Journal of Business & Economic Statistics, July 1992, Vol. 10, No. 3

Tests for Parameter Instability in Regressions with I(1) Processes

Bruce E. Hansen
Department of Economics, University of Rochester, Rochester, NY 14627

This article derives the large-sample distributions of Lagrange multiplier (LM) tests for parameter instability against several alternatives of interest in the context of cointegrated regression models. The fully modified estimator of Phillips and Hansen is extended to cover general models with stochastic and deterministic trends. The test statistics considered include the SupF test of Quandt, as well as the LM tests of Nyblom and of Nabeya and Tanaka. It is found that the asymptotic distributions depend on the nature of the regressor processes—that is, if the regressors are stochastic or deterministic trends. The distributions are noticeably different from the distributions when the data are weakly dependent. It is also found that the lack of cointegration is a special case of the alternative hypothesis considered (an unstable intercept), so the tests proposed here may also be viewed as a test of the null of cointegration against the alternative of no cointegration. The tests are applied to three data sets—an aggregate consumption function, a present value model of stock prices and dividends, and the term structure of interest rates.

KEY WORDS: Cointegration; Fully modified estimation; Quandt statistic; Parameter constancy; Structural change.

One potential problem with time series regression models is that the estimated parameters may change over time. A form of model misspecification, parameter nonconstancy, may have severe consequences on inference if left undetected. In consequence, many applied econometricians routinely apply tests for parameter change. The most common test is the sample split or Chow test (Chow 1960). This test is simple to apply, and the distribution theory is well developed. The test is crippled, however, by the need to specify a priori the timing of the (one-time) structural change that occurs under the alternative. It is hard to see how any nonarbitrary choice can be made independently of the data. In practice, the selection of the breakpoint is chosen either with historical events in mind or after time series plots have been examined. This implies that the breakpoint is selected conditional on the data and therefore conventional critical values are invalid. One can only conclude that inferences may be misleading.

An alternative testing procedure was proposed by Quandt (1960), who suggested specifying the alternative hypothesis as a single structural break of unknown timing. The difficulty with Quandt's test is that the distributional theory was unknown until quite recently. A distributional theory for this test statistic valid for weakly dependent regressors was presented independently by Andrews (1990), Chu (1989), and Hansen (1990). Chu considered as well the case of a simple linear time trend.

Another testing approach has developed in the statistics literature that specifies the coefficients under the alternative hypothesis as random walks. Recent expositions were given by Nabeya and Tanaka (1988), Nyblom (1989), and Hansen (1990).

The preceding works did not consider models with integrated regressors. This article makes such an extension. The test statistics mentioned previously are examined here in the context of cointegrating regressions, making use of the fully modified estimation method of Phillips and Hansen (1990). The asymptotic distributions of the test statistics are found to depend on the stochastic process describing the regressors. It emerges as an important conclusion that it is necessary to know the stochastic process of the regressors before one can apply the tests considered here.

An additional finding is that, since the alternative hypothesis of a random walk in the intercept is identical to *no cointegration*, the test statistics are tests of the null of cointegration against the alternative of no cointegration.

A related research effort by Zivot and Andrews (1992) and Banerjee, Lumsdaine, and Stock (1992) developed a distributional theory for the test of the unit-root hypothesis employed by Perron (1989). Perron specified the alternative to be a single structural break of known timing, but the aforementioned articles specify the time of the break as unknown. These articles address a different question (testing the unit-root hypothesis), although using similar methods.

Section 1 sets up the structure of the model, allowing for quite general stochastic and deterministic trends in the regressors. This model builds on and extends the setup used by Phillips and Hansen (1990) and Hansen

(1992a). Section 2 describes the method of fully modified estimation for the model. Section 3 describes the test statistics for parameter instability. Section 4 gives the asymptotic distribution theory for the test statistics under the null hypothesis. Section 5 discusses the alternative of no cointegration. Section 6 reports three simple applications of the tests. Section 7 concludes. Proofs are in the Appendix.

A GAUSS program that implements the estimation methods and test statistics is available on request from the author.

1. A COINTEGRATED REGRESSION MODEL

We will be working with a fairly standard multiple-regression model under cointegration. The cointegrating equations are

$$y_t = Ax_t + u_{1t}, \qquad t = 1, \ldots, n, \tag{1}$$

where the process $x_t = (x_{1t}' \; x_{2t}')'$ is determined by the equations

$$\begin{aligned} x_{1t} &= k_{1t} \\ x_{2t} &= \prod_1 k_{1t} + \prod_2 k_{2t} + x_{2t}^o \\ x_{2t}^o &= x_{2t-1}^o + u_{2t}. \end{aligned} \tag{2}$$

Define the vectors

$$\underset{m_1 \quad m_2 + p_2}{u_t' = (\, u_{1t}' \qquad u_{2t}' \,)}; \quad \underset{p_1 \quad p_2}{k_t' = (k_{1t}' \quad k_{2t}')}.$$

$\{u_t\}$ is a sequence of mean zero random vectors, but the elements of k_t are nonnegative integer powers of time. Note that x_{1t} has p_1 elements and x_{2t} has $m_2 + p_2$ elements. More specific assumptions regarding u_t and k_t will be made in Section 3. This model places the *trends* k_{1t} directly in the regression equation (I will speak of k_t as trends although it may contain a constant). The trends k_{2t} determine the behavior of the stochastic regressors x_{2t} but are excluded from the regression. If k_t contains a constant, we assume that it is an element of k_{1t} and thus enters the levels regression.

The notation of (2) may appear confusing at first. It turns out to be especially convenient for development of a full theory for the cointegrated regression model (1). In earlier work on cointegrated models, such as that of Johansen (1988), Park and Phillips (1989), and Phillips and Hansen (1990), allowing for general deterministic trends was intentionally excluded to allow the clean development of a large-sample distributional theory. Although theoretically elegant, this approach excludes many applications of interest.

In most applications, it seems most reasonable to specify the stochastic regressors x_{2t} as I(1) with a deterministic trend. In this case, k_t equals a constant and a linear time trend. If y_t and x_{2t} are *deterministically cointegrated*, using the terminology of Ogaki and Park (1990), then the levels regression need only contain a constant, so $x_{1t} = k_{1t} = 1$ and $k_{2t} = t$. On the other hand, if a time trend is required in the levels regression, then $x_{1t} = k_{1t} = k_t = (1, t)'$ and there is no k_{2t}. Another common specification is that there are no trends in the system, so $k_{1t} = k_t = 1$ and there is no k_{2t}.

Some applied researchers have considered using *breaking trend* functions in addition to the simple integer powers of time considered here. Although in principle it is straightforward to define the estimators and test statistics with these more general trend functions, this extension will not be considered in this article for two reasons. First, the restriction to powers of time simplifies the asymptotic theory. Standardized powers of time converge uniformly to limiting functions, but this is not true of discontinuous trend functions. See Zivot and Andrews (1992) for an econometric example of weak convergence with discontinuous functions. Second, breaking trends may only make sense in a probability model if the timing and magnitude of the break is allowed to be random. This simply reintroduces I(1) or I(2) components into the system that are already captured in system (1)–(2).

The following nuisance parameters play an important role in the formulation of the statistics we will be considering. Define the matrices

$$\begin{aligned} \Omega &= \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{n} E(u_j u_t') \\ \Lambda &= \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{t} E(u_j u_t') \end{aligned} \tag{3}$$

partitioned in conformity with u:

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}; \quad \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}.$$

When the vector u_t is weakly stationary, Ω is proportional to the spectral density matrix evaluated at frequency 0. It is sometimes referred to as the *long-run covariance matrix*.

We also define

$$\Omega_{1\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} \tag{4}$$

and

$$\Lambda_{21}^{+} = \Lambda_{21} - \Lambda_{22}\Omega_{22}^{-1}\Omega_{21}.$$

One may loosely call $\Omega_{1\cdot 2}$ the long-run variance of u_{1t}, conditional on u_{2t}. Λ_{21}^{+} represents the bias due to endogeneity of the regressors after the *fully modified correction* discussed in Section 2.

2. FULLY MODIFIED ESTIMATION

The constancy tests we will discuss require an estimate of A in (1) that has a mixture normal asymptotic distribution. For concreteness, we will consider the fully modified (FM) estimator of Phillips and Hansen (1990). Alternative estimators with the same asymptotic distribution include the maximum likelihood estimator (MLE) of Johansen (1988) or the "leads and lags" es-

timator of Saikkonen (1991) and Stock and Watson (1991). (See Phillips and Loretan [1991] for an informative review.)

2.1 Estimation of Covariance Parameters

The semiparametric method of Phillips and Hansen (1990) is a two-step estimator in which the first step estimates the covariance parameters $\Omega_{1\cdot 2}$ and Λ_{21}^{+} defined in (4). Our suggestion is to use a prewhitened kernel estimator with the plug-in bandwidth recommended by Andrews and Monahan (in press). We outline the procedure in this subsection.

First, estimate (1) by ordinary least squares (OLS), yielding the parameter estimates $\hat{A}$ and the residuals $\hat{u}_{1t} = y_t - \hat{A}x_t$. Second, estimate (2) by OLS in differences: $\Delta x_{2t} = \hat{\Pi}_1 \Delta k_{1t} + \hat{\Pi}_2 \Delta k_{2t} + \hat{u}_{2t}$, yielding the residuals $\hat{u}_{2t}$. Set $\hat{u}_t' = (\hat{u}_{1t}', \hat{u}_{2t}')$.

