R(x)$, $q_{R,1-\alpha}$, $\widehat{q}_{R,1-\alpha}$, and R_∞ analogously with R in place of P.

The asymptotic null behavior of the P and R tests is given in the following theorem.

Theorem 1. Suppose that Assumption S holds. Then, under H_0, as $T \to \infty$, we have (a) $P \to_d P_\infty$, (b) $\widehat{F}_{P,T}(x) \to_p F_P(x)$ for all x in a neighborhood of $q_{P,1-\alpha}$, (c) $\widehat{q}_{P,1-\alpha} \to_p q_{P,1-\alpha}$, and (d) $\Pr(P > \widehat{q}_{P,1-\alpha}) \to \alpha$, and (e) parts (a)–(d) hold with R in place of P.

Comment. Parts (d) and (e) of Theorem 1 show that the asymptotic null rejection rate of the tests based on P and R is α, as desired.

Theorem 2, given in the Appendix, establishes the asymptotic distributions of the P and R test statistics under the alternative hypothesis H_1. Theorem 2 shows that the random critical values $\widehat{q}_{P,1-\alpha}$ and $\widehat{q}_{R,1-\alpha}$ have the same asymptotic behavior under H_1 as under H_0. This is desirable for the power of the tests. Theorem 2 also shows that P and R do not diverge to infinity as $T \to \infty$ under H_1. Hence P and R do not yield consistent tests. This is due to the assumption that the number, m, of postbreakdown 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 is an asymptotically unbiased test, and likewise for R.

5. BREAKDOWN AT THE BEGINNING OR IN THE MIDDLE OF THE SAMPLE

The tests introduced previously 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 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 earlier 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 $t = t_0, \ldots, t_0 + m - 1$ when the total number of observations is $T + m$. The null and alternative hypotheses are given by

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

$$H_1: \begin{cases} y_t = x_t'\beta_0 + u_t \text{ for all } t = 1, \ldots, t_0 - 1,\ t_0 + \\ m, \ldots, T+m \text{ and } y_t = x_t'\beta_t + u_t \text{ with } \beta_t \neq \beta_0 \\ \text{for some } t = t_0, \ldots, t_0 + m - 1 \text{ and/or the dis-} \\ \text{tribution of } \{u_{t_0}, \ldots, u_{t_0+m-1}\} \text{ differs from that} \\ \text{of error sequences } \{u_t, \ldots, u_{t+m-s}\} \text{ that do not} \\ \text{overlap with it.} \end{cases} \tag{29}$$

One can construct tests for these hypotheses by moving the observations $\{(y_t, x_t) : t = t_0, \ldots, t_0 + m - 1\}$ to the end of the sample and moving the observations indexed $t = T + 1, \ldots, T + m$ up to fill the gap. The observations originally indexed by $t = t_0, \ldots, t_0 + m - 1$ are subsequently indexed by $t = T, \ldots, T + m$, and the tests defined previously can be used to test the hypotheses in (29).

6. MONTE CARLO EXPERIMENT

In this section we describe some Monte Carlo results that are designed to assess and compare the null rejection rates and power properties of the P and R tests.

6.1 Experimental Design

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

$$y_t = x_t'\beta_0 + u_t \quad \text{for } t = 1, \ldots, T + m, \tag{30}$$

with $\beta_0 = 0$. We consider two values of T: 100 and 250. For the main results, we consider three values of m: 10, 5, and 1. To show what happens when m is relatively large compared with T, we also report some results with $m = 25$ and 50. In the base model that we consider, we take

$$y_t = \beta_{1,0} + t\beta_{2,0} + x_{1,t}'\beta_{3,0} + x_{2,t}'\beta_{4,0} + u_t, \tag{31}$$

where $x_{1,t}$ is a vector of unit root regressors and $x_{2,t}$ is a vector of stationary mean-0 regressors. (The dimensions of $x_{1,t}$ and $x_{2,t}$ are two for all cases considered except one; see later.) 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: (a) standard normal [N(0, 1)], (b) chi-squared with 2 df (χ_2^2) recentered and rescaled to have mean 0 and variance 1, (c) t with 3 df (t_3) rescaled to have variance 1, and (d) uniform (U) on $[-\sqrt{12}/2, \sqrt{12}/2]$, which has mean 0 and variance 1. 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, denoted as $\rho_{Dx,u}$. 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 used are

$$\{(\psi_t^*, \xi_t^{*\prime}, \eta_t^*, x_{2,t}^{*\prime})' : t = 1, \ldots, T + m\}, \tag{32}$$

where $\psi_t^*, \eta_t^* \in \mathbb{R}$, $\xi_t^* \in \mathbb{R}^{d_{x_1}}$, $x_{2,t}^* \in \mathbb{R}^{d_{x_2}}$, and d_{x_1} and d_{x_2} denote the dimensions of $x_{1,t}$ and $x_{2,t}$. 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 before. The AR(1) processes based on these innovations are

$$\begin{aligned} \psi_t &= \rho\psi_{t-1} + \psi_t^*, \\ \xi_t &= \rho\xi_{t-1} + \xi_{,t}^*, \\ \eta_t &= \rho\eta_{t-1} + \eta_{,t}^*, \\ x_{2,t} &= \rho x_{2,t-1} + x_{2,t}^*, \end{aligned} \tag{33}$$

for $t = 1, \ldots, T + m$. 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

$$\begin{aligned} u_t &= (1 - \rho_{Dx,u})^{1/2}\psi_t + \rho_{Dx,u}^{1/2}\eta_t, \\ Dx_{1,t} &= (1 - \rho_{Dx,u})^{1/2}\xi_t + \rho_{Dx,u}^{1/2}\eta_t 1_{d_{x_1}}, \end{aligned} \tag{34}$$

where $1_{d_{x_1}}$ denotes a d_{x_1}-vector of 1's. 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,

$$\begin{aligned} &\text{Base model:} \\ &\text{BC(a) } x_t = (1, t, x_{1,t}', x_{2,t}')' \text{ and } x_{1,t}, x_{2,t} \in \mathbb{R}^2 \\ &\text{BC(b) } \rho_{Dx,u} = 0 \\ &\text{BC(c) } G = \text{N}(0, 1). \end{aligned} \tag{35}$$

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 $\rho_{Dx,u} = .8$. 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 with no time trend. Models 2–7 are summarized as follows:

$$\begin{aligned}
&\text{Model 2} \quad (\chi^2_2 \text{ distribution}):\\
&\qquad \text{BC(a) and BC(b) hold and } G = \chi^2_2,\\
&\text{Model 3} \quad (t_3 \text{ distribution}):\\
&\qquad \text{BC(a) and BC(b) hold and } G = t_3,\\
&\text{Model 4} \quad (U \text{ distribution}):\\
&\qquad \text{BC(a) and BC(b) hold and } G = U,\\
&\text{Model 5} \quad (\rho_{Dx,u} = .4):\\
&\qquad \text{BC(a) and BC(c) hold and } \rho_{Dx,u} = .4,\\
&\text{Model 6} \quad (\rho_{Dx,u} = .8):\\
&\qquad \text{BC(a) and BC(c) hold and } \rho_{Dx,u} = .8,\\
&\text{Model 7} \quad (\text{no stationary regressors}):\\
&\qquad \text{BC(b) and BC(c) hold, } x_t = (1, t, x'_{1,t})' \text{ and } x_{1,t} \in \mathbb{R}^4,\\
&\text{Model 8} \quad (\text{no time trend}):\\
&\qquad \text{BC(b) and BC(c) hold, } x_t = (1, x'_{1,t}, x'_{2,t})', \text{ and}\\
&\qquad x_{1,t}, x_{2,t} \in \mathbb{R}^2.
\end{aligned} \tag{36}$$

For each of the eight models, the main results are based on three values of ρ, two values of T, and three values of m (10, 5, and 1).

