



ELSEVIER

Journal of Econometrics 70 (1996) 99–126

---

---

JOURNAL OF  
Econometrics

---

---

## Residual-based tests for cointegration in models with regime shifts

Allan W. Gregory<sup>a</sup>, Bruce E. Hansen<sup>\*, b, c</sup>

<sup>a</sup> Department of Economics, Queen's University, Kingston, Ont. K7L 3N6, Canada

<sup>b</sup> Department of Economics, Boston College, Chestnut Hill, MA 02167-3806, USA

<sup>c</sup> Department of Economics, University of Rochester, Rochester, NY 14627, USA

---

### Abstract

In this paper we examine tests for cointegration which allow for the possibility of regime shifts. We propose  $ADF$ -,  $Z_t$ -, and  $Z_t$ -type tests designed to test the null of no cointegration against the alternative of cointegration in the presence of a possible regime shift. In particular we consider cases where the intercept and/or slope coefficients have a single break of unknown timing. A formal proof is provided for the limiting distributions of the various tests for the regime shift model (both a level and slope change). Critical values are calculated for the tests by simulation methods and a simple Monte Carlo experiment is conducted to evaluate finite-sample performance. In the limited set of experiments, we find that the tests can detect cointegrating relations when there is a break in the intercept and/or slope coefficient. For these same experiments, the power of the conventional ADF test with no allowance for regime shifts falls sharply. As an illustration we test for structural breaks in the U.S. long-run money-demand equation using annual and quarterly data.

*Key words:* Level shift; Regime shift; Cointegration; Brownian motion

*JEL classification:* C12; C15; C22; C52

---

### 1. Introduction

It is now routine for researchers to test for cointegration when working with multivariate time series. The most widely applied tests are residual-based ones in

---

\*Corresponding author.

We thank Gregor Smith for helpful comments and Kim Cook for excellent research assistance. We also thank the co-editors Jean-Marie Dufour and Eric Ghysels and two referees for helpful comments which improved the presentation of our results. The first author acknowledges financial support from the Social Sciences and Humanities Research Council of Canada. The second author thanks the National Science Foundation for financial support.

which the null hypothesis of no cointegration is tested against the alternative that the relation is cointegrated in the sense of Engle and Granger (1987), meaning that a linear combination of the integrated variables has a stationary distribution. A large sample distribution theory for this class of tests has been studied by Phillips and Ouliaris (1990). Rejection of the null hypothesis in this context implies the strong result that the variables are cointegrated. Acceptance of the null hypothesis is often taken as evidence of the lack of cointegration.

While these tests and the associated distributional theory are appropriate for the precise question of no cointegration versus cointegration, there are many related questions which may appear quite similar, but actually require a different set of tests and distributional theory. In this paper we are concerned with the possibility of a more general type of cointegration, where the cointegrating vector is allowed to change at a single unknown time during the sample period. While our null hypothesis (no cointegration) is the same, our alternative hypothesis is different than the conventional tests. Indeed, we extend the class of models under consideration, since our alternative hypothesis contains the Engle–Granger model as a special subcase.

The motivation for the class of tests considered here derives from the conventional notion of regime change. In some empirical exercises, a researcher may wish to entertain the possibility that the series are cointegrated, in the sense that a linear combination of the nonstationary variables is stationary, but that this linear combination (the cointegrating vector) has shifted at one unknown point in the sample. In this context, the standard tests for cointegration are not appropriate, since they presume that the cointegrating vector is time-invariant under the alternative hypothesis. A class of residual-based tests for cointegration are needed which include in the alternative hypothesis the models considered here.

Specifically, we propose extensions of the  $ADF$ ,  $Z_{\alpha}$ , and  $Z_t$  tests for cointegration. Our tests allow for a regime shift in either the intercept alone or the entire coefficient vector, and are noninformative with respect to the timing of the regime shift. This prevents informal data analysis (such as the visual examination of time series plots) from contaminating the choice of breakpoint. The tests of this paper can be viewed as multivariate extensions of the univariate tests of Perron (1989), Banerjee, Lumsdaine, and Stock (1992), Perron and Vogelsang (1992), and Zivot and Andrews (1992). These papers tested the null of a unit root in a univariate time series against the alternative of stationarity, while allowing for a structural break in the deterministic component of the series. In fact, some results of these papers can be viewed as special cases of our results, when the number of stochastic regressors is taken to be zero.