The covariance matrices Ω and Λ could be estimated directly from the residuals $\hat{u}_t$ via a kernel. In most applications, the cointegrating residuals $\hat{u}_{1t}$ have a significant degree of serial correlation. In this event, the kernel estimate will be highly biased, unless a large bandwidth parameter is used, which increases the variance of the estimator. In such cases, an estimator based on prewhitening is often preferable in moderate sample sizes. We suggest using a (vector autoregressive) VAR(1), although a higher order VAR could also be used. We first fit a VAR to the residuals $\hat{u}_t$: $\hat{u}_t = \hat{\phi}\hat{u}_{t-1} + \hat{e}_t$. A kernel estimator is then applied to the whitened residuals $\hat{e}_t$. These take the form

$$\hat{\Lambda}_e = \sum_{j=0}^{n} w(j/M) \frac{1}{n} \sum_{t=j+1}^{n} \hat{e}_{t-j}\hat{e}_t'$$

and

$$\hat{\Omega}_e = \sum_{j=-n}^{n} w(j/M) \frac{1}{n} \sum_{t=j+1}^{n} \hat{e}_{t-j}\hat{e}_t',$$

where $w(\cdot)$ is a weight function (or kernel) that yields positive semi-definite estimates and M is a bandwidth parameter. The estimator $\hat{\Omega}_e$ can be seen as a scaled estimate of the spectral density of e_t (when e_t is covariance stationary) and has its origin in the literature on spectral density estimation, which dates back to Parzen (1957).

The covariance parameter estimates of interest can be obtained by *recoloring*: $\hat{\Omega} = (I - \hat{\phi})^{-1}\hat{\Omega}_e(I - \hat{\phi}')^{-1}$ and $\hat{\Lambda} = (I - \hat{\phi})^{-1}\hat{\Lambda}_e(I - \hat{\phi}')^{-1} - (I - \hat{\phi})^{-1}\hat{\phi}\hat{\Sigma}$, where $\hat{\Sigma} = (1/n)\sum_{i=1}^{n}\hat{u}_t\hat{u}_t'$.

These estimates require a choice of kernel and bandwidth parameter. Any kernel that yields positive semi-definite estimates can be used. These include the Bartlett, Parzen, and quadratic spectral (QS) kernels. Andrews (1991) recommended the QS kernel, which takes the form

$$w(x) = \frac{25}{12\pi^2x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right).$$

In most applications, it appears that the choice of kernel is much less important than the choice of bandwidth parameter. For current consistency proofs (see Hansen 1992b), it is required that $M \to \infty$ at some rate slower than sample size. This, however, does not provide a useful guide for the selection of the bandwidth parameter in a particular application. In a recent article, Andrews (1991) provided some useful guidelines (based on the minimization of asymptotic truncated mean squared error) to their selection. He recommended a plug-in bandwidth estimator. For the Bartlett, Parzen, and QS kernels, the choices are

$$\text{Bartlett:} \quad \hat{M} = 1.1147(\hat{\alpha}(1)n)^{1/3}$$

$$\text{Parzen:} \quad \hat{M} = 2.6614(\hat{\alpha}(2)n)^{1/5}$$

$$\text{QS:} \quad \hat{M} = 1.3221(\hat{\alpha}(2)n)^{1/5},$$

where $\hat{\alpha}(1)$ and $\hat{\alpha}(2)$ are obtained from approximating parametric models. A particularly attractive choice suggested by Andrews is univariate (autoregressive) AR(1) models for each element, $\hat{e}_{at}$ $(a = 1, \ldots, p)$ of $\hat{e}_t$. Let $(\hat{\rho}_a, \hat{\sigma}_a^2)$ denote the autoregressive and innovation variance estimates for the ath element. Then

$$\hat{\alpha}(1) = \sum_{a=1}^{p} \frac{4\hat{\rho}_a^2\hat{\sigma}_a^2}{(1-\hat{\rho}_a)^6(1+\hat{\rho}_a)^2} \Bigg/ \sum_{a=1}^{p} \frac{\hat{\sigma}_a^2}{(1-\hat{\rho}_a)^4}$$

and

$$\hat{\alpha}(2) = \sum_{a=1}^{p} \frac{4\hat{\rho}_a^2\hat{\sigma}_a^2}{(1-\hat{\rho}_a)^8} \Bigg/ \sum_{a=1}^{p} \frac{\hat{\sigma}_a^2}{(1-\hat{\rho}_a)^4}.$$

The use of a plug-in bandwidth parameter has several advantages. First, it removes the arbitrariness associated with the choice of bandwidth. Many applied researchers have been frustrated with the semiparametric branch of the unit-root literature because the test outcomes sometimes depend on the choice of bandwidth. Second, simulation results of Park and Ogaki (1991) demonstrated that its use can dramatically improve the mean squared error of semiparametric estimates of cointegrating relationships.

2.2 Estimation of the Regression Parameters

Partition $\hat{\Lambda}$ and $\hat{\Omega}$ as Λ and Ω, set $\hat{\Omega}_{1\cdot 2} = \hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}$ and $\hat{\Lambda}_{21}^{+} = \hat{\Lambda}_{21} - \hat{\Lambda}_{22}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}$ and define the transformed dependent variable $y_t^{+} = y_t - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{u}_{2t}$. The FM estimator of A is then given by

$$\hat{A}^{+} = \left(\sum_{t=1}^{n}(y_t^{+}x_t' - (0 \quad \hat{\Lambda}_{21}^{+\prime}))\right)\left(\sum_{t=1}^{n} x_t x_t'\right)^{-1}.$$

Associated with these parameter estimates are the residuals $\hat{u}_{1t}^{+} = y_t^{+} - \hat{A}^{+}x_t$. One interesting feature of the FM estimates that will be important for our later developments is that

$$\frac{1}{n}\sum_{t=1}^{n} x_t\hat{u}_{1t}^{+\prime} = \begin{pmatrix} 0 \\ \hat{\Lambda}_{21}^{+} \end{pmatrix},$$

although in OLS regression the sum of the products of the regressors and residuals is identically 0. Thus the *scores* of the problem are the variables

$$\hat{s}_t = \left(x_t \hat{u}_{1t}^{+\prime} - \begin{pmatrix} 0 \\ \hat{\Lambda}_{21}^{+} \end{pmatrix} \right), \tag{5}$$

which satisfy $\Sigma_{t=1}^{n} \hat{s}_t = 0$.

3. TESTS FOR PARAMETER INSTABILITY

Hansen (1990) outlined a general theory of testing for parameter instability in econometric models. The test statistics can be derived as Lagrange multiplier (LM) tests in correctly specified likelihood problems. In this section, we describe these test statistics in the context of fully modified estimation of cointegrated regression models.

We can modify (1) to incorporate possible parameter instability by allowing A to depend on time:

$$y_t = A_t x_t + u_t. \tag{1}'$$

For all of the tests, the null hypothesis is that the coefficient A_t in (1)′ is constant, although the tests differ in the treatment of alternative hypotheses.

The first two tests model A_t is obeying a single structural break at time t, where $1 < t < n$:

$$\begin{aligned} A_i &= A_1, \qquad i \le t \\ &= A_2, \qquad i > t. \end{aligned}$$

The null hypothesis is $H_0 : A_1 = A_2$. For the first test, the timing of the structural break is known under the alternative $H_1 : A_1 \neq A_2$, t known. A test for H_0 against H_1 is given by the statistic

$$\begin{aligned} F_{nt} &= \text{vec}(S_{nt})'(\hat{\Omega}_{1\cdot 2} \otimes V_{nt})^{-1}\text{vec}(S_{nt}) \\ &= \text{tr}\{S_{nt}' V_{nt}^{-1} S_{nt} \hat{\Omega}_{1\cdot 2}^{-1}\}, \end{aligned}$$

where

$$S_{nt} = \sum_{i=1}^{t} \hat{s}_i, \tag{6}$$

$$V_{nt} = M_{nt} - M_{nt} M_{nn}^{-1} M_{nt}, \tag{7}$$

and

$$M_{nt} = \sum_{i=1}^{t} x_i x_i'. \tag{8}$$

For the second test, the timing of the structural break is treated as unknown: $H_2 : A_1 \neq A_2$, $[t/n] \in \mathcal{T}$, where $\mathcal{T}$ is some compact subset of (0, 1), and $[\cdot]$ denotes "integer part." This test statistic is simply

$$\text{SupF} = \sup_{t/n \in \mathcal{T}} F_{nt}.$$

The third and fourth tests model the parameter A_t as a martingale process: $A_t = A_{t-1} + \varepsilon_t$; $E(\varepsilon_t | \mathcal{T}_{t-1}) = 0$, $E(\varepsilon_t \varepsilon_t') = \delta^2 G_t$. In this context, the null hypothesis can be written as the constraint that the variance of the martingale differences is 0: $H_0 : \delta^2 = 0$. One alternative hypothesis is $H_3 : \delta^2 > 0$, $G_t = (\hat{\Omega}_{1\cdot 2} \otimes V_{nt})^{-1}$; $t/n \in \mathcal{T}$, with test statistic

$$\text{MeanF} = \frac{1}{n^*} \sum_{t/n \in \mathcal{T}} F_{nt}, \qquad \text{where } n^* = \sum_{t/n \in \mathcal{T}} 1.$$

The final alternative is $H_4 : \delta^2 > 0$, $G_t = (\hat{\Omega}_{1\cdot 2} \otimes M_{nn})^{-1}$, with test statistic

$$L_c = \text{tr}\left\{ M_{nn}^{-1} \sum_{t=1}^{n} S_t \hat{\Omega}_{1\cdot 2}^{-1} S_t' \right\}.$$

The F_{nt} test (fixed t) is computationally simple, corresponding to the classical Chow test or sample split test. The test statistic is computationally equivalent to estimating A_1 and A_2 on the two subsamples and testing their equivalence using a Wald test, using the variance estimate for the full-sample estimates. This can be easily seen if we consider the special case of least squares estimation on a single equation ($m_1 = 1$). Then note that

$$\begin{aligned} M_{nt}^{-1} S_{nt} &= \left(\sum_{1}^{t} x_i x_i' \right)^{-1} \sum_{1}^{t} x_i \hat{u}_i \\ &= \left(\sum_{1}^{t} x_i x_i' \right)^{-1} \sum_{1}^{t} x_i y_i - \left(\sum_{1}^{t} x_i x_i' \right)^{-1} \sum_{1}^{t} x_i x_i' \hat{A} = \hat{A}_t - \hat{A}; \end{aligned}$$

that is, the score from the first part of the sample, evaluated at the estimate from the full sample, is proportional to the difference between the estimates obtained from just the first part of the sample and the full sample. It follows (with a little algebra) that our statistic F_{nt} is essentially equivalent to the Wald statistic that tests the equivalence of $\hat{A}_t$ and $\hat{A}$. The only difference arises due to the choice of the variance estimates. It is well known that this Wald statistic is algebraically equivalent to the classic Chow statistic, which is based on the difference between the estimates obtained from the two subsamples. For example, see Snow and Im (1991).