For each of the eight models, we report the actual rejection rates of the nominal 5% P and R tests. 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 $t = T + 1, \ldots, T + m$, the error is given by

$$u_t = (1 - \rho_{Dx,u})^{1/2}\psi_t + \rho_{Dx,u}\eta_t + \sqrt{2}\sum_{s=1}^{t-T}\left[(1 - \rho_{Dx,u})^{1/2}\widetilde{\psi}_s + \rho_{Dx,u}\widetilde{\eta}_s\right], \tag{37}$$

where $\{(\widetilde{\psi}_s, \widetilde{\eta}_s) : s = 1, \ldots, m\}$ has the same distribution as $\{(\psi_s, \eta_s) : s = 1, \ldots, m\}$ 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 $t = 1, \ldots, T + m$, 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 $t = T + 1, \ldots, T + m$, the true parameter β_0 is proportional to a vector of 1's 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 other 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 used 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 nonrandom critical value.

All of 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 Null Rejection Rates. Table 1 presents the null rejection rate results for nominal .05 tests with $m = 10$, 5, and 1. The first two rows of Table 1 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 two 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.

In the discussion paper version of this article (Andrews and Kim 2003), tables A-I and A-II give the rejection rates for each of the 72 model/(m, ρ) combinations. Some of the results stated herein are based on these more detailed tables. When $m = 1$, $P = R$ and separate results are not given for P and R.

The main results are as follows:

1. The P test has the best performance under the null overall. Its average rejection rate is close to .05 when $T = 100$ or 250. The deviation of the rejection rate from .05 is at most .014 when $T = 250$ over the wide range of models considered. The R test underrejects 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 $T = 100$.
2. Not surprisingly, both tests perform noticeably better in terms of null rejection rate when $T = 250$ than when $T = 100$. In particular, the range of rejection rates for each test shrinks considerably when T is increased.
3. The null rejection rates of the P and R tests are not very sensitive to the values of ρ and m.
4. For the P test, 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 test, there is no clear pattern of variation of the rejection rates with the type of distribution.
5. For the P and R 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 and R 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.

Table 1. Average and Range of Null Rejection Rates Over Eight Models for Nominal .05 Tests: m = 10, 5, and 1

			$T = 100$		$T = 250$	
m	ρ	Test	Average	Range	Average	Range
Average over all nine		P	.052	[.028, .081]	.051	[.040, .064]
(m, ρ) values		R	.046	[.020, .077]	.047	[.032, .065]
10	0	P	.064	[.046, .081]	.057	[.049, .064]
		R	.034	[.020, .045]	.041	[.032, .047]
10	.4	P	.056	[.041, .075]	.051	[.042, .063]
		R	.038	[.026, .051]	.042	[.033, .050]
10	.8	P	.054	[.044, .066]	.048	[.040, .057]
		R	.053	[.041, .067]	.050	[.044, .059]
5	0	P	.045	[.034, .051]	.049	[.043, .054]
		R	.032	[.024, .038]	.042	[.036, .047]
5	.4	P	.037	[.028, .046]	.046	[.040, .052]
		R	.034	[.026, .042]	.044	[.039, .049]
5	.8	P	.047	[.039, .057]	.052	[.047, .060]
		R	.051	[.043, .061]	.053	[.051, .057]
1	0	$P (= R)$	.044	[.035, .048]	.045	[.042, .049]
1	.4	$P (= R)$	.050	[.044, .053]	.048	[.047, .049]
1	.8	$P (= R)$	.074	[.072, .077]	.062	[.060, .065]

To conclude, we find that the P and R tests have similar null rejection rate performance. The R test tends to underreject the null too often compared with the P test. Hence the P test has the best overall properties under the null. Considering the very wide range of models and (m, ρ) values considered, which range from t_3 distributions to $\rho_{Dx,u} = .8$, the null rejection rate performance of the P test seems quite good. This is especially true for $T = 250$.

6.2.2 Power. Tables 2 and 3 provide 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 1. In the discussion paper version of this article (Andrews and Kim 2003), tables A-III–A-VI give the rejection rates for each of the 72 model/(m, ρ) combinations.

The main findings are as follows:

1. The simulation results indicate that the P test has considerable power against unit root alternatives, even though it is designed for parameter shift alternatives. Likewise, the R test has considerable power against parameter shift alternatives even though it is designed for unit root alternatives. In fact, paradoxically, the P test tends to outperform the R test for unit root alternatives and vice versa with parameter shift alternatives; but the differences are not large.
2. For both tests, power increases sharply with m. This occurs because m determines the amount of information available regarding the postbreak time period.
3. For both tests, power increases by a small amount (roughly .03 on average) as T increases from 100 to 250 for unit root alternatives. Power increases by a substantial amount (roughly .15 on average) as T increases from 100 to 250 for parameter shift alternatives.
4. The 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

Table 2. Unit Root Alternatives: Average and Range of Nonnull Rejection Rates Over Eight Models for Size-Corrected .05 Tests for m = 10, 5, and 1

			$T = 100$		$T = 250$	
m	ρ	Test	Average	Range	Average	Range
Average over all nine		P	.58	[.17, .95]	.61	[.16, .97]
(m, ρ) values		R	.56	[.17, .84]	.59	[.16, .86]
10	0	P	.82	[.57, .95]	.87	[.58, .97]
		R	.80	[.75, .84]	.85	[.80. .86]
10	.4	P	.87	[.66, .94]	.91	[.77, .96]
		R	.78	[.73, .82]	.84	[.80, .85]
10	.8	P	.85	[.73, .89]	.89	[.80, .92]
		R	.75	[.69, .78]	.81	[.76, .83]
5	0	P	.62	[.37, .81]	.69	[.47, .85]
		R	.64	[.56, .69]	.69	[.62, .72]
5	.4	P	.67	[.46, .79]	.73	[.55, .82]
		R	.63	[.56, .66]	.67	[.59, .70]
5	.8	P	.66	[.54, .73]	.68	[.58, .76]
		R	.62	[.54, .67]	.67	[.59, .72]
1	0	$P (= R)$	.23	[.17, .33]	.25	[.16, .36]
1	.4	$P (= R)$	.23	[.18, .30]	.24	[.18, .32]
1	.8	$P (= R)$	.25	[.19, .29]	.24	[.19, .28]

Table 3. Parameter Shift Alternatives: Average and Range of Nonnull Rejection Rates Over Eight Models for Size-Corrected .05 Tests for m = 10, 5, and 1