The asymptotic distributions of the test statistics are derived. We find that the asymptotic distributions of the proposed test statistics are free of nuisance parameter dependencies, other than the number of stochastic and deterministic regressors. The distributional theory is more involved than the theory for the

conventional cointegration model (see Phillips and Ouliaris, 1990) due to the inclusion of dummy variables and the explicit minimization over the set of possible breakpoints. It should be emphasized that our results are more general than the other results which have appeared in the literature on breaking trends and unit roots. Zivot and Andrews (1992) provided a rigorous proof for the simple Dickey–Fuller statistic in the univariate case, under the assumption of iid innovations. In contrast, we examine the more cumbersome Phillips  $Z_\alpha$  and  $Z_t$  tests in the multivariate case, while allowing for general forms of serial correlation in the innovations through the use of mixing conditions. (We do not, however, examine the most difficult test – the Augmented Dickey–Fuller statistic.)

Since there are no closed-formed solutions for the limiting distributions, critical values for up to four regressors are calculated for the tests by simulation methods. Also, since the computational requirements from recursive calculations are extremely high, preventing simulations with large sample sizes, we follow MacKinnon (1991) and estimate response surfaces to approximate the appropriate critical values. We evaluate the finite-sample performance of the tests using Monte Carlo methods based upon the experimental design of Engle and Granger (1987). In a limited set of experiments, we find the tests can detect cointegrating relations when there is a break in the intercept and/or slope coefficient. For these same experiments, the power of the conventional ADF test with no allowance for regime shifts falls sharply.

The tests of the present paper are clearly useful in helping lead an applied researcher to a correct model specification. Many researchers start a cointegration analysis with the usual augmented Dickey–Fuller (ADF) test, and proceed only if the statistic rejects the null of no cointegration. If the model is indeed cointegrated with a one-time regime shift in the cointegrating vector, the standard ADF test may not reject the null and the researcher will falsely conclude that there is no long-run relationship. Indeed, Gregory, Nason, and Watt (1994) have shown that the power of the conventional ADF test falls sharply in the presence of a structural break. In contrast, if the tests of the present paper are employed, there is a better chance of rejecting the null hypothesis, leading to a correct model formulation.

The tests of this paper are complementary to those of Hansen (1992a) and Quintos and Phillips (1993). These papers developed tests of the hypothesis of time invariance of the coefficients of a cointegrating relation. Their null hypothesis is Engle–Granger cointegration, while our null hypothesis is no cointegration. The tests of Hansen and Quintos–Phillips are best viewed as specification tests for the Engle–Granger cointegration model. In contrast, the tests of this paper are tests for cointegration, and are therefore best viewed as pre-tests akin to the conventional residual-based cointegration tests.

As an illustration of the techniques, we test for structural breaks in the U.S. long-run money-demand equation using annual and quarterly data. Our results

are consistent with those of Gregory and Nason (1992), who found some evidence of instability in the long-run relationship.

The organization of the paper is as follows. In Section 2 we develop several single-equation regression models which allow for cointegration with structural change. In Section 3 we describe various tests for the null of no cointegration with power against the structural change alternatives outlined in Section 2. Section 4 contains the asymptotic distribution theory for the tests. Critical values are calculated using simulation methods. Section 5 assesses the finite-sample properties of the structural change tests in a simple Monte Carlo experiment, and Section 6 examines an illustrative empirical example based upon money demand. Finally, in Section 7 we close with some concluding remarks.

## 2. Model

In this section we develop single-equation regression models which allow for cointegration with structural change. The observed data is  $y_t = (y_{1t}, y_{2t})$ , where  $y_{1t}$  is real-valued and  $y_{2t}$  is an  $m$ -vector. We commence with the standard model of cointegration with no structural change.

*Model 1: Standard cointegration*

$$y_{1t} = \mu + \alpha^T y_{2t} + e_t \quad t = 1, \dots, n, \quad (2.1)$$

where  $y_{2t}$  is  $I(1)$  and  $e_t$  is  $I(0)$ . In this model the parameters  $\mu$  and  $\alpha$  describe the  $m$ -dimensional hyperplane towards which the vector process  $y_t$  tends over time. Engle and Granger (1987) describe cointegration as a useful model for 'long-run equilibrium'.