The distributional theory developed for this test (asymptotic chi-squared) is only valid when t can be chosen independently of the sample. This is a restrictive assumption in practice and may be valid only when t is chosen in an arbitrary way, such as $t = n/2$. In this event, the test might have low power against many alternatives of interest.

The SupF test dates back to Quandt (1960). Several recent works have explored the distributional theory in several contexts—those of Andrews (1990), Chu (1989), and Hansen (1990). The only difficulty in implementation is the choice of the region $\mathcal{T}$. As pointed out by Anderson and Darling (1952) and emphasized by Andrews (1990), the region $\mathcal{T}$ must not include the endpoints 0 and 1; otherwise the test statistic will diverge to infinity almost surely. The fix suggested by Andrews is to select $\mathcal{T} = [.15, .85]$. Although a reasonable approach, this introduces an element of arbitrariness that dilutes the appeal of the test.

The MeanF test statistic is derived from a different hypothesis structure but is seen to be simply the average F_{nt} test. Although in principle the averaging can include all values of τ for which F_{nt} can be computed, in practice some trimming will be required (since F_{nt} will not be defined over all t). Thus the arbitrariness associated with the SupF test is not completely avoided. For the remainder of the article, we set $\mathcal{T} = [.15, .85]$ as for the SupF test.

The L_c test has a long history in statistical theory, although it has not been fully understood until quite recently. It was first proposed by Gardner (1969) as a Bayes test for structural change. It was later independently proposed by Pagan and Tanaka (1981), Nyblom and Makelainen (1983), and King (1987). These works all concerned tests on a single coefficient in a Gaussian linear regression model. First attempts at a large-sample distributional theory were made by Nyblom and Makelainen (1983), Nabeya and Tanaka (1988), and Leybourne and McCabe (1989). A fairly complete theory for maximum likelihood was given by Nyblom (1989) and was extended to general econometric estimators by Hansen (1990). It has the advantage that it is much easier to compute than the SupF and MeanF tests and requires no decisions for trimming, hence excluding any form of arbitrariness.

The three proposed tests—SupF, MeanF, and L_c—are all tests of the same null hypothesis but differ in their choice of alternative hypothesis. In practice, all of the tests will tend to have power in similar directions, so the choice may be made on the computational grounds that L_c is much easier to calculate. But the appropriate test statistic for a particular application should also depend on the purpose of the test. If the desire is to discover whether there was a swift *shift in regime*, then the SupF test is appropriate. On the other hand, if one is simply interested in testing whether or not the specified model is a good model that captures a stable relationship, the notion of martingale parameters is more appropriate, since it captures the notion of an unstable model that gradually shifts over time. If the likelihood of parameter variation is relatively constant throughout the sample, then the L_c statistic is the appropriate choice.

4. DISTRIBUTIONAL THEORY

The assumptions we require for the asymptotic distribution theory are summarized in the following. Let $\{\alpha_m\}$ denote the α-mixing (strong-mixing) coefficients for $\{u_t\}$.

Assumptions. For some $q > \beta > 5/2$,

1. $E(u_t) = 0$;
2. α_m are of size $-q\beta/[2(q - \beta)]$;
3. $\sup_{t\geq 1} E|u_t|^q < \infty$;
4. Ω as defined in (3) exists with finite elements;
5. $\Omega_{22} > 0$ and $\Omega_{1\cdot 2} > 0$;
6. $\text{rank}(\Pi_2) = p_2$;
7. $k_t = (1, t, t^2, \ldots, t^p)'$, $p = p_1 + p_2 - 1$; and
8. $M^5/n = O(1)$.

For the random sequence $\{u_t\}$, the assumptions impose weak dependence through fairly mild conditions on the strong mixing coefficients. The moment conditions are only slightly stronger than square integrability. Assumption 5 implies that the elements of x_{2t}^o are not mutually cointegrated, that x_{2t} does not contain a lagged dependent variable, and that x_{2t}^o and the error u_{1t} are not multicointegrated (see Granger and Lee 1990). Assumption 6 says that the vector k_2 plays a role in the asymptotic behavior of x_{2t}. Assumption 7 restricts the vector k_t to integer powers of time.

Set $Y_t = \Sigma_{j=1}^t u_j$. Our assumptions are sufficient for the following results:

$$(1/\sqrt{n})Y_{[nr]} \Rightarrow B(r) \equiv \text{BM}(\Omega); \tag{9}$$

$$(1/n)\sum_{t=1}^{[nr]} Y_t u_{t+1}' \Rightarrow \int_0^r B\,dB' + r\Lambda; \tag{10}$$

and

$$\hat{\Lambda} \to_p \Lambda, \ \hat{\Omega} \to_p \Omega. \tag{11}$$

Here and elsewhere, "$\Rightarrow$" denotes weak convergence of the associated probability measures with respect to the uniform metric, and $\text{BM}(\Omega)$ denotes a Brownian motion with covariance matrix Ω. The invariance principle (9) was shown by Herrndorf (1984). Convergence to the matrix stochastic integral (10) was shown by Hansen (1992c). Consistent covariance parameter estimation (11) was shown by Hansen (1992b).

We need to find a sequence of weight matrices that will appropriately standardize the regressors x_t and the estimates $\hat{A}^+$. We adopt a method from Hansen (1992a). Set $\delta_n = \text{diag}(1, n^{-1}, n^{-2}, \ldots, n^{-p})$ and $k(r) = (1, r, r^2, \ldots, r^p)'$. Thus

$$\delta_n k_{[nr]} \to k(r) \text{ as } n \to \infty \tag{12}$$

uniformly in r. Partition $\delta_n = \text{diag}(\delta_{1n}, \delta_{2n})$ and $k(\text{r}) = (k_1(r), k_2(r))'$ in conformity with k_t.

Equation (2) specifies that the stochastic regressors x_{2t} are driven by the processes k_{1t}, k_{2t}, and x_{2t}^o. Since k_{1t} is also in the levels regression, least squares will project x_{2t} orthogonal to k_{1t}, leaving only k_{2t} and x_{2t}^o. We would like to isolate the effects of the stochastic trends from the deterministic trends. Construct an $(m_2 + p_2) \times m_2$ matrix Π_2^* in the null space of Π_2. The matrix Π_2^* will then annihilate the remaining deterministic component k_{2t} from x_{2t}. Now define the weight matrix for x_{2t},

$$\Gamma_{2n} = \begin{pmatrix} \delta_{2n}(\Pi_2'\Pi_2)^{-1}\Pi_2' \\ (1/\sqrt{n})(\Pi_2^{*\prime}\Omega_{22}\Pi_2^*)^{-1/2}\Pi_2^{*\prime} \end{pmatrix},$$

and the weight matrix for x_t,

$$\Gamma_n = \begin{pmatrix} \delta_{1n} & 0 \\ -\Gamma_{2n}\Pi_1 & \Gamma_{2n} \end{pmatrix}.$$

To see that this is a good choice for weighting matrix, note that

$$\Gamma_n x_t = \begin{pmatrix} \delta_{1n} k_{1t} \\ \Gamma_{2n}(\Pi_2 k_{2t} + x_{2t}^o) \end{pmatrix}$$

$$= \begin{pmatrix} \delta_{1n} k_{1t} \\ \delta_{2n} k_{2t} + \delta_{2n}(\Pi_2'\Pi_2)^{-1}\Pi_2' x_{2t}^o \\ (1/\sqrt{n})(\Pi_2^{*\prime}\Omega_{22}\Pi_2^*)^{-1/2}\Pi_2^{*\prime} x_{2t}^o \end{pmatrix}, \quad (13)$$

so by (9) and (12),

$$\Gamma_n x_{[nr]} \Rightarrow \begin{pmatrix} k_1(r) \\ k_2(r) \\ W_2(r) \end{pmatrix} = X(r), \text{ say}, \quad (14)$$

where $W_2(r) = (\Pi_2^{*\prime}\Omega_{22}\Pi_2^*)^{-1/2}\Pi_2^{*\prime}B(r) \equiv \text{BM}(I_{m2})$. Since $\int_0^\tau XX' > 0$ for all $\tau > 0$ (Phillips and Hansen 1990, lemma A.2), Γ_n is an appropriate weighting matrix for the process x_t. Equation (14) says that x_t is asymptotically dominated by the trend processes $k_1(r)$ and $k_2(r)$ and an m_2-dimensional stochastic trend (W_2).

The test statistics of Section 3 are functions of partial sample sums. It will be convenient to express these sums as functions of the space [0, 1]. Specifically, define

$$M_n(\tau) = M_{n[n\tau]} = \sum_{i=1}^{[n\tau]} x_i x_i$$

and $V_n(\tau) = V_{n[n\tau]} = M_n(\tau) - M_n(\tau)M_n(1)^{-1}M_n(\tau)$. We can now find the function-space distributional limits of these random functionals.

Theorem 1.