			$T = 100$		$T = 250$	
m	ρ	Test	Average	Range	Average	Range
Average over all nine		P	.63	[.34, .84]	.80	[.57, .93]
(m, ρ) values		R	.67	[.43, .84]	.82	[.61, .90]
10	0	P	.50	[.34, .65]	.75	[.57, .86]
		R	.65	[.55, .72]	.84	[.83, .86]
10	.4	P	.61	[.47, .74]	.83	[.76, .90]
		R	.68	[.61, .73]	.85	[.83, .88]
10	.8	P	.74	[.64, .82]	.89	[.85, .93]
		R	.69	[.54, .77]	.88	[.86, .90]
5	0	P	.54	[.37, .67]	.75	[.65, .82]
		R	.65	[.61, 70]	.80	[.79, .84]
5	.4	P	.64	[.52, .74]	.81	[.76, .87]
		R	.68	[.63, .73]	.83	[.81, .86]
5	.8	P	.77	[.73, .82]	.88	[.86, .91]
		R	.76	[.73, .81]	.87	[.86, .90]
1	0	$P (= R)$	.46	[.43, .56]	.65	[.61, .71]
1	.4	$P (= R)$	.62	[.59, .69]	.76	[.74, .80]
1	.8	$P (= R)$	.80	[.78, .84]	.87	[.86, .90]

model compared with the base model for parameter shift alternatives.

5. The power for the P test is more sensitive to the distribution compared with the power for the R test. For the P test, the power is lower for the t_3 and χ^2_2 distributions than for the normal and higher for the uniform than for the normal.

Overall, the power of the P and R tests is fairly comparable. The P test has somewhat higher power than the R test against unit root alternatives, but this situation is reversed for parameter shift alternatives. The R test tends to have power that is less variable across changes in the model, such as changes in the distribution, than the P tests. Consequently, the R test is deemed to have the better overall power properties than the P test, but by only a small margin.

Combining the size and power results, we find that the choice of the best test is not clear-cut. The P test has somewhat better size properties because the R test is somewhat under-sized in a number of cases. On the other hand, the R test has power that is less variable across different distributions than the P test. On balance, the P test seems to be preferable because of its size properties. Consequently, we recommend using the P test.

6.2.3 Null Rejection Rates: Large m. Table 4 presents null rejection rate results for nominal .05 tests with $m = 25$ and 50 (with the other parameters as before). The point of Table 4 is to show what happens when m is relatively large compared with T. The rows in the table give the average and range of the null rejection rates over the eight models for each (m, ρ) value and each test.

The main findings are as follows:

1. The P test tends to overreject the null when m is large relative to T, whereas the R test tends to underreject. The extent of overrejection and underrejection increases with m/T. Thus the worst case is when $m = 50$ and $T = 100$, which yields $m/T = .5$; the best case is when $m = 25$ and $T = 250$, which yields $m/T = .1$; and the two intermediate cases are when $m = 25$ and $T = 100$, which yields $m/T = .25$, and $m = 50$ and $T = 250$, which yields $m/T = .2$. In the best case, both the P and R tests perform reasonably well. In the intermediate cases, the P test cannot be recommended because the magnitude of its overrejection is too large, but the R test can be applied. In the worst case, neither test can be recommended, because the overrejection of the P tests is too large and the underrejection of the R test is sufficiently large that the power of the test is likely to be quite low.
2. The null rejection rates of both tests improve as ρ increases from 0 to .4 to .8. Presumably this occurs because increased correlation increases the variability of the test statistics with known true parameter values more than it increases the variability of the actual test statistics that is due to parameter estimation.
3. For the P and R tests, the rejection rates are higher for the χ^2_2 and t_3 distributions than for the normal and uniform distributions. Given that the P test overrejects the null for all distributions, this means that the P test performs better for the normal and uniform distributions than for the χ^2_2 and t_3 distributions. The opposite is true for the R test.
4. The sensitivity of both tests to $\rho_{Dx,u}$ is low.
5. Compared with the base model, the P test performs better in the model with no stationary regressors and worse in the model with no time trend. The opposite is true for the R test.

To conclude, it is evident that the null rejection rate performance of the P and R tests deteriorates as m/T increases. The R test performs well for m/T ratios of $\leq .1$ and adequately for ratios as high as .2 or .25. The P test works well for values as high as .1, but less than adequately for ratios of .2 or .25. Of course, when using these results, one should take into account the fact that they are based on the specific models used in the simulation experiment and may not be applicable generally.

Table 4. Average and Range of Null Rejection Rates Over Eight Models for Nominal .05 Tests: Large m

			$T = 100$		$T = 250$	
m	ρ	Test	Average	Range	Average	Range
25	0	P	.110	[.070, .156]	.073	[.056, .088]
		R	.022	[.006, .048]	.032	[.020, .045]
25	.4	P	.081	[.010, .137]	.064	[.053, .082]
		R	.028	[.009, .057]	.035	[.023, .048]
25	.8	P	.076	[.056, .097]	.052	[.044, .068]
		R	.046	[.022, .082]	.042	[.030, .055]
50	0	P	.175	[.115, .254]	.112	[.087, .143]
		R	.012	[.001, .043]	.029	[.012, .055]
50	.4	P	.139	[.080, .211]	.101	[.081, .138]
		R	.016	[.002, .054]	.033	[.015, .060]
50	.8	P	.106	[.054, .144]	.079	[.062, .107]
		R	.029	[.004, .076]	.044	[.024, .071]

ACKNOWLEDGMENTS

Andrews gratefully acknowledges research support from the National Science Foundation (grants SES-0001706 and SES-0417911). The authors thank the editor and referees for helpful comments.

APPENDIX A: GENERAL ASYMPTOTIC RESULTS

Here we provide general results concerning the asymptotic properties of the P and R tests under H_0 and H_1. Theorem 1 is a special case of the results provided here.

A.1 Assumptions

To simplify the theoretical analysis, 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 used only in the theoretical analysis of the tests. Let

$$z_t = H'x_t = \begin{pmatrix} z_{1,t} \\ z_{2,t} \end{pmatrix},$$

$$\gamma_0 = H^{-1}\beta_0 = \begin{pmatrix} \gamma_{1,0} \\ \gamma_{2,0} \end{pmatrix},$$

$$\gamma_t = H^{-1}\beta_t = \begin{pmatrix} \gamma_{1,t} \\ \gamma_{2,t} \end{pmatrix}, \qquad \text{and} \tag{A.1}$$

$$\widehat{\gamma}_{2(j)} = H^{-1}\widehat{\beta}_{2(j)} = \begin{pmatrix} \widehat{\gamma}_{1,2(j)} \\ \widehat{\gamma}_{2,2(j)} \end{pmatrix},$$

where H is a nonrandom nonsingular $k \times k$ matrix; $z_{\ell,t}$, $\gamma_{\ell,0}$, $\gamma_{\ell,t}$, $\widehat{\gamma}_{\ell,2(j)} \in \mathbb{R}^{k_\ell}$ 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-0 random variables.

The model can be rewritten as

$$y_t = \begin{cases} z_t'\gamma_0 + u_t & \text{for } t = 1, \ldots, T \\ z_t'\gamma_t + u_t & \text{for } t = T+1, \ldots, T+m. \end{cases} \tag{A.2}$$

Let w_t denote the vector of errors and stationary regressors,

$$w_t = \begin{pmatrix} u_t \\ z_{2,t} \end{pmatrix}. \tag{A.3}$$

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

We make the following assumptions.