In many cases, if model 1 is to capture a long-run relationship, we will want to consider  $\mu$  and  $\alpha$  as time-invariant. But in other applications, it may be desirable to think of cointegration as holding over some (fairly long) period of time, and then shifting to a new 'long-run' relationship. We treat the timing of this shift as unknown. The structural change would be reflected in changes in the intercept  $\mu$  and/or changes to the slope  $\alpha$ .

To model structural change, it is useful to define the dummy variable:

$$\varphi_{t\tau} = \begin{cases} 0 & \text{if } t \leq [n\tau], \\ 1 & \text{if } t > [n\tau], \end{cases}$$

where the unknown parameter  $\tau \in (0, 1)$  denotes the (relative) timing of the change point, and  $[ \ ]$  denotes integer part.

Structural change can take several forms, of which three are discussed here. A simple case is that there is a level shift in the cointegrating relationship, which

can be modeled as a change in the intercept  $\mu$ , while the slope coefficients  $\alpha$  are held constant. This implies that the equilibrium equation has shifted in a parallel fashion. We call this a *level shift* model denoted by C.

*Model 2: Level shift (C)*

$$y_{1t} = \mu_1 + \mu_2 \varphi_{t\tau} + \alpha^\top y_{2t} + e_t, \quad t = 1, \dots, n. \quad (2.2)$$

In this parameterization  $\mu_1$  represents the intercept before the shift, and  $\mu_2$  represents the change in the intercept at the time of the shift. We can also introduce a time trend into the level shift model.

*Model 3: Level shift with trend (C/T)*

$$y_{1t} = \mu_1 + \mu_2 \varphi_{t\tau} + \beta t + \alpha^\top y_{2t} + e_t, \quad t = 1, \dots, n. \quad (2.2)$$

Another possible structural change allows the slope vector to shift as well. This permits the equilibrium relation to rotate as well as shift parallel. We call this the *regime shift* model.

*Model 4: Regime shift (C/S)*

$$y_{1t} = \mu_1 + \mu_2 \varphi_{t\tau} + \alpha_1^\top y_{2t} + \alpha_2^\top y_{2t} \varphi_{t\tau} + e_t, \quad t = 1, \dots, n. \quad (2.3)$$

In this case  $\mu_1$  and  $\mu_2$  are as in the level shift model,  $\alpha_1$  denotes the cointegrating slope coefficients before the regime shift, and  $\alpha_2$  denotes the change in the slope coefficients.

The standard methods to test the null hypothesis of no cointegration (derived in the context of model 1) are residual-based. The candidate cointegrating relation is estimated by ordinary least squares (OLS), and a unit root test is applied to the regression errors. In principle the same approach could be used for testing models 2–4, if the timing of the regime shift  $\tau$  were known *a priori*. We take the view that such break points are unlikely to be known in practice without some appeal to the data. Indeed much of the debate about whether there was a regime shift in U.S. GNP around 1929 (as identified in Perron, 1989) can only be resolved conditional on the data (see Banerjee, Lumsdaine, and Stock, 1992; Christiano, 1992; Zivot and Andrews, 1992). Similar problems occur in testing for regime shifts in cointegrated models and so we develop tests procedures that do not require information regarding the timing of or indeed the occurrence of a break.

This completes the description of the structural change models under cointegration. In the next section we analyze some tests designed to detect cointegration in the possible presence of such breaks.

### 3. Testing the null of no cointegration

Hansen (1992a) constructed tests of model 1 against the alternative of model 4. A statistically significant test statistic in this context would be taken as evidence against the standard cointegration model in favor of the regime shift model. However, before calculating such a test, an applied econometrician might wish to apply a conventional test for cointegration, such as the ADF test, in the context of model 1. If the true process is represented by model 4, and not model 1, the distributional theory used to assess the significance of the ADF test statistic is not same. In this section we develop such tests of the null of no cointegration against the alternatives in models 2–4.