(a) $\frac{1}{n}\Gamma_n M_n(\tau)\Gamma_n' \Rightarrow M(\tau) = \int_0^\tau XX'$;

(b) $\frac{1}{n}\Gamma_n V_n(\tau)\Gamma_n' \Rightarrow V(\tau) = M(\tau) - M(\tau)M(1)^{-1}M(\tau)$;

(c) $\frac{1}{\sqrt{n}}\Gamma_n \sum_{i=1}^{[n\tau]} x_t u_{1t}^{+\prime} \Rightarrow \int_0^\tau X dB_{1\cdot 2}' + \tau\begin{pmatrix} 0 \\ \Lambda_{21}^* \end{pmatrix}$,

where $u_{1t}^+ = u_{1t} - \Omega_{12}\Omega_{22}^{-1}u_{2t}$;

(d) $\sqrt{n}\Gamma_{2n}\hat{\Lambda}_{21}^+ \to_p \begin{pmatrix} 0 \\ \Lambda_{21}^* \end{pmatrix}$;

(e) $\sqrt{n}(\hat{A}^+ - A)\Gamma_n^{-1} \Rightarrow \int_0^1 dB_{1\cdot 2} X' \left(\int_0^1 XX'\right)^{-1}$,

where $B_{1\cdot 2} \equiv \text{BM}(\Omega_{1\cdot 2})$ is independent of $X(r)$, and $\Lambda_{21}^* = (\Pi_2^{*\prime}\Omega_{22}\Pi_2^*)^{-1/2}\Pi_2^{*\prime}\Lambda_{21}^+$.

All of the test statistics considered in Section 3 were functions of the stochastic process S_{nt}. Three of the tests were also functions of the process F_{nt}. It will be convenient to write these also as functionals on [0, 1]. Define

$$S_n(\tau) = S_{n[n\tau]} = \sum_{i=1}^{[n\tau]} \hat{s}_i$$

$$F_n(\tau) = F_{n[n\tau]}$$

$$= \text{vec}(S_n(\tau))'(\hat{\Omega}_{1\cdot 2} \otimes V_n(\tau))^{-1}\text{vec}(S_n(\tau)).$$

We are now in a position to analyze the asymptotic distribution of these processes.

Theorem 2.

(a) $\frac{1}{\sqrt{n}}\Gamma_n S_n(\tau) \Rightarrow S^*(\tau)\Omega_{1\cdot 2}^{1/2}$;

(b) $F_n(\tau) \Rightarrow F(\tau)$ on $\tau \in \mathcal{T}$;

where $S^*(\tau) = S(\tau) - M(\tau)M(1)^{-1}S(1)$, $S(\tau) = \int_0^\tau X dW_1$, $F(\tau) = tr(S^*(\tau)'V(\tau)^{-1}S^*(\tau))$, and $W_1 = \Omega_{1\cdot 2}^{1/2}B_{1\cdot 2} \equiv \text{BM}(I_{m1})$, independent of X.

The process $S^*(\tau)$ is a tied-down version of the process $S(\tau)$, which is a continuous time martingale. Conditional on $\mathcal{F}_x = \sigma(X(r): 0 \le r \le 1)$, the sigma field generated by the process $X(r)$, both are Gaussian processes. Their conditional covariance functions are given, for $\tau_1 \le \tau_2$:

$$E(\text{vec}(S(\tau_1))\text{vec}(S(\tau_2))'|\mathcal{F}_x) = I_{m2} \otimes M(\tau_1)$$

$$E(\text{vec}(S^*(\tau_1))\text{vec}(S^*(\tau_2))'|\mathcal{F}_x) = I_{m2} \otimes (M(\tau_1) - M(\tau_1)M(1)^{-1}M(\tau_2)).$$

The distributional theory for the limit process of $S_n(\tau)$ is analogous to the distributional theory that arises in models without trends; for example, see Nyblom (1989) or Hansen (1990). In models without trends, we find

$$\frac{1}{\sqrt{n}}S_n(\tau) \Rightarrow B^*(\tau) = B(\tau) - (\tau M)M^{-1}B(1)$$

$$= B(\tau) - \tau B(1),$$

a Brownian bridge. In this expression, $M > 0$ is a constant matrix. The difference between this result and Theorem 2 arises because, in models without trends, sample covariance matrices converge to constant matrices. In models with stochastic trends, sample covariance matrices are random variables that change over time. Thus the expression for $S^*(\tau)$ depends on the matrix process $M(\tau)$, representing the sample covariance structure of the regressors.

We can now give expressions for the asymptotic distribution of the test statistics from Section 3.

Theorem 3.

(a) $F \to_d \chi_b^2$, $b = (1 + p + m_2)m_1$;

(b) $\text{supF} \to_d \sup_{\tau \in \mathcal{T}} F(\tau)$;

(c) $\text{MeanF} \to_d \int_{\mathcal{T}} F(r)dr$;

(d) $L_c \to_d \int_0^1 tr(S^*(\tau)'M(1)^{-1}S^*(\tau))$.

The asymptotic distribution of the standard F test is chi-squared. This test, however, as suggested earlier, has limited applicability due to the restrictive nature of the alternative hypothesis involved. The other test sta-

tistics are nonstandard and depend on the nature of the trends in X (i.e., p, m_1, and m_2).

Theorem 3 shows that it is important to know the trend properties of the regressors before a parameter constancy test can be mounted. Since the asymptotic distributions only depend on a few parameters, appropriate critical values can be tabulated. If X contains only deterministic trends, then analytic methods are available. Nabeya and Tanaka (1988) derived the asymptotic distribution of L_c for $x_t = k_t$ (in our notation) and iid errors u_{1t}. In this case, they have found expressions for the characteristic function of the limiting distribution. Their method does not immediately extend to stochastic trends, so here I resort to simulation methods.

The asymptotic distributions are approximated by draws from samples of size 1,000 using iid normal pseudorandom numbers. The calculations were made in GAUSS386 using its random-number generator. Critical values for the three tests are tabulated in Tables 1, 2, and 3 for the single equation setting ($m_1 = 1$). The tables include $p = 0, 1, 2$ and $m_2 = 0, 1, 2, 3, 4$. For $m_2 = 0, 1, 2$, 25,000 replications were made. For $m_2 = 3, 4$, 10,000 replications were made. The critical values are noticeably different from those for the case of weakly dependent data. (For the supF statistic, see Andrews [1990, table 1]; for the MeanF and L_c statistics, see Hansen [1990, table 1].

The critical values of Table 1 are useful but require applied researchers to frequently look up tables when making computations. It is more convenient to have computer packages produce p values along with test statistics. What we want is a function p value $= p(x)$, which maps an observed test statistic x into the appropriate value in the range [0, 1]. This is of special interest when the p value falls into the range [0, .2]. Suppose that we can well approximate the function $p(x)$ by a low-order polynomial: $p(x) \doteq a_0 + a_1x + a_2x^2 + a_3x^3$. Then if we can obtain the parameters $a = (a_0, a_1, a_2, a_3)$, we can easily generate approximate asymptotic p values automatically in the course of calculating the statistic.

To calculate the parameters, I evaluated 38 upper percentiles from the Monte Carlo distributions, from .20 to .015 in steps of .005. Then I regressed the percentiles on a third-order polynomial in the associate critical values. For all cases, the fit was very good over this region. Experimentation with extending up to the .010 percentile indicated a worsened fit, so it was not done. The estimated parameters are reported in Table 1. On their own, they are not interesting. But when incorporated into a computer program, they reduce the need to use tables. The polynomials should only be viewed as approximations that can produce p values over the region [.20, .015]. This is not a major handicap, since a p value below .20 is rarely termed "significant."

The distributional theory of this section is asymptotic. An investigation of the behavior of the test statistics in finite samples was undertaken by Gregory and Nason (1991). These authors assessed the testing procedures described in this article by applying the tests in the context of a linear-quadratic model. Their Monte Carlo design involved sample sizes of 100, 200, and 500. They found that the tests exhibited very little size distortion in these samples. They also found that the tests had good power against simple structural breaks at the first, second, and third quarter of the sample. The power of the tests depended on a cost-of-adjustment parameter, which induces serial correlation into the cointegrating error. As the degree of serial correlation in the error increases, the power decreases. This is not entirely surprising, because a highly serially correlated error is close to a random walk, which is equivalent to a random walk in the intercept. The ability of the test to discriminate between these two cases breaks down, and the power falls. Overall, Gregory and Nason's study casts a favorable light on the finite-sample performance of the test statistics advocated here.

Table 1. Asymptotic Critical Values for SupF

		Significance level			p-value coefficients			
m_2	p	1%	5%	10%	a_0	a_1	a_2	a_3
0	1	16.4	12.9	11.2	1.954	−.373	.0245	−.00055
0	2	20.0	15.8	14.1	2.487	−.400	.0219	−.00041
1	0	16.2	12.4	10.6	1.960	−.350	.0213	−.00044
1	1	19.0	15.2	13.4	2.666	−.424	.0230	−.00043
1	2	22.0	17.8	15.9	3.480	−.505	.0250	−.00042
2	0	18.6	14.8	13.0	3.182	−.491	.0258	−.00046
2	1	21.4	17.3	15.3	3.652	−.511	.0243	−.00039
2	2	23.9	19.7	17.7	4.003	−.508	.0219	−.00032
3	0	21.0	17.2	15.3	2.882	−.403	.0193	−.00031
3	1	23.9	19.3	17.3	3.248	−.397	.0163	−.00023
3	2	26.0	21.4	19.4	4.488	−.523	.0206	−.00027
4	0	23.6	19.0	17.1	3.522	−.449	.0194	−.00028
4	1	25.2	21.2	19.1	4.030	−.472	.0187	−.00025
4	2	28.0	23.2	21.0	5.341	−.594	.0224	−.00029

NOTE: These tables are for the single equation model ($m_1 = 1$). Critical values are calculated by Monte Carlo simulation using samples of size 1,000. 25,000 replications were made for $m_2 = 0, 1, 2$. 10,000 replications were made for $m_2 = 3, 4$.