Assumption 1. $\{w_{0,t} : t \geq 1\}$ are mean 0, stationary, and ergodic random vectors under H_0 and H_1. The distribution of $\{z_{1,t} : t = 1, \ldots, T\}$ is the same under H_0 and H_1. Under H_1, the distribution of $\{w_{T,t} : t = T+1, \ldots, T+m\}$ 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 nonrandom positive-definite diagonal $k_1 \times k_1$ matrices $\{B_T : T \geq 1\}$ under H_0 and H_1.

Assumption 4. $\|B_T(\widehat{\gamma}_{1,1:(T+m)} - \gamma_{1,0})\| \to_p 0$, $\|\widehat{\gamma}_{2,1:(T+m)} - \gamma_{2,0}\| \to_p 0$, $\max_{j=1,\ldots,T-m+1} \|B_T(\widehat{\gamma}_{1,2(j)} - \gamma_{1,0})\| \to_p 0$, and $\max_{j=1,\ldots,T-m+1} \|\widehat{\gamma}_{2,2(j)} - \gamma_{2,0}\| \to_p 0$ with m fixed, under H_0 and H_1, where B_T is as in Assumption 3.

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 earlier and the second condition of Assumption 1, the joint distribution of all of the variables for time periods $t = 1, \ldots, T$ 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 that 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, $z_{1,t}$, be properly normalized. The diagonal element of B_T that corresponds to a unit root variable in $z_{1,t}$ with mean 0 and "asymptotically weakly dependent" innovations is $T^{1/2}$. 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 asymptotically equivalent, because m does not depend on T. We choose $T^{1/2}$ for notational simplicity. 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 0 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 A.2. 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 1. 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 that $z_{1,t} = (1, t, r_t')'$, where r_t is a p vector of unit root variables with mean 0 asymptotically weakly dependent innovations. Then we have

$$B_T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{1/2} I_p \end{pmatrix}. \tag{A.4}$$

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 (see, e.g., Sowell 1990, thms. 1 and 2).

Assumption 4 concerns the behavior of the transformed estimators $\widehat{\gamma}_{1:(T+m)}$ 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-0 asymptotically weakly dependent innovations) just needs 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.

A.2 Least Squares Estimation

In this section we give sufficient conditions for Assumption 4 for the case where the estimator used is the LS estimator defined

in (7). 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 pseudometric on the space of functions on $[0, 1]$. We use the uniform metric, as done by 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) $Ez_{2,t}u_t = 0$, $E|u_t|^{1+\delta} < \infty$, $E\|u_t \times z_{2,t}\|^{1+\delta} < \infty$, and $E\|z_{2,t}\|^{2+\delta} < \infty$ for some $\delta > 0$ for $t \le T$.

(b) $\nu_T(\cdot) \Rightarrow \nu(\cdot)$ as $T \to \infty$, where $\nu_T(r) = B_T^{-1} z_{1,[Tr]}$ for $r \in [0, 1]$, $\{B_T : T \ge 1\}$ are nonrandom 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 (d) involve variables indexed by $t \le T$. These variables have the same distribution under H_0 and H_1 by the definition of $\{w_t : t \le T\}$ given earlier and Assumption 1. Consequently, 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, used to obtain uniformity of $\widehat{\gamma}_{2(j)} - \gamma_0 \to_p 0$ 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 here, 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 later.

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 (a) the unit root and deterministic regressors, $z_{1,t}$, come from a sequence rather than a triangular array and (b) $B_T^{-1} B_{T+m} = O(1)$. The reason that conditions (a) and (b) are sufficient for Assumption LS(d) is that Assumption LS(b) and the CMT imply that $\sup_{r\in[0,1]} \|\nu_T(r)\| \to_d \sup_{r\in[0,1]} \|\nu(r)\| < \infty$ a.s. The left side equals $\max_{t\le T} \|B_T^{-1} \times z_{1,t}\|$. If $z_{1,t}$ comes from a sequence, then this implies that $\max_{t\le T+m} \|B_{T+m}^{-1} z_{1,t}\| = O_p(1)$. Combined with condition (b), this yields Assumption LS(d). Condition (a) 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 (b) 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:

(a) $z_{1,t}$ contains a vector of polynomials in t with nonnegative exponents and/or a unit root random vector $z_{1,t}^*$ that satisfies $z_{1,t}^* = z_{1,t-1}^* + v_t$ for $t = 1, 2, \ldots$, where $z_{1,0}^* = O_p(1)$.
(b) $\sup_{t\ge1} E\|(w_{0,t}', v_t')\|^{\beta+\varepsilon} < \infty$ for some $\beta > 2$ and $\varepsilon > 0$.
(c) $\{(w_{0,t}', v_t')' : t \ge 1\}$ is a weakly stationary strong mixing sequence of mean-0 random variables with strong mixing numbers that satisfy $\sum_{r=1}^{\infty} \alpha^{1-2/\beta}(r) < \infty$. (A.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 $a \ge 0$, is r^a for $r \in [0, 1]$. In addition, the subvector 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

$$\Omega^* = Ev_1v_1' + \sum_{k=2}^{\infty} Ev_1v_k' + \sum_{k=2}^{\infty} Ev_kv_1' \tag{A.6}$$

for v_t defined in (A.5).

Sufficiency of (A.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 $a \ge 0$, we take the corresponding element of B_T to be T^a, and the polynomial element of $\nu_T(r)$ converges to the nonrandom 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 1, and hence a weak law of large numbers for triangular arrays of mean-0 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:

(a) condition (a) of (A.5) holds.
(b) $(w_{0,t}', v_t')' = \sum_{j=-\infty}^{\infty} C_j \varepsilon_{t-j}$ and $\{\varepsilon_t : t \ge 1\}$ are iid with mean 0 and variance $\Delta > 0$.
(c) $\sum_{j=-\infty}^{\infty} \|C_j\|_* < \infty$ and $\sum_{k=1}^{\infty} (\|\sum_{j=k}^{\infty} C_j\|_* + \|\sum_{j=k}^{\infty} C_{-j}\|_*) < \infty$, where $\|C_j\|_* = \max_k |\sum_\ell C_{j,k,\ell}|$ and $C_{j,k,\ell} = [C_j]_{k,l}$. (A.7)

The limit random vector $\nu(r)$ that arises in Assumption LS(b) is the same in this case as that defined in the paragraph containing (A.5). Sufficiency of the conditions in (A.7) follows from the theorem and its proof given by 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 before.

Assumption LS(e) is standard in the literature. It rules out the case in which one or more regressors 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 use this condition because it is not very restrictive, and its elimination would complicate the results and the proofs.

Lemma A.1. Assumptions 1 and LS imply that Assumptions 2–4 hold.

Comment. Analogs of Lemma A.1 could be established for other estimators, such as fully modified, ML, and various other asymptotically efficient estimators mentioned earlier. For brevity, we do not do this here.