Define the innovation vector  $u_t = \Delta y_t$ , its cumulative process  $S_t = \sum_{i=1}^t u_i$  (so  $y_t = y_0 + S_t$ ), and its long-run variance  $\Omega = \lim_n n^{-1} E S_n S_n^T$ . When  $u_t$  is covariance-stationary,  $\Omega$  is proportional to the spectral density matrix evaluated at the zero frequency.

Our null hypothesis is that model 1 holds with  $e_t \equiv I(1)$ . This implies that  $\Omega > 0$ . We include this aspect of the null hypothesis in the following regularity conditions:

*Assumptions:*

- (a)  $\{u_t\}$  is mean-zero and strong mixing with mixing coefficients of size  $-p\beta/(p - \beta)$ , and  $E|u_t|^p < \infty$  for some  $p > \beta > \frac{5}{2}$ .
- (b) The matrix  $\Omega$  exists with finite elements and  $\Omega > 0$ .
- (c)  $y_0$  is a random vector with  $E|y_0| < \infty$ .

The solution we adopt to handling regime shifts is similar to that of Banerjee, Lumsdaine, and Stock (1992) and Zivot and Andrews (1992). We compute the cointegration test statistic for each possible regime shift  $\tau \in T$ , and take the smallest value (the largest negative value) across all possible break points. In principle the set  $T$  can be any compact subset of  $(0, 1)$ . In practice, it will need to be small enough so that all of the statistics discussed here can be calculated. For example  $T = (0.15, 0.85)$  seems a reasonable suggestion, following the earlier literature. Although  $T$  contains an uncountable number of points, all the statistics that we consider are step functions on  $T$ , taking jumps only on the points  $\{i/n, i \text{ integer}\}$ . For computational purposes, the test statistic is computed for each break point in the interval  $([0.15n], [0.85n])$ .

We now describe the computation of the test statistics. For each  $\tau$ , estimate one of the models 2–4 (depending upon the alternative hypothesis under consideration) by OLS, yielding the residual  $\hat{e}_{\tau}$ . The subscript  $\tau$  on the residuals denotes the fact that the residual sequence depends on the choice of change point  $\tau$ . From these residuals, calculate the first-order serial

correlation coefficient

$$\hat{\rho}_\tau = \frac{\sum_{t=1}^{n-1} \hat{e}_{t\tau} \hat{e}_{t+1\tau}}{\sum_{t=1}^{n-1} \hat{e}_{t\tau}^2}.$$

The Phillips (1987) test statistics are formed using a bias-corrected version of the first-order serial correlation coefficient. Define the second-stage residuals

$$\hat{v}_{t\tau} = \hat{e}_{t\tau} - \hat{\rho}_\tau \hat{e}_{t-1\tau}.$$

The correction involves the following estimate of a weighted sum of autocovariances:

$$\hat{\lambda}_\tau = \sum_{j=1}^M w\left(\frac{j}{M}\right) \hat{\gamma}_\tau(j),$$

where

$$\hat{\gamma}_\tau(j) = \frac{1}{n} \sum_{t=j+1}^n \hat{v}_{t-j\tau} \hat{v}_{t\tau},$$

and  $M = M(n)$  is the bandwidth number selected so that  $M \rightarrow \infty$  and  $M/n^5 = O(1)$ . This latter requirement may be stronger than necessary, but simplifies our proof, and should not be taken as a serious implication as a practical selection guideline. We allow the kernel weights  $w(\cdot)$  to satisfy the standard conditions for spectral density estimators. The estimate of the long-run variance of  $\hat{v}_{t\tau}$  is

$$\hat{\sigma}_\tau^2 = \hat{\gamma}_\tau(0) + 2\hat{\lambda}_\tau.$$

In this paper the long-run variance is estimated using a prewhitened quadratic spectral kernel with a first-order autoregression for the prewhitening and an automatic bandwidth estimator (see Andrews, 1991; Andrews and Monahan, 1992; for details).