Table 2. Asymptotic Critical Values for MeanF

		Significance level			p-value coefficients			
m_2	p	1%	5%	10%	a_0	a_1	a_2	a_3
0	1	6.83	4.48	3.73	1.080	−.511	.0843	−.00478
0	2	8.85	6.22	5.11	1.595	−.613	.0818	−.00374
1	0	6.78	4.57	3.73	1.008	−.470	.0773	−.00438
1	1	8.61	6.22	5.20	1.386	−.501	.0629	−.00271
1	2	10.4	7.76	6.50	1.641	−.479	.0479	−.00163
2	0	8.50	6.17	5.18	1.477	−.557	.0729	−.00326
2	1	10.3	7.69	6.58	1.818	−.556	.0607	−.00223
2	2	11.9	9.12	7.88	2.121	−.550	.0489	−.00148
3	0	10.1	7.68	6.66	1.448	−.397	.0370	−.00117
3	1	12.0	9.21	7.89	2.22	−.580	.0520	−.00159
3	2	13.4	10.4	9.15	2.640	−.609	.0480	−.00128
4	0	11.7	9.08	7.87	2.162	−.563	.0505	−.00154
4	1	13.3	10.6	9.28	2.440	−.551	.0426	−.00113
4	2	15.0	11.9	10.4	3.287	−.702	.0512	−.00127

NOTE: See Note to Table 1.

5. A TEST OF COINTEGRATION AGAINST NO COINTEGRATION

Many applied econometricians believe that it is important that an econometric model be able to survive statistical tests of the assumptions underlying that model. In the case of estimating a cointegrating relationship, a natural hypothesis to test is that of cointegration itself. In contrast, most *cointegration tests*, such as those of Engle and Granger (1987), Stock and Watson (1988), Johansen (1988), and Phillips and Ouliaris (1990), take the null to be no cointegration. The one notable exception is the spurious regressor test of Park, Ouliaris, and Choi (1988).

The specification tests developed in Section 3 are clearly tests of the model of cointegration proposed by Granger (1981) and developed by Engle and Granger (1987). It is, of course, possible to generalize the definition of cointegration to allow a nonstationary linear relationship between the variables, but this would be a radical departure from the idea Granger originally put forward. But does the model of *no cointegration*, conventionally defined, fall into the set of alternatives considered by the specification tests considered here?

For simplicity, rewrite Model (1) as

$$y_t = A_1 + A_2 x_{2t} + u_t; \tag{15}$$

that is, assume that k_{1t} is simply a constant. Assume that y_t and x_{2t} are not cointegrated. This is equivalent to the statement that the error u_t is $I(1)$. Now we can decompose u_t as $u_t = W_t + v_t$, where W_t is a random walk (ΔW_t is white noise) and v_t is stationary. Thus Equation (15) can be written as

$$y_t = A_{1t} + A_2 x_{2t} + v_t, \tag{16}$$

where $A_{1t} = A_1 + W_t$. Equation (16) is a special case of (1)′, which is our model of cointegration with nonstationary coefficients. Specifically, we can see that no cointegration is equivalent to one coefficient, the intercept, following a random walk. This is a special case of the alternative hypothesis for which the L_c statistic is an LM test statistic. We conclude that L_c is a test of the null of cointegration against the alternative of no cointegration.

Table 3. Asymptotic Critical Values for L_c

		Significance level			p-value coefficients			
m_2	p	1%	5%	10%	a_0	a_1	a_2	a_3
0	1	.723	.468	.361	.927	−3.536	4.771	−2.225
0	2	.758	.480	.382	1.120	−3.644	4.155	−1.628
1	0	.898	.575	.450	.769	−3.432	5.471	−3.041
1	1	.959	.623	.497	.996	−3.493	4.311	−1.834
1	2	.999	.654	.520	1.171	−3.421	3.507	−1.240
2	0	1.03	.690	.556	.855	−3.829	6.085	−3.342
2	1	1.13	.778	.625	1.074	−3.658	4.358	−1.778
2	2	1.19	.814	.666	1.263	−3.511	3.404	−1.133
3	0	1.18	.834	.680	1.247	−3.393	3.235	−1.066
3	1	1.29	.901	.752	1.430	−3.623	3.185	−.959
3	2	1.33	.954	.793	1.496	−3.636	3.075	−.894
4	0	1.31	.934	.780	1.451	−3.515	2.942	−.841
4	1	1.45	1.03	.866	1.694	−3.835	2.992	−.795
4	2	1.51	1.10	.922	1.726	−3.729	2.792	−.716

NOTE: See Note to Table 1.

The SupF and MeanF statistics are not specifically targeted for the alternative of random-walk coefficients, but they will have asymptotic power against this alternative as well.

Interestingly, there is a connection between the SupF test, the MeanF test, and the spurious regressor test of Park et al. (1988). For each t, F_{nt} is the F test for the significance of sample-split slope dummies. In the terminology of Park et al. (1988), these are "spurious trends." By the same argument used by these authors, each F_{nt} statistic is consistent against the alternative of no cointegration. Therefore, the SupF and MeanF statistics will be also.

There is another intuitive concept linking no cointegration and parameter instability. Under the null of cointegration, regression coefficient estimates converge uniformly in different parts of the sample space to the cointegrating relationship. Under the alternative of no cointegration, however, the regression estimates converge to random variables, which will take on different values in different samples. See Phillips (1986). Thus sequential parameter estimates will display apparent parameter instability. This simple observation implies an important message. Rejection of the null of constant parameters does not imply the particular alternative the test was designed to detect. There are many possibilities. If the SupF test rejects, for example, it would be quite inappropriate to conclude (on this piece of evidence alone) that there were two *cointegrating regimes*, which shifted at a particular point in the sample. The only statistically justified conclusion is that the standard model of cointegration, including its implicit assumption of long-run stability of the cointegrating relationship, is rejected by the data.

6. APPLICATIONS

We now apply this testing method to three applications. In each example, the fully modified estimation method is used. The covariance parameters are estimated using a QS kernel on residuals prewhitened with a VAR(1). (All of the reported regressions were also estimated using the Parzen and Bartlett kernels, and the results were nearly identical.) The bandwidth parameter was selected according to the recommendations of Andrews (1991), using univariate AR(1) approximating models. In all regressions reported, the estimates and standard errors are the fully modified estimates of Phillips and Hansen (1990). The estimated plug-in bandwidth parameter ($\hat{M}$) is reported. All of the SupF and MeanF statistics are calculated using the trimming region [.15, .85]. The constancy test statistics are reported along with their asymptotic p values (in parentheses), which are calculated according to the method of Section 4.

6.1 Aggregate Consumption Function

The notion of an aggregate relationship between consumption and income has a long history in macroeconomics. In a recent article, Campbell (1987) showed that a strict infinite-horizon permanent-income model yields a cointegrating relationship between aggregate consumption and aggregate disposable income. We can now test the constancy of this cointegrating relationship.

The data, from Blinder and Deaton (1985), are seasonally adjusted aggregate quarterly U.S. consumption and total disposable income (DI_t) in real per capita units, for the period 1953:2–1984:4. Campbell estimated the equation for both total consumption (TC_t) and nondurables and services consumption (NDS_t). In the following regressions, a constant and a time trend were included:

$$\begin{aligned} TC_t &= \underset{(194)}{-113} + \underset{(1.83)}{-1.02t} + \underset{(.088)}{.982DI_t} \\ \hat{M} &= .89 \\ \text{SupF} &= 12.3 \quad (.15) \\ \text{MeanF} &= 6.2 \quad (.05) \\ L_C &= .51 \quad (.09); \end{aligned}$$

$$\begin{aligned} NDS_t &= \underset{(103)}{518} + \underset{(.97)}{2.96t} + \underset{(.047)}{.526DI_t} \\ \hat{M} &= .97 \\ \text{SupF} &= 8.7 \quad (>.20) \\ \text{MeanF} &= 3.0 \quad (>.20) \\ L_C &= .14 \quad (>.20). \end{aligned}$$

First, examine the fully modified estimates. It appears that the corrections are having an important effect. The OLS estimates of the first equation, for example, yield a coefficient for disposable income of .85 rather than the economically more plausible .98. Note that the estimated bandwidth parameter for both equations is less than 1, indicating that nearly all of the serial correlation in the residuals was captured by the prewhitening procedure.

The tests when applied to the first equation do not yield clear results, with p values ranging from .05 to .15. Although the evidence suggests that the relationship may indeed be unstable, the data are not sufficiently informative to be able to reject the null of stability. On the other hand, the second equation (for nondurables and services consumption) does not suggest instability at all, since none of the test statistics are significant at the 20% level. These results (for both equations) are robust to the choice of kernel and whether the equations are estimated after taking logarithms.

It is informative to visually examine the sequence of F statistics for structural change. Figures 1 and 2 display these sequences for each regression, along with 5% critical values for its largest value (SupF), its average value (MeanF), and for a fixed known breakpoint.