A.3 Asymptotic Results

We now state the asymptotic results that justify using the subsample critical values introduced earlier. Define $\widehat{F}_{P,T}(x)$, $F_P(x)$, $q_{P,1-\alpha}$, $\widehat{q}_{P,1-\alpha}$, and P_∞ as in Section 4.2 (and analogously with R in place of P). 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 $\{w_{T,t} : t = T+1, \ldots, T+m\}$ is assumed to be independent of T, which is appropriate for fixed alternatives.

The main theoretical result of the article is the following.

Theorem 2. Suppose that Assumptions 1–5 hold. Then, as $T \to \infty$, (a) $P \to_d P_\infty$ under H_0 and H_1, (b) $\widehat{F}_{P,T}(x) \to_p F_P(x)$ for all x in a neighborhood of $q_{P,1-\alpha}$ under H_0 and H_1, (c) $\widehat{q}_{P,1-\alpha} \to_p q_{P,1-\alpha}$ under H_0 and H_1, (d) $\Pr(P > \widehat{q}_{P,1-\alpha}) \to \alpha$ under H_0, and (e) parts (a)–(d) hold with R in place of P.

Comments. 1. The asymptotic distribution of P under H_0 and H_1 is given in part (a) of Theorem 2.

2. Part (c) of Theorem 2 shows that the random critical value $\widehat{q}_{P,1-\alpha}$ 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 Theorem 2 shows that the asymptotic null rejection rate of the test is α, as desired.

4. Part (a) shows that P does not diverge to infinity as $T \to \infty$ under H_1. Hence P is not a consistent test. This is due to the assumption that the number, m, of postbreakdown 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 is an asymptotically unbiased test.

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

APPENDIX B: PROOFS

Proof of Theorem 1

Theorem 1 is a special case of Theorem 2 because Assumption S implies that Assumption 1, assumptions (a)–(c) of (A.5), Assumptions LS(a) and LS(e), and Assumption 5 hold with $z_{1,t} = (1, t, x'_{1,t})'$, $z_{2,t} = x_{2,t}$, $H = I_k$, and $B_T = \operatorname{diag}\{1, T, T^{1/2}, \ldots, T^{1/2}\}$, which in turn implies that Assumptions 1–5 hold by Lemma A.1 and (A.5). Assumption 1 is implied by Assumption S(b). Assumptions (a)–(c) of (A.5) are implied by Assumptions S(a), S(d), and S(b). Assumption LS(a) is implied by Assumptions S(c) and S(d). Assumption LS(e) is implied by Assumptions S(a), S(b), S(d), and S(e), with $\nu(r)$ being a vector Brownian motion with variance matrix Ω^*. Assumption 5 is implied by Assumption S(f).

Proof of Theorem 2

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

$$P_j(\gamma, \Omega) = \big(\mathbf{Y}_{j:(j+m-1)} - \mathbf{Z}_{j:(j+m-1)}\gamma\big)' \times \Omega\big(\mathbf{Y}_{j:(j+m-1)} - \mathbf{Z}_{j:(j+m-1)}\gamma\big) \quad \text{and} \quad (B.1)$$

$$P_j(\gamma) = P_j(\gamma, I_m), \quad \text{where } \mathbf{Z}_{r:s} = (z_r, \ldots, z_s)'.$$

Similarly, in the proof we take $\widehat{F}_{P,T}(x)$ and $F_P(x)$ to be defined with $P_j(\widehat{\beta}_{2(j)})$ and $P_j(\beta_0)$ replaced by $P_j(\widehat{\gamma}_{2(j)})$ and $P_j(\gamma_0)$.

We start by bounding the difference $P_j(\widehat{\gamma}) - P_j(\gamma_0)$, where $\widehat{\gamma}$ denotes $\widehat{\gamma}_{1:(T+m)}$ or $\widehat{\gamma}_{2(j)}$. For $\varepsilon > 0$, define the event $L_{1,T}(\varepsilon)$ by

$$\begin{aligned} L_{1,T}(\varepsilon) = \big\{ &\big\|B_T(\widehat{\gamma}_{1,1:(T+m)} - \gamma_{1,0})\big\| \le \varepsilon, \\ &\big\|\widehat{\gamma}_{2,1:(T+m)} - \gamma_{2,0}\big\| \le \varepsilon, \\ &\big\|B_T\big(\widehat{\gamma}_{1,2(j)} - \gamma_{1,0}\big)\big\| \le \varepsilon, \big\|\widehat{\gamma}_{2,2(j)} - \gamma_{2,0}\big\| \le \varepsilon, \\ &\forall j = 1, \ldots, T - m + 1\big\}. \end{aligned} \quad (B.2)$$

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

$$L_{2,T}(c) = \Big\{ \max_{t \le T+m} \|B_T^{-1} z_{1,t}\| \le c \Big\}. \quad (B.3)$$

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

$$L_T = L_{1,T}(\varepsilon_T) \cap L_{2,T}(c_T). \quad (B.4)$$

Then we have

$$\Pr(L_T) \to 1 \quad \text{and} \quad \Pr(\overline{L}_T) \to 0 \quad \text{as } T \to \infty, \quad (B.5)$$

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

Now, for $\widehat{\gamma} = (\widehat{\gamma}_1', \widehat{\gamma}_2')'$ equal to $\widehat{\gamma}_{1:(T+m)}$ or $\widehat{\gamma}_{2(j)}$ and for $j = 1, \ldots, T+1$, we have, on the set L_T,

$$\begin{aligned}
&|P_j(\widehat{\gamma}) - P_j(\gamma_0)| \\
&= \left| -2 \sum_{t=j}^{j+m-1} u_t z_t'(\widehat{\gamma} - \gamma_0) + \sum_{t=j}^{j+m-1} (z_t'(\widehat{\gamma} - \gamma_0))^2 \right| \\
&\le 2 \sum_{t=j}^{j+m-1} |u_t| \max_{s \le T+m} \|B_T^{-1} z_{1,s}\| \cdot \|B_T(\widehat{\gamma}_1 - \gamma_0)\| \\
&\quad + 2 \sum_{t=j}^{j+m-1} \|u_t z_{2,t}\| \cdot \|\widehat{\gamma}_2 - \gamma_0\| \\
&\quad + \sum_{t=j}^{j+m-1} \Big(\max_{s \le T+m} \|B_T^{-1} z_{1,s}\| \cdot \|B_T(\widehat{\gamma}_1 - \gamma_0)\| \\
&\qquad\qquad + \|z_{2,t}\| \cdot \|\widehat{\gamma}_2 - \gamma_0\| \Big)^2 \\
&\le 2 \sum_{t=j}^{j+m-1} |u_t| c_T \varepsilon_T + 2 \sum_{t=j}^{j+m-1} \|u_t z_{2,t}\| \varepsilon_T \\
&\quad + \sum_{t=j}^{j+m-1} (c_T \varepsilon_T + \|z_{2,t}\| \varepsilon_T)^2 \\
&= g_j(\varepsilon_T, c_T), \qquad \text{(B.6)}
\end{aligned}$$