The bias-corrected first-order serial correlation coefficient estimate is given by

$$\hat{\rho}_\tau^* = \frac{\sum_{t=1}^{n-1} (\hat{e}_{t\tau} \hat{e}_{t+1\tau} - \hat{\lambda}_\tau)}{\sum_{t=1}^{n-1} \hat{e}_{t\tau}^2}.$$

The Phillips test statistics can be written as

$$Z_\alpha(\tau) = n(\hat{\rho}_\tau^* - 1),$$

$$Z_t(\tau) = (\hat{\rho}_\tau^* - 1)/\hat{s}_\tau, \quad \hat{s}_\tau^2 = \hat{\sigma}_\tau^2 / \sum_1^{n-1} \hat{e}_{t\tau}^2.$$

The final statistic we discuss is the augmented Dickey–Fuller (ADF) statistic. This is calculated by regressing  $\Delta \hat{e}_{t\tau}$  upon  $\hat{e}_{t-1\tau}$  and  $\Delta \hat{e}_{t-1\tau}, \dots, \Delta \hat{e}_{t-K\tau}$  for some

suitably chosen lag truncation  $K$ . The ADF statistic is the  $t$ -statistic for the regressor  $\hat{e}_{t-1\tau}$ . We denote this by

$$ADF(\tau) = tstat(\hat{e}_{t-1\tau}).$$

These test statistics are now standard tools for the analysis of cointegrating regressions without regime shifts. Our statistics of interest, however, are the *smallest* values of the above statistics, across all values of  $\tau \in T$ . We examine the smallest values since small values of the test statistics constitute evidence against the null hypothesis. These test statistics are

$$Z_{\alpha}^* = \inf_{\tau \in T} Z_{\alpha}(\tau), \quad (3.1)$$

$$Z_t^* = \inf_{\tau \in T} Z_t(\tau), \quad (3.2)$$

$$ADF^* = \inf_{\tau \in T} ADF(\tau). \quad (3.3)$$

#### 4. Asymptotic distributions

We follow much of the recent literature and give asymptotic distributions for the test statistics which are expressed as functionals of Brownian motions. This gives simple expressions for the limit distributions. Since they are not given in closed-form, however, we use simulation methods to obtain critical values.

The theorem we give below applies to the models discussed in Section 2, although our formal proof is for a specialized case to minimize space. We also examine formally only the  $Z_{\alpha}^*$  and  $Z_t^*$  tests, since these are more straightforward to analyze. We expect that the limiting distribution of  $ADF^*$  is identical to  $Z_t^*$ .

The proof of the theorem is contained in the Appendix. Some comments on the proof method are warranted. If we were fixing  $\tau$  a priori, then the proof would be a fairly straightforward extension of Phillips and Ouliaris (1990). Our statistics, however, are a function of every pointwise test statistic, considered as a function of  $\tau$ . This requires the distributional results to hold uniformly over  $\tau$ . This difficulty was rigorously addressed by Zivot and Andrews (1992) in the context of testing for an autoregressive unit root. They handled the problem by considering the test statistics as functions of the indicator function  $\varphi_{t\tau}$ . Since the latter is discontinuous (even asymptotically), this required that they avoid the uniform metric, which complicated their analysis. Their approach also made the handling of serial correlation (and the appropriate corrections) more difficult, so their proof only dealt with iid error terms. As an alternative to the Zivot–Andrews proof method, we construct our proof in such a way as to avoid reference to the discontinuous function  $\varphi_{t\tau}$ , allowing the use of the uniform



metric, and allowing for general forms of serial correlation (and the appropriate corrections).

*Theorem.* Under the null hypothesis,

$$Z_{\alpha}^* \longrightarrow_d \inf_{\tau \in T} \int_0^1 W_{\tau} dW_{\tau} / \int_0^1 W_{\tau}^2$$

and

$$Z_{\tau}^* \longrightarrow_d \inf_{\tau \in T} \int_0^1 W_{\tau} dW_{\tau} / \left[ \int_0^1 W_{\tau}^2 \right]^{1/2} [1 + \kappa_{\tau}^{\top} D_{\tau} \kappa_{\tau}]^{1/2},$$

where

$$W_{\tau}(r) = W_1(r) - \int_0^1 W_1 W_{2\tau}^{\top} \left[ \int_0^1 W_{2\tau} W_{2\tau}^{\top} \right]^{-1} W_{2\tau}(r),$$