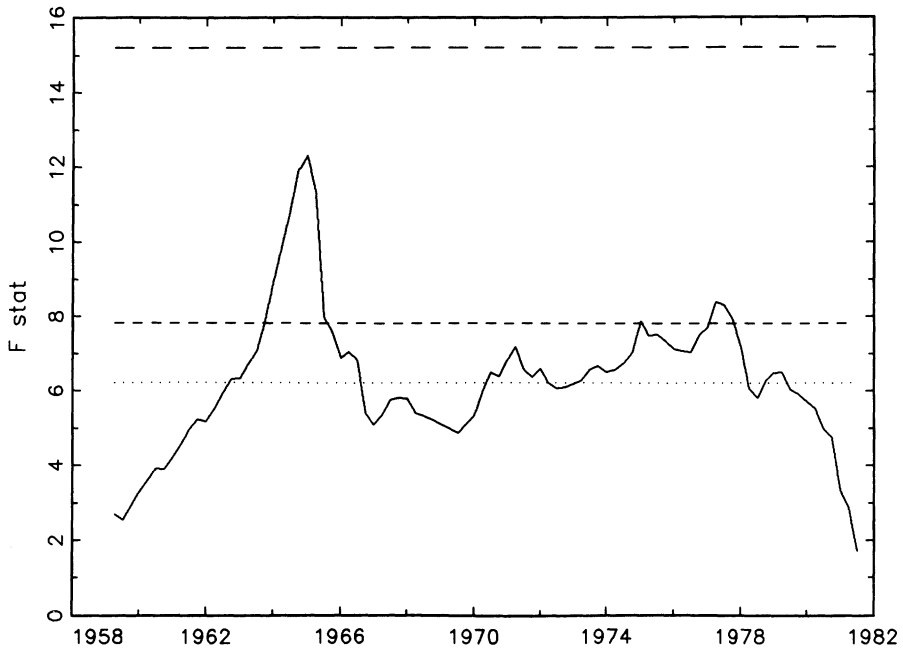

Figure 1. Total Consumption, 1953–1984: ——, F Statistic Sequence; – –, 5% Critical, SupF; . . . . . , 5% Critical, MeanF; ---, 5% Critical, Known Break.

6.2 Present-Value Model

Campbell and Shiller (1987) argued that a standard rational-expectations model of asset markets implies that real stock prices and dividends should be cointegrated. Using price (P_t) and (D_t) indexes for the period 1871–1986, they found evidence to support this claim. In a later series of articles, Campbell and Shiller (1988a,b) argued for a logarithmic approximation that implicitly assumes that the logarithms of the price and dividend indexes are cointegrated. Using their data, we can test

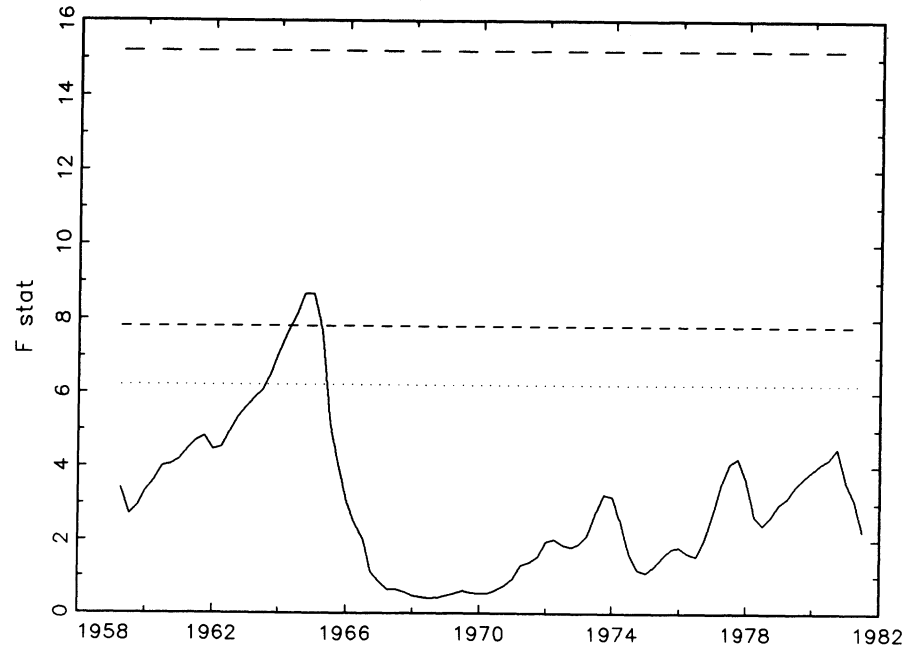

Figure 2. NDS Consumption, 1953–1984: ——, F Statistic Sequence; – –, 5% Critical, SupF; . . . . . , 5% Critical, MeanF; ---, 5% Critical, Known Break.

the stability of each specification:

$$\begin{aligned} P_t &= -.15 + 32.1\mathrm{D}_t \\ &\quad (.06) \quad (3.5) \\ \hat{M} &= .78 \\ \text{SupF} &= 6.9 \quad (>.20) \\ \text{MeanF} &= 3.3 \quad (.14) \\ L_C &= .30 \quad (>.20); \end{aligned}$$

$$\begin{aligned} \ln(P_t) &= 4.44 + 1.33 \ln(\mathrm{D}_t) \\ &\quad (.51) \quad (.12) \\ \hat{M} &= 1.07 \\ \text{SupF} &= 11.7 \quad (.06) \\ \text{MeanF} &= 5.1 \quad (.03) \\ L_C &= .35 \quad (.18). \end{aligned}$$

The levels equation yields estimates very close to those from OLS (which gives a slope coefficient of 31.1). This corresponds to a long-run real-interest rate of 3.1%, which, as noted by Campbell and Shiller (1987), is below the sample mean return of 8.2%. The relationship appears very stable, however, with no significant test statistics. The plot of the sequence of F statistics is displayed in Figure 3.

The logarithmic specification does not perform as well. The model predicts that the slope coefficient should be unity, but the point estimate is significantly above this value. The SupF and MeanF tests statistics suggests that the relationship is not stable. The plot of the sequence of F statistics is displayed in Figure 4.

This evidence suggests that stock prices and dividends are indeed cointegrated, but the logarithmic approximation used by Campbell and Shiller (1988a,b) may be misspecified.

6.3 Term Structure of Interest Rates

The theory of the term structure of interest rates suggests that, if interest rates can be characterized as $I(1)$ processes, then they should be cointegrated. Stock and Watson (1988), for example, tested for cointegration among three postwar U.S. interest rates and found evidence of two cointegrating vectors (i.e., only one common trend). They used monthly data from January 1960 to August 1979, presumably to exclude a possible regime shift in the term structure due to the change in the Federal Reserve's operating procedures in 1979. We now test the hypothesis that these relationships are stable over the entire period from January 1960 to March 1990. We use the same series—the federal funds rate (FF), the 90-day treasury-bill rate (TB3), and the one-year treasury-bill rate (TB12)—and obtained the series from the Citibase data base.

We report the results of two fully modified regressions, TB3 on TB12 and TB3 on FF (the regression of TB12 on FF yields results very similar to the regression

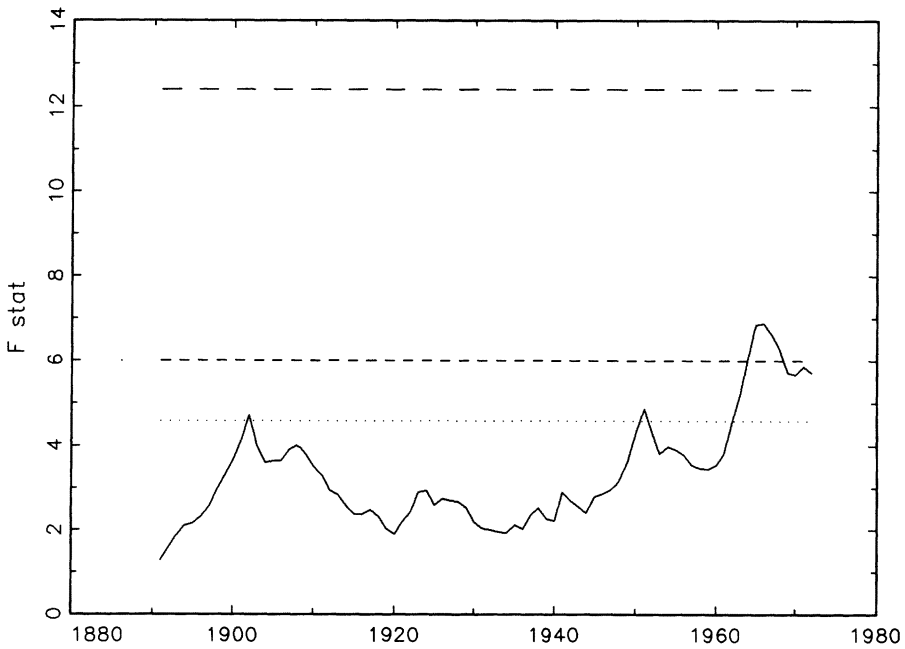

Figure 3. Stock Prices and Dividends, 1871–1986: ——, F Statistic Sequence; – –, 5% Critical, SupF; . . . . ., 5% Critical MeanF; ---, 5% Critical, Known Break.