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

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

$$\begin{aligned}
&\Pr\big(P_{T+1}\big(\widehat{\gamma}_{1:(T+m)}\big) \le x\big) \\
&= \Pr\big(\big\{P_{T+1}\big(\widehat{\gamma}_{1:(T+m)}\big) \le x\big\} \cap L_T\big) \\
&\quad + \Pr\big(\big\{P_{T+1}\big(\widehat{\gamma}_{1:(T+m)}\big) \le x\big\} \cap \overline{L}_T\big) \\
&\le \Pr\big(\{P_{T+1}(\gamma_0) \le x + g_j(\varepsilon_T, c_T)\} \cap L_T\big) + o(1) \\
&= \Pr(P_{T+1}(\gamma_0) \le x) + o(1) \\
&= \Pr(P_\infty \le x) + o(1), \qquad \text{(B.7)}
\end{aligned}$$

where the inequality holds by (B.5) and (B.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 (B.7) also holds with $\ge$ in place of $\le$ 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 nonrandom vectors $\{\gamma_j : j = 1, \ldots, T-m+1\}$, let $\widehat{F}_T(x, \{\gamma_j\})$ denote the empirical df based on $\{P_j(\gamma_j) : j = 1, \ldots, T-m+1\}$, that is,

$$\widehat{F}_T(x, \{\gamma_j\}) = \frac{1}{T-m+1} \sum_{j=1}^{T-m+1} \mathbb{1}(P_j(\gamma_j) \le x) \qquad \text{(B.8)}$$

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

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 $\mathbb{R}$ with bounded derivative 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, we could take $k(x) = \cos(\pi x)/2 + 1/2$ for $x \in (0, 1)$. For $\{\gamma_j\}$ as before, we define the smoothed df,

$$\widehat{F}_T(x, \{\gamma_j\}, h_T) = \frac{1}{T-m+1} \sum_{j=1}^{T-m+1} k\left(\frac{P_j(\gamma_j) - x}{h_T}\right), \qquad \text{(B.9)}$$

where $\{h_T : T \ge 1\}$ 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 we can take $h_T = \varepsilon_T^{1/4}$.

We have

$$\begin{aligned}
&|\widehat{F}_{P,T}(x) - F_P(x)| \le \sum_{i=1}^{4} D_{i,T}, \quad \text{where} \\
&D_{1,T} = \big|\widehat{F}_{P,T}(x) - \widehat{F}_T\big(x, \big\{\widehat{\gamma}_{(j)}\big\}, h_T\big)\big|, \\
&D_{2,T} = \big|\widehat{F}_T\big(x, \big\{\widehat{\gamma}_{(j)}\big\}, h_T\big) - \widehat{F}_T(x, \{\gamma_0\}, h_T)\big|, \\
&D_{3,T} = \big|\widehat{F}_T(x, \{\gamma_0\}, h_T) - \widehat{F}_T(x, \{\gamma_0\})\big|, \qquad \text{and} \\
&D_{4,T} = |\widehat{F}_T(x, \{\gamma_0\}) - F_P(x)|. \qquad \text{(B.10)}
\end{aligned}$$

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

We have

$$D_{1,T} \le \frac{1}{T-m+1} \sum_{j=1}^{T-m+1} \mathbb{1}\big(P_j\big(\widehat{\gamma}_{2(j)}\big) - x \in (0, h_T)\big), \qquad \text{(B.11)}$$

because $\widehat{F}_{P,T}(x)$ and $\widehat{F}_T(x, \{\widehat{\gamma}_{2(j)}\}, h_T)$ differ only when $(P_j(\widehat{\gamma}_{2(j)}) - x)/h_T \in (0, 1)$.

Now, for all $\delta > 0$,

$$\begin{aligned}
&\Pr(D_{1,T} > \delta) \\
&\le \Pr(\{D_{1,T} > \delta\} \cap L_T) + \Pr(\overline{L}_T) \\
&\le \Pr\Bigg(\frac{1}{T-m+1} \sum_{j=1}^{T-m+1} \mathbb{1}\big(P_j(\gamma_0) - x \\
&\qquad \in \big(-g_j(\varepsilon_T, c_T), h_T + g_j(\varepsilon_T, c_T)\big)\big) > \delta \Bigg) + o(1)
\end{aligned}$$

$$\le E\mathbb{1}\big(P_1(\gamma_0) - x \in \big(-g_1(\varepsilon_T, c_T),\ h_T + g_1(\varepsilon_T, c_T)\big)\big)/\delta + o(1)$$
$$= o(1), \tag{B.12}$$

where the second inequality holds using (B.5), (B.6), and (B.11), the third inequality uses Markov's inequality and the identical distributions of $P_j(\gamma_0)$ for $j = 1, \ldots, T-m+1$, 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 $\Pr(P_1(\beta_0) \ne x) = 1$ 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), it remains to show that $D_{2,T} \to_p 0$. By mean-value expansions about $P_j(\gamma_0)$, we have: in the set L_T,

$$D_{2,T} = \left| \frac{1}{T-m+1} \sum_{j=1}^{T-m+1} k'\left(\frac{\widetilde{P}_j - x}{h_T}\right) \frac{P_j(\widehat{\gamma}_{2(j)}) - P_j(\gamma_0)}{h_T} \right|$$
$$\le \frac{B}{T-m+1} \sum_{j=1}^{T-m+1} \frac{g_j(\varepsilon_T, c_T)}{h_T}, \tag{B.13}$$

where $k'(\cdot)$ denotes the derivative of $k(\cdot)$, $\widetilde{P}_j$ lies between $P_j(\widehat{\gamma}_{2(j)})$ and $P_j(\gamma_0)$, $B < \infty$ denotes the bound on the derivative of $k(\cdot)$, and the inequality holds by (B.6).

By the dominated convergence theorem,

$$\frac{Eg_1(\varepsilon_T, c_T)}{h_T} \to 0 \quad \text{as } T \to \infty, \tag{B.14}$$

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

$$\Pr(D_{2,T} > \delta) \le \Pr(\{D_{2,T} > \delta\} \cap L_T) + \Pr(\overline{L}_T)$$
$$\le \Pr\left(\frac{B}{T-m+1} \sum_{j=1}^{T-m+1} \frac{g_j(\varepsilon_T, c_T)}{h_T} > \delta\right) + o(1)$$
$$\le \delta^{-1} BE \frac{g_1(\varepsilon_T, c_T)}{h_T} + o(1)$$
$$= o(1), \tag{B.15}$$

where the second inequality holds by (B.5) and (B.13), the third inequality holds by Markov's inequality and the identical distributions of $\{g_j(\varepsilon_T, c_T) : j = 1, \ldots, T-m+1\}$, and the equality holds by (B.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 $\varepsilon > 0$, $\widehat{F}_{P,T}(q_{P,1-\alpha} - \varepsilon) \to_p F_P(q_{P,1-\alpha} - \varepsilon) < 1 - \alpha$ and $\widehat{F}_{P,T}(q_{P,1-\alpha} + \varepsilon) \to_p F_P(q_{P,1-\alpha} + \varepsilon) > 1 - \alpha$.