$$\kappa_{\tau} = \left[ \int_0^1 W_{2\tau} W_{2\tau}^{\top} \right]^{-1} \int_0^1 W_{2\tau} W_1,$$

and  $W_{2\tau}(r)$  and  $D_{\tau}$  depend on the model. If the residuals are from OLS estimation of model 2, then

$$W_{2\tau}(r) = [1, \varphi_{\tau}(r), W_2^{\top}(r),]^{\top},$$

where  $\varphi_{\tau}(r) = \{r \geq \tau\}$  (the braces  $\{ \}$  refer to the indicator function), and

$$D_{\tau} = \begin{pmatrix} 0 & 0 \\ 0 & I_m \end{pmatrix}.$$

If the residuals are from OLS estimation of model 3, then

$$W_{2\tau}(r) = [1, r, \varphi_{\tau}(r), W_2^{\top}(r),]^{\top}$$

and

$$D_{\tau} = \begin{pmatrix} 0 & 0 \\ 0 & I_m \end{pmatrix}.$$

If the residuals are from OLS estimation of model 4, then

$$W_{2\tau}(r) = [1, \varphi_{\tau}(r), W_2^{\top}(r), W_2^{\top}(r) \varphi_{\tau}(r)]^{\top}$$

and

$$D_{\tau} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_m & (1 - \tau)I_m \\ 0 & (1 - \tau)I_m & (1 - \tau)I_m \end{pmatrix}.$$

### Calculating critical values

One standard way in which critical values have been obtained in situations where no closed-form expressions exist is to simulate a test statistic for a large sample size for a large number of replications. In the present example, we are unable to use large sample sizes since the recursive calculations which are required over the sample are particularly time-consuming on even fast computers. For instance, with  $n = 300$  and 10,000 replications it took a GATEWAY2000 486/33C over a week to do the relevant calculations for the one regressor case ( $m = 1$ ). To reduce the computational requirements we adopt a procedure due to MacKinnon (1991). Using 10,000 replications for each of the sample sizes  $n = 50, 100, 150, 200, 250,$  and  $300$  we obtain critical values,  $Crt(n, p, m)$ , where  $p$  is the percent quantile and  $m$  is the number of regressors in the equation (excluding a constant and/or trend). We then estimate by ordinary least squares for each  $p$  and  $m$  the response surface

$$Crt(n, p, m) = \psi_0 + \psi_1 n^{-1} + error.$$

Various other functional relations were tried (involving  $n^{-1/2}, n^{-2}, n^{-3/2}$ ) but this one appeared to have the best fit ( $R^2$ 's were generally over 0.98). The asymptotic critical value is taken to be the OLS estimate  $\hat{\psi}_0$ . While this specification of the critical values seemed best overall, estimates of  $\psi_0$  changed on occasion by a factor of 2 under alternative specifications. Results for  $p = 0.01, 0.025, 0.05, 0.10,$  and  $0.975$  and  $m = 1, 2, 3, 4$  are presented in Table 1. The symbols  $C, C/T,$  and  $C/S$  refer to breaks in the constant ( $C$ ) and slope coefficient ( $S$ ) as defined in models 2–4.

### 5. A simple Monte Carlo experiment

In order to gauge the finite-sample properties of the proposed test, we conduct a simple Monte Carlo experiment based upon the design of Granger and Engle (1987) and also used in Banerjee, Dolado, Hendry, and Smith (1986). The model in the absence of structural change is

$$y_{1t} = 1 + 2y_{2t} + \varepsilon_t, \quad \varepsilon_t = \rho\varepsilon_{t-1} + \vartheta_t, \quad \vartheta_t \sim \text{NID}(0, 1),$$

$$y_{2t} = y_{2t} + \eta_t, \quad \eta_t = \eta_{t-1} + \omega_t, \quad \omega_t \sim \text{NID}(0, 1),$$

where  $y_{2t}$  is scalar ( $m = 1$ ). We first consider the size of the various tests with  $\rho = 1$ , so that the null of no cointegration is true. In Table 2 we report the rejection frequencies in 2500 replications at the 5% level of significance (we also did 1% and 10%, and these are available upon request) using critical values from Table 1. Two sample sizes ( $n = 50$  and  $100$ ) are considered.  $Z_\alpha^*, Z_t^*$ , and  $ADF^*$  are the test statistics defined in Eqs. (3.1)–(3.3) respectively, and  $ADF$  is