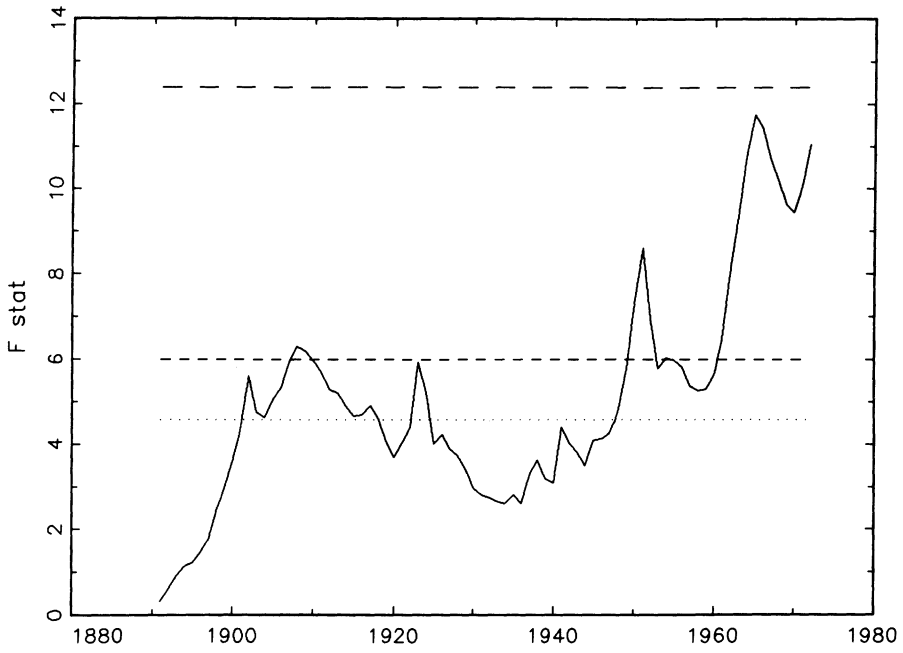

Figure 4. Logged Stock Prices and Dividends, 1871–1986: ——, F Statistic Sequence; – –, 5% Critical SupF; . . . . . , 5% Critical, MeanF; ---, 5% Critical, Known Break.

of TB3 on FF):

$$\begin{aligned} TB3_t &= -.62 + 1.06TB12_t \\ &\quad (.22) \quad (.03) \\ \hat{M} &= 2.54 \\ \text{SupF} &= 3.6 \quad (>.20) \\ \text{MeanF} &= 1.6 \quad (>.20) \\ L_C &= .21 \quad (>.20); \end{aligned}$$

$$\begin{aligned} TB3_t &= .49 + .83FF_t \\ &\quad (.14) \quad (.02) \\ \hat{M} &= 2.51 \\ \text{SupF} &= 22.8 \quad (.01) \\ \text{MeanF} &= 8.4 \quad (.01) \\ L_C &= .45 \quad (.06). \end{aligned}$$

Over the entire period, it appears that the two treasury-bill rates are cointegrated with a stable relationship, with a near-unity slope coefficient. This is strong support for the theory of the term structure. In contrast, the relationship of the treasury-bill rate with the federal-funds rate appears unstable, with the SupF and MeanF statistics highly significant. Figures 5 and 6 display the sequences of F statistics for the two regressions. The sequence for the second regression crosses the 5% SupF critical value several times, achieving its maximal value approximately in 1980. This supports the conjecture that the change in the Federal Reserve's operating procedures altered the relationship between some interest rates. It is interesting that this regime shift only appears to have affected the relationship between the federal-funds rate and the treasury-bill rates but not the relationship between the treasury-bill rates of different maturities.

7. CONCLUSION

As shown by example in Section 6, in some applications the three test statistics (SupF, MeanF, and L_c) may appear to be in conflict. There is no reason why all three tests should reject (or not reject) at a particular level of significance in a particular sample. The tests are looking in different directions and will have more power against some alternatives than others. All of the tests, however, will have asymptotic power against the same set of alternatives. The possibility of conflicting test statistics is not new to applied economists. There are many tests for heteroscedasticity, for unit roots, for cointegration, and so forth. The same care needs to be exercised in the present context. Calculation of all three test statistics seems the most judicious suggestion at this time.

The tests were described here using the Phillips–Hansen fully modified estimator. This is not the only possibility. It is quite straightforward to calculate the test statistics for other asymptotically efficient estimates of cointegrating vectors, such as the MLE due to Johansen (1988) or the "leads and lags" estimator of Saikkonen (1991) and Stock and Watson (1991). Since the estimators are asymptotically equivalent, the test statistics would have the same asymptotic distributions as those tabulated in this article. It is quite likely, however, that the asymptotic proofs would be more difficult.

This article only discussed joint tests on all of the

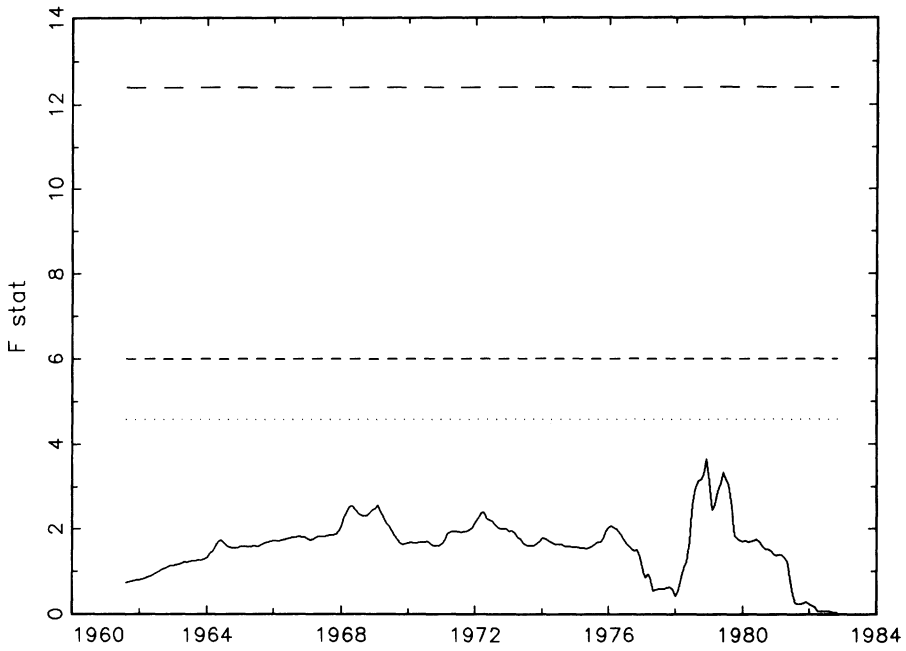

Figure 5. Three-Month and One-Year Treasury-Bill Rates, 1960–1990: ——, F Statistic Sequence; – –, 5% Critical SupF; . . . . . , 5% Critical MeanF; ---, 5% Critical, Known Break.

regression parameters in a cointegrating regression. It should be possible to extend these results to tests on a subset of the parameters as well. This will be left to future research.

ACKNOWLEDGMENTS

An earlier version of this article was entitled "Testing for Structural Change of Unknown Form in Models With Non-stationary Regressors." My thanks go to Don Andrews, Bill Brown, David Hendry, Peter Phillips, Adrian Pagan, Jim Stock, the associate editor, the referee, and seminar participants at the University of Rochester, Indiana University, and UCLA for helpful discussions and comments on earlier drafts. This research was supported by the National Science Foundation.

APPENDIX: PROOFS OF THE THEOREMS

Proof of Theorem 1

(a) The finite-dimensional result is immediate from (14). Weak convergence follows from the continuous

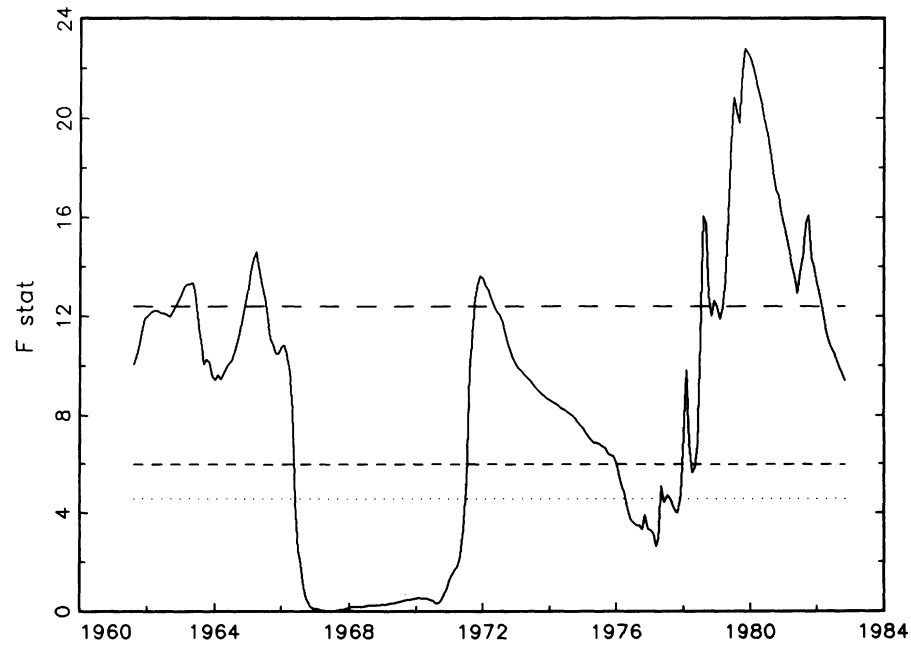

Figure 6. Three-Month Treasury-Bill and Federal Funds Rates, 1960–1990: ——, F Statistic Sequence; – –, 5% Critical, SupF; . . . . , 5% Critical, MeanF; ---, 5% Critical, Known Break.

mapping theorem (see Billingsley 1968, p. 30) since $M(\tau)$ is a continuous function of τ and $X(\cdot)$.

(b) This follows from part (a) and continuity.