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

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

$$P_j(\gamma, A_m) = \big(C\mathbf{Y}_{j:(j+m-1)} - C\mathbf{Z}_{j:(j+m-1)}\gamma\big)' \times \big(C\mathbf{Y}_{j:(j+m-1)} - \mathbf{CZ}_{j:(j+m-1)}\gamma\big). \tag{B.16}$$

Define

$$(\widetilde{u}_j, \ldots, \widetilde{u}_{j+m-1})' = \widetilde{\mathbf{U}}_{j:(j+m-1)} = C\mathbf{U}_{j:(j+m-1)},$$
$$(\widetilde{z}_j, \ldots, \widetilde{z}_{j+m-1})' = \widetilde{\mathbf{Z}}_{j:(j+m-1)} = C\mathbf{Z}_{j:(j+m-1)}, \tag{B.17}$$
$$\widetilde{z}_t = (\widetilde{z}'_{1,t}, \widetilde{z}'_{2,t})',$$

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

$$\widetilde{z}_{1,t} = \widetilde{\mathbf{Z}}'_{1,j:(j+m-1)} c_{t-j+1}, \quad \text{where } C = (c_1, \ldots, c_m)', \tag{B.18}$$

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

An analog of (B.6) holds with $P_j(\gamma)$ replaced by $P_j(\gamma, A_m)$ by replacing u_t, z_t, $z_{2,t}$, and $\max_{s \le T+m} \|B_T^{-1} z_{1,s}\|$ by $\widetilde{u}_t$, $\widetilde{z}_t$, $\widetilde{z}_{2,t}$, and $m^2 \max_{s \le T+m} \|B_T^{-1} z_{1,s}\|$, provided that $L_{2T}(c)$ is defined with c replaced by c/m^2 on the right side of (B.3). This holds because

$$\max_{j=1,\ldots,T-m+1} \max_{t=j,\ldots,j+m-1} \|B_T^{-1} \widetilde{z}_{1,t}\|$$
$$= \max_{j=1,\ldots,T-m+1} \max_{t=j,\ldots,j+m-1} \|B_T^{-1}(z_{1,j}, \ldots, z_{1,j+m-1}) c_{t-j+1}\|$$
$$\le m^2 \max_{s \le T+m} \|B_T^{-1} z_{1,s}\|, \tag{B.19}$$

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 (B.6) but with u_t and $z_{2,t}$ replaced by $\widetilde{u}_t$ and $\widetilde{z}_{2,t}$. 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).

Proof of Lemma A.1

To establish Assumption 4, we start by showing that $B_T(\widehat{\gamma}_{1,1:T} - \gamma_{1,0}) \to_p 0$, where $\widehat{\gamma}_{\ell,1:T}$ denotes the LS estimator based on $t = 1, \ldots, T$ for $\ell = 1, 2$. First, note that

$$T^{-1} \sum_{t=1}^{T} B_T^{-1} z_{1,t} z'_{1,t} B_T^{-1} = \int_0^1 \nu_{1,T}(r) \nu_{1,T}(r)' \, dr, \tag{B.20}$$

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,

$$B_T(\widehat{\gamma}_{1,1:T} - \gamma_{1,0})$$
$$= \left(\int_0^1 \nu_{1,T}(r) \nu_{1,T}(r)' \, dr - \nu_{1,2,T} \left(T^{-1} \sum_{t=1}^{T} z_{2,t} z'_{2,t} \right)^{-1} \nu'_{1,2,T} \right)^{-1}$$
$$\times \left(T^{-1} \sum_{t=1}^{T} B_T^{-1} z_{1,t} u_t - \nu_{1,2,T} \left(T^{-1} \sum_{t=1}^{T} z_{2,t} z'_{2,t} \right)^{-1} T^{-1} \sum_{t=1}^{T} z_{2,t} u_t \right)$$
$$= o_p(1), \tag{B.21}$$

where the second equality holds because (a) $\nu_{1,2,T} \to_p 0$ by Assumption LS(c); (b) $T^{-1}\sum_{t=1}^{T} z_{2,t}z_{2,t}' \to_p \Sigma_{2,0} > 0$ by the ergodic theorem and Assumptions 1, LS(a), and LS(e); (c) 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; (d) $T^{-1} \times \sum_{t=1}^{T} B_T^{-1} z_{1,t}u_t \to_p 0$ by Assumption LS(c); and (e) $T^{-1} \times \sum_{t=1}^{T} z_{2,t}u_t \to_p 0$ by the ergodic theorem and Assumptions 1 and LS(a).

Similarly, we have

$$\begin{aligned}
&\widehat{\gamma}_{2,1:T} - \gamma_{2,0} \\
&\quad = \Bigg(T^{-1}\sum_{t=1}^{T} z_{2,t}z_{2,t}' \\
&\qquad - \nu_{1,2,T}'\Bigg(\int_0^1 \nu_{1,T}(r)\nu_{1,T}(r)'\,dr\Bigg)^{-1}\nu_{1,2,T}\Bigg)^{-1} \\
&\qquad \times \Bigg(T^{-1}\sum_{t=1}^{T} z_{2,t}u_t - \nu_{1,2,T}'\Bigg(\int_0^1 \nu_{1,T}(r)\nu_{1,T}(r)'\,dr\Bigg)^{-1} \\
&\qquad\qquad \times T^{-1}\sum_{t=1}^{T} B_T^{-1} z_{1,t}u_t\Bigg) \\
&\quad = o_p(1). \qquad \text{(B.22)}
\end{aligned}$$

Now, to establish Assumption 4 for $\widehat{\gamma}_{\ell,1:(T+m)}$ for $\ell = 1, 2$, it suffices to show that

$$\begin{aligned}
K_{1,T} &= \Bigg\| T^{-1}\sum_{t=T+1}^{T+m} B_T^{-1} z_{1,t}z_{1,t}' B_T^{-1} \Bigg\| = o_p(1), \\
K_{2,T} &= \Bigg\| T^{-1}\sum_{t=T+1}^{T+m} z_{2,t}z_{2,t}' \Bigg\| = o_p(1), \\
K_{3,T} &= \Bigg\| T^{-1}\sum_{t=T+1}^{T+m} z_{2,t}u_t \Bigg\| = o_p(1), \qquad \text{(B.23)} \\
K_{4,T} &= \Bigg\| T^{-1}\sum_{t=T+1}^{T+m} B_T^{-1} z_{1,t}z_{2,t}' \Bigg\| = o_p(1), \\
K_{5,T} &= \Bigg\| T^{-1}\sum_{t=T+1}^{T+m} B_T^{-1} z_{1,t}u_t \Bigg\| = o_p(1).
\end{aligned}$$

These conditions are sufficient because (B.21) and (B.22) show that the differences between $B_T(\widehat{\gamma}_{\ell,1:T} - \gamma_{\ell,0})$ and $B_T(\widehat{\gamma}_{\ell,1:(T+m)} - \gamma_{\ell,0})$ are captured by the terms in (B.23).

We have

$$K_{1,T} \le T^{-1} m \max_{t=T+1,\ldots,T+m} \|B_T^{-1} z_{1,t}\|^2 = o_p(1), \qquad \text{(B.24)}$$

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

$$\begin{aligned}
&\Pr\Bigg(\Bigg\| T^{-1}\sum_{i=1}^{m} z_{2,T+i}z_{2,T+i}' \Bigg\| > \varepsilon\Bigg) \\
&\quad = \Pr\Bigg(\Bigg\| \sum_{i=1}^{m} z_{2,T+i}z_{2,T+i}' \Bigg\| > T\varepsilon\Bigg) \\
&\quad = o(1), \qquad \text{(B.25)}
\end{aligned}$$

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

$$K_{4,T} \le T^{-1}\sum_{t=T+1}^{T+m} \|z_{2,t}\| \cdot \max_{r\in[0,1]} \|\nu_T(r)\| = o_p(1), \qquad \text{(B.26)}$$

where the equality holds by Assumption LS(b) and the argument in (B.25). An analogous argument with $\|z_{2,t}\|$ replaced by $|u_t|$ gives $K_{5,T} = o_p(1)$. This establishes the result of Assumption 4 for $\widehat{\gamma}_{\ell,1:(T+m)}$.