Table 1  
Approximate asymptotic critical values

	Level	0.01	0.025	0.05	0.10	0.975
$m = 1$	$ADF^*, Z_t^*$					
	C	-5.13	-4.83	-4.61	-4.34	-2.25
	C/T	-5.45	-5.21	-4.99	-4.72	-2.72
	C/S	-5.47	-5.19	-4.95	-4.68	-2.55
	$Z_x^*$					
	C	-50.07	-45.01	-40.48	-36.19	-10.63
	C/T	-57.28	-52.09	-47.96	-43.22	-15.90
	C/S	-57.17	-51.32	-47.04	-41.85	-13.15
	$m = 2$	$ADF^*, Z_t^*$				
C		-5.44	-5.16	-4.92	-4.69	-2.61
C/T		-5.80	-5.51	-5.29	-5.03	-3.01
C/S		-5.97	-5.73	-5.50	-5.23	-3.12
$Z_x^*$						
C		-57.01	-51.41	-46.98	-42.49	-14.27
C/T		-64.77	-58.57	-53.92	-48.94	-19.19
C/S		-68.21	-63.28	-58.33	-52.85	-19.72
$m = 3$		$ADF^*, Z_t^*$				
	C	-5.77	-5.50	-5.28	-5.02	-2.96
	C/T	-6.05	-5.79	-5.57	-5.33	-3.33
	C/S	-6.51	-6.23	-6.00	-5.75	-3.65
	$Z_x^*$					
	C	-63.64	-57.96	-53.58	-48.65	-18.20
	C/T	-70.27	-64.26	-59.76	-54.94	-22.72
	C/S	-80.15	-73.91	-68.94	-63.42	-26.64
	$m = 4$	$ADF^*, Z_t^*$				
C		-6.05	-5.80	-5.56	-5.31	-3.26
C/T		-6.36	-6.07	-5.83	-5.59	-3.59
C/S		-6.92	-6.64	-6.41	-6.17	-4.12
$Z_x^*$						
C		-70.18	-64.41	-59.40	-54.38	-22.04
C/T		-76.95	-70.56	-65.44	-60.12	-26.46
C/S		-90.35	-84.00	-78.52	-72.56	-33.69

These critical values are based on the response surface

$$Crt = \psi_0 + \psi_1 n^{-1} + error,$$

where  $Crt$  is the critical value obtained from 10,000 replications at sample size  $n = 50, 100, 150, 200, 250, 300$ . The asymptotic critical value is the ordinary least squares estimate (OLS) of  $\psi_0$ .  $Z_x^*$ ,  $Z_t^*$ , and  $ADF^*$  are the test statistics defined in Eqs. (3.1)–(3.3), respectively. The symbols C, C/T, and C/S refer to models 2–4, respectively.

the usual augmented Dickey–Fuller statistic ( $\tau = 1$ ). The symbols  $C$  and  $C/T$  for the usual  $ADF$  refer to regressions with a constant and a constant and a trend. For  $ADF^*$  and  $ADF$  the lag length  $K$  is selected on the basis of a  $t$ -test following a procedure similar to Perron and Vogelsang (1992). We set  $K_{\max}$  to 6 and then test downward (reducing  $K$ ) until the last lag of the first difference included is significant at the 5% level using normal critical values. Typically over all the experiments including those with regime shifts,  $K$  was 0 or 1 but on occasion this procedure led to lags of 6 being chosen.

From the size experiments ( $\rho = 1$ ), we see that the  $ADF^*$ ,  $Z_t^*$ , and the standard  $ADF$  tests are all biased away from the null. On the other hand,  $Z_\alpha$  is biased towards the null, particularly at sample size  $n = 50$ . Also the size distortion is larger for the  $ADF^*$  compared to the standard  $ADF$ . In light of the different size properties of the tests, in the remaining experiments where the variables are cointegrated (with possible breaks), we calculate both nominal and size-adjusted (in parentheses in the tables) power. The size adjustments are made on the basis of the  $\rho = 1$  results of Table 2.