(c) By (13) and theorem 4.1 of Hansen (1992c),

$$\frac{1}{\sqrt{n}}\sum_1^{[n\tau]}\Gamma_n x_t u_{1t}^{+\prime} = \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_1^{[n\tau]}\delta_n k_t u_{1t}^{+} + o_p(1) \\ \frac{1}{n}\sum_1^{[n\tau]}\left(\Pi_2^{*\prime}\Omega_{22}\Pi_2^{*}\right)^{-1}\Pi_2^{*\prime}x_{2t}^{o}u_{1t}^{+\prime} \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \int_0^\tau k dB_{1\cdot 2}' \\ \left(\Pi_2^{*\prime}\Omega_{22}\Pi_2^{*}\right)^{-1}\Pi_2^{*\prime}\left(\int_0^\tau B_2 dB_{1\cdot 2}' + \tau\Lambda_{21}^{+}\right) \end{pmatrix}$$

$$= \begin{pmatrix} \int_0^\tau k dB_{1\cdot 2}' \\ \int_0^\tau W_2 dB_{1\cdot 2}' + \tau\Lambda_{21}^{*} \end{pmatrix}.$$

(d) By (11),

$$\sqrt{n}\,\Gamma_{2n}\hat{\Lambda}_{21}^{+} = \begin{pmatrix} \sqrt{n}\,\delta_{2n}\left(\Pi_2'\Pi_2\right)^{-1}\Pi_2' \\ \left(\Pi_2^{*\prime}\Omega_{22}\Pi_2^{*}\right)^{-1/2}\Pi_2^{*\prime} \end{pmatrix}(\hat{\Lambda}_{21} - \hat{\Lambda}_{22}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21})$$

$$\to_p \begin{pmatrix} 0 \\ \left(\Pi_2^{*\prime}\Omega_{22}\Pi_2^{*}\right)^{-1/2}\Pi_2^{*\prime} \end{pmatrix}(\Lambda_{21} - \Lambda_{22}\Omega_{22}^{-1}\Omega_{21})$$

$$= \begin{pmatrix} 0 \\ \Lambda_{21}^{*} \end{pmatrix}.$$

(e)

$$\sqrt{n}\,(\hat{A}^{+} - A)\Gamma_n^{-1} = \left(\frac{1}{\sqrt{n}}\sum_1^n u_{1t}^{+}x_t'\Gamma_n' - \sqrt{n}\,(0\ \hat{\Lambda}_{21}^{+\prime})\Gamma_n' - (\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \Omega_{12}\Omega_{22}^{-1})\frac{1}{\sqrt{n}}\sum_1^n \hat{u}_{2t}x_t'\Gamma_n'\right) \times \left(\frac{1}{n}\Gamma_n\sum_1^n x_t x_t'\Gamma_n'\right)^{-1}$$

$$\Rightarrow \left(\int_0^1 dB_{1\cdot 2}X' + (0\ \Lambda_{21}^{+\prime}) - (0\ \Lambda_{21}^{+\prime})\right)\left(\int_0^1 XX'\right)^{-1}$$

$$= \int_0^1 dB_{1\cdot 2}X'\left(\int_0^1 XX'\right)^{-1}$$

by parts (a), (c), and (d), and (11).

Proof of Theorem 2

(a)

$$\frac{1}{\sqrt{n}}\Gamma_n S_n(\tau) = \frac{1}{\sqrt{n}}\Gamma_n\sum_{t=1}^{[n\tau]}\hat{s}_t = \frac{1}{\sqrt{n}}\Gamma_n\sum_{t=1}^{[n\tau]}\left(x_t\hat{u}_{1t}^{+\prime} - \begin{pmatrix} 0 \\ \hat{\Lambda}_{21}^{+} \end{pmatrix}\right)$$

$$= \frac{1}{\sqrt{n}}\Gamma_n\sum_{t=1}^{[n\tau]}x_t u_{1t}^{+\prime} - \frac{1}{n}\Gamma_n\sum_{t=1}^{[n\tau]}x_t x_t'\Gamma_n'^{-1}(\hat{A}^{+} - A)'\sqrt{n} - \frac{[n\tau]}{\sqrt{n}}\begin{pmatrix} 0 \\ \Gamma_{2n}\hat{\Lambda}_{21}^{+} \end{pmatrix}$$

$$\Rightarrow \int_0^\tau X dB_{1\cdot 2}' + \tau\begin{pmatrix} 0 \\ \Lambda_{21}^{*} \end{pmatrix} - \int_0^\tau XX'\left(\int_0^1 XX\right)^{-1}\int_0^1 X dB_{1\cdot 2}' - \tau\begin{pmatrix} 0 \\ \Lambda_{21}^{*} \end{pmatrix}$$

$$= S(\tau)\Omega_{1\cdot 2}^{1/2} - M(\tau)M(1)^{-1}S(1)\Omega_{1\cdot 2}^{1/2}$$

$$= S^{*}(\tau)\Omega_{1\cdot 2}^{1/2}$$

by Theorem 1 and the continuous mapping theorem.

(b)

$$F_n(\tau) = \mathrm{tr}\{S_n(\tau)'V_n(\tau)^{-1}S_n(\tau)\hat{\Omega}_{1\cdot 2}^{-1}\}$$

$$= \mathrm{tr}\left\{\frac{1}{\sqrt{n}}S_n(\tau)'\Gamma_n'\left(\frac{1}{n}\Gamma_n V_n(\tau)\Gamma_n'\right)^{-1}\frac{1}{\sqrt{n}}\Gamma_n S_n(\tau)\hat{\Omega}_{1\cdot 2}^{-1}\right\}$$

$$\Rightarrow \mathrm{tr}\{\Omega_{1\cdot 2}^{1/2}S^{*}(\tau)'V(\tau)^{-1}S^{*}(\tau)\Omega_{1\cdot 2}^{1/2}\Omega_{1\cdot 2}^{-1}\}$$

$$= \mathrm{tr}\{S^{*}(\tau)'V(\tau)^{-1}S^{*}(\tau)\}.$$

Proof of Theorem 3

Part (a) follows from Theorem 2 (b) and the fact that conditional on $\mathcal{F}_x$, for any $\tau \in \mathcal{T}$, $\mathrm{vec}(S^{*}(\tau)) \equiv N(0, \Omega_{1\cdot 2} \otimes V(\tau))$, and therefore

$$F(\tau)|_{\mathcal{F}_x} \equiv \chi_b^2.$$

Since this distribution is independent of $\mathcal{F}_x$, it is the unconditional distribution as well.

Part (b) follows from Theorem 2(b) and the continuous mapping theorem.

Part (c) follows from Theorem 2(b), the continuous mapping theorem and the fact that $\int_{\mathcal{T}}F(\tau)d\tau$ is well defined.

Part (d) follows from Theorem 1(a), Theorem 2(a), and (11).

[*Received January 1991. Revised January 1992.*]

REFERENCES

Anderson, T. W., and Darling, D. A. (1952), "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes," *The Annals of Mathematical Statistics*, 23, 193–212.

Andrews, D. W. K. (1990), "Tests for Parameter Instability and Structural Change With Unknown Change Point," Discussion Paper 943, Yale University, Cowles Foundation for Research in Economics.

——— (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Andrews, D. W. K., and Monahan, J. C. (in press), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60.

Banerjee, A., Lumsdaine, R. L., and Stock, J. H. (1992), "Recursive and Sequential Tests of the Unit Root and Trend Break Hypothesis Theory and International Evidence. *Journal of Business & Economic Statistics*, 10, 271–287.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: John Wiley.

Blinder, A. A., and Deaton, A. S. (1985) "The Time Series Consumption Function Revisited," *Brookings Papers on Economic Activity*, 465–511.

Campbell, J. Y. (1987), "Does Saving Anticipate Declining Labor

Income? An Alternative Test of the Permanent Income Hypothesis," *Econometrica*, 55, 1249–1273.

Campbell, J. Y., and Shiller, R. J. (1987), "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1088.

——— (1988a), "Stock Prices, Earning and Expected Dividends," *The Journal of Finance*, 43, 661–676.

——— (1988b), "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195–228.

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

Chu, C.-S. J. (1989), "New Tests for Parameter Constancy in Stationary and Nonstationary Regression Models," unpublished manuscript, University of California at San Diego, Dept. of Economics.

Engle, R. F., and Granger, C. W. J. (1987), "Co-integration and Error-correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

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

Granger, C. W. J. (1981), "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121–130.

Granger, C. W. J., and Lee, T.-H. (1990), "Multicointegration," *Advances in Econometrics*, 8, 71–84.

Gregory, A. W., and Nason, J. M. (1991), "Testing for Structural Breaks in Cointegrated Relationships," unpublished manuscript, Queen's University, Dept. of Economics.

Hansen, B. E. (1990), "Lagrange Multiplier Tests for Parameter Instability in Non-linear Models," unpublished manuscript, University of Rochester, Dept. of Economics.

——— (in press a), "Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends," *Journal of Econometrics*, 52.

——— (in press b), "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica*, 60.

——— (in press c), "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8.

Herrndorf, N. (1984), "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *The Annals of Probability*, 12, 829–839.

Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

King, M. L. (1987), "An Alternative Test for Regression Coefficient Stability," *The Review of Economics and Statistics*, 69, 379–381.

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

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. (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*, 84, 856–864.

Ogaki, M., and Park, J. Y. (1990), "A Cointegration Approach to Estimating Preference Parameters," Paper 236, Rochester Center for Economic Research.

Pagan, A. R., and Tanaka, K. (1981), "A Further Test for Assessing the Stability of Regression Coefficients," unpublished manuscript.

Park, J. Y., and Ogaki, M. (1991), "Inference in Cointegrated Models Using VAR Prewhitening to Estimate Shortrun Dynamics," Working Paper 281, Rochester Center for Economic Research.

Park, J. Y., Ouliaris, S., and Choi, B. (1988), "A New Approach to Testing for a Unit Root," mimeo, Cornell University, Dept. of Economics.

Park, J. Y., and Phillips, P. C. B. (1989), "Statistical Inference in Regressions With Integrated Processes, Part 1," *Econometric Theory*, 4, 468–497.

Parzen, E. (1957), "On Consistent Estimates of the Spectrum of a Stationary Time Series," *The Annals of Mathematical Statistics*, 28, 329–348.

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

Phillips, P. C. B. (1986), "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340.

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.

Phillips, P. C. B., and Ouliaris, S. (1990), "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165–193.

Quandt, R. (1960), "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324–30.

Saikkonen, P. (1991), "Asymptotically Efficient Estimation of Cointegrating Regressions," *Econometric Theory*, 7, 1–21.

Snow, M. S., and Im, E. I. (1991), "The Equivalence of Two Test Statistics for Testing the Constancy of Regression Coefficients," *Econometric Theory*, 7, 419–420.

Stock, J. H., and Watson, M. W. (1988), "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

——— (1991), "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," unpublished manucript.

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