To obtain the result of Assumption 4 for $\widehat{\gamma}_{\ell,2(j)}$ for $\ell = 1, 2$, it suffices to show that the conditions in (B.23) hold with the sums being over $t = j, \ldots, j + M - 1$, where $M = \lceil m/2 \rceil$, rather than $t = T + 1, \ldots, T + m$, and with the max over $j = 1, \ldots, T + M$ added. These conditions are sufficient because (B.21) and (B.22) show that the differences between $B_T(\widehat{\gamma}_{\ell,1:(T+m)} - \gamma_{\ell,0})$ and $B_T(\widehat{\gamma}_{\ell,2(j)} - \gamma_{\ell,0})$ are captured by the terms in (B.23) with the adjustments just described. Let $K_{i,T}^* = o_p(1)$ for $i = 1, \ldots, 5$ denote the conditions in (B.23) with these changes.

We have

$$K_{1,T}^* \le T^{-1} M \sup_{r\in[0,1]} \|\nu_T(r)\|^2 = T^{-1}O_p(1) = o_p(1), \qquad \text{(B.27)}$$

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

To establish the conditions of (B.23) for $K_{2,T}^*$–$K_{5,T}^*$, we use the following result. Suppose that $\{\xi_t : t \ge 1\}$ is a sequence of mean-0 random variables and that $\sup_{t\ge1} E\|\xi_t\|^{1+\delta} < \infty$ for some $\delta > 0$. Let $\tau_j = \sum_{t=j}^{j+M-1} \xi_t$. Then, for all $\varepsilon > 0$,

$$\begin{aligned}
&\Pr\Big(T^{-1}\max_{j\le T-M+1} \|\tau_j\| > \varepsilon\Big) \\
&\quad = \Pr\Bigg(\bigcup_{j=1}^{T-M+1} \{\|\tau_j\| > T\varepsilon\}\Bigg) \\
&\quad \le \sum_{j=1}^{T-M+1} \Pr(\|\tau_j\| > T\varepsilon) \\
&\quad \le (T - M + 1)E\|\tau_j\|^{1+\delta} T^{-(1+\delta)} \varepsilon^{-(1+\delta)} \\
&\quad = o(1), \qquad \text{(B.28)}
\end{aligned}$$

where the second inequality uses Markov's inequality. Hence

$$\max_{j\le T-M+1} \Bigg\| T^{-1}\sum_{t=j}^{j+M-1} \xi_t \Bigg\| \to_p 0. \qquad \text{(B.29)}$$

Applying (B.29) with $\xi_t = z_{2,t}z_{2,t}' - Ez_{2,t}z_{2,t}'$ gives $K_{2,T}^* \to_p 0$ 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 $t \le T$ by Assumption 1.

Applying (B.29) with $\xi_t = z_{2,t}u_t$ gives $K^*_{3,T} \to_p 0$ using the fact that $E\|z_{2,t}u_t\|^{1+\delta} < \infty$ by Assumption LS(a).

For $K^*_{4,T}$, we have

$$K^*_{4,T} \le \max_{j=1,\ldots,T+M-1} T^{-1} \sum_{t=j}^{j+M-1} \|z_{2,t}\| \cdot \max_{r\in[0,1]} \|\nu_T(r)\| = o_p(1), \tag{B.30}$$

where the equality holds by Assumption LS(b) and by applying (B.29) 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 $K^*_{5,T} \to_p 0$ because $E|u_t|^{1+\delta} < \infty$ by Assumption LS(a). This establishes the result of Assumption 4 for $\widehat{\gamma}_{\ell,2(j)}$ for $\ell = 1, 2$.

Assumption LS(a) obviously implies Assumption 2.

Finally, Assumption LS(b) and the CMT imply that

$$\max_{t\le T} \|B_T^{-1} z_{1,t}\| = \sup_{r\in[0,1]} \|\nu_T(r)\| \to_d \sup_{r\in[0,1]} \|\nu(r)\| < \infty \quad \text{a.s.} \tag{B.31}$$

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

[Received January 2004. Revised March 2006.]

REFERENCES

Ahn, S. K., and Reinsel, G. C. (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.

——— (2003), "End-of-Sample Instability Tests," *Econometrica*, 71, 1661–1694.

Andrews, D. W. K., and Kim, J.-Y. (2003), "End-of-Sample Cointegration Breakdown Tests," Cowles Foundation Discussion Paper 1404, Yale University, available at *http://cowles.econ.yale.edu*.

Carstensen, K. (2006), "Stock Market Downswing and the Stability of EMU Money Demand," *Journal of Business & Economic Statistics*, 24, 395–402.

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

Dufour, J.-M., Ghysels, E., and Hall, A. (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.

Gardner, L. A. (1969), "On Detecting Changes in the Mean of Normal Variates," *The Annals of Mathematical Statistics*, 40, 116–126.

Hansen, B. E. (1992), "Tests for Parameter Instability in Regressions With $I(1)$ Processes," *Journal of Business & Economic Statistics*, 10, 321–336.

——— (2000), "Testing for Structural Change in Conditional Models," *Journal of Econometrics*, 97, 93–115.

Hansen, H., and Johansen, S. (1999), "Some Tests for Parameter Constancy in Cointegrated VAR Models," *Econometrics Journal*, 2, 306–333.

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.

——— (2002), "A Small-Sample Correction of the Test for Cointegration Rank in the Vector Autoregressive Model," *Econometrica*, 70, 1929–1962.

Kim, J.-Y. (1999), "Testing Structural Change in a Cointegration Relation," unpublished manuscript, State University of New York Albany, Dept. of Economics.

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

Kim, J.-Y., Belaire-Franch, J., and Amador, R. B. (2002), "Corrigendum to 'Detection of Change in Persistence of Linear Time Series'," *Journal of Econometrics*, 109, 389–392.

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

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

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

MacNeill, I. B. (1978), "Properties of Sequences of Partial Sums of Polynomial Regression Residuals With Application to Tests for Change of Regression at Unknown Times," *The Annals of Statistics*, 6, 422–433.

Nabeya, S., and Tanaka, K. (1988), "Asymptotic Theory of a Test for the Constancy of Regression Coefficients Against the Random-Walk Alternative," *The 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 Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

Nyblom, J., and Makelainen, T. (1983), "Comparisons of Tests for the Presence of 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.

Perron, P. (1991), "A Test for Changes in a Polynomial Trend Function for a Dynamic Time Series," unpublished manuscript, Princeton University, Dept. of Economics.

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 $\int_0^1 B\,dB'$," *Journal of Multivariate Analysis*, 24, 252–264.

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

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

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

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

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

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

Quintos, C. E., and Phillips, P. C. B. (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 Watson, M. (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.