Turning to power in the usual cointegration setting (model 1) we let  $\rho = 0$  and  $\rho = 0.5$  (Table 2). With  $\rho = 0$ , the  $Z_t^*$  and  $Z_\alpha^*$  have highest size-adjusted power with the standard  $ADF$  and the  $ADF^*$  being roughly comparable and much lower. Typically, the addition of the dummy variables lowered the lag length chosen in these tests. Clearly from the perspective of power, faulty inclusion (i.e., including unnecessary regressors to capture breaks that do not exist) is not a problem. As we would expect, all tests have lower power at sample size  $n = 50$  when the error in the cointegrating regression is serially correlated (see Gregory, 1994). Interestingly with  $\rho = 0.5$ , the power (both nominal and size-adjusted) are close for all the tests with the exception of the much smaller nominal power for  $Z_\alpha$  at  $n = 50$ . There is also the tendency with more serial correlation for the relative power decline to be larger at the smaller sample size for the structural break tests for cointegration than the standard  $ADF$  tests.

We postulate a simple structural break for the intercept and then just the slope (Table 3),

$$\begin{aligned}
 y_{1t} &= \mu_t + \alpha_t y_{2t} + \varepsilon_t, \\
 \mu_t &= \mu_1, \quad \alpha_t = \alpha_1 \quad \text{if } t \leq [\tau n], \\
 &= \mu_2, \quad \alpha_t = \alpha_2 \quad \text{if } t > [\tau n], \\
 \varepsilon_t &= 0.5\varepsilon_{t-1} + \vartheta_t, \\
 y_{1t} &= y_{2t} + \eta_t, \quad \eta_t = \eta_{t-1} + \omega_t.
 \end{aligned}$$

The two errors  $\vartheta_t$  and  $\omega_t$  are uncorrelated and distributed as  $NID(0,1)$ . Since in applied work the errors are likely to be serially correlated, we make the break experiments similar in structure to those in Table 2 with  $\rho = 0.5$ . A break point occurs at  $\tau = 0.25, 0.5$ , and  $0.75$ .

Table 2  
Testing for cointegration with no regime shifts

$$y_{1t} = 1 + 2y_{2t} + \varepsilon_t, \quad \varepsilon_t = \rho\varepsilon_{t-1} + \vartheta_t, \quad \vartheta_t \sim \text{NID}(0, 1)$$

$$y_{1t} = y_{2t} + \eta_t, \quad \eta_t = \eta_{t-1} + \omega_t, \quad \omega_t \sim \text{NID}(0, 1)$$

	$\rho = 1.0$		$\rho = 0$		$\rho = 0.5$	
	$n = 50$	$n = 100$	$n = 50$	$n = 100$	$n = 50$	$n = 100$
<i>ADF*</i>						
C	0.17	0.13	0.85 (0.77)	0.86 (0.82)	0.49 (0.24)	0.89 (0.81)
C/T	0.16	0.13	0.82 (0.64)	0.86 (0.82)	0.40 (0.17)	0.87 (0.72)
C/S	0.13	0.10	0.87 (0.79)	0.89 (0.86)	0.40 (0.22)	0.88 (0.82)
<i>Z<sub>t</sub>*</i>						
C	0.11	0.09	0.99 (0.99)	1.0 (1.0)	0.58 (0.37)	0.98 (0.95)
C/T	0.14	0.12	0.98 (0.92)	1.0 (1.0)	0.52 (0.26)	0.96 (0.88)
C/S	0.11	0.09	0.99 (0.97)	1.0 (1.0)	0.50 (0.32)	0.97 (0.93)
<i>Z<sub>x</sub>*</i>						
C	0.01	0.04	0.82 (0.96)	1.0 (1.0)	0.11 (0.36)	0.93 (0.94)
C/T	0.00	0.03	0.39 (0.89)	1.0 (1.0)	0.02 (0.27)	0.79 (0.85)
C/S	0.00	0.03	0.57 (0.95)	1.0 (1.0)	0.03 (0.33)	0.84 (0.92)
<i>ADF</i>						
C	0.08	0.07	0.64 (0.61)	0.81 (0.78)	0.56 (0.45)	0.84 (0.82)
C/T	0.11	0.08	0.60 (0.54)	0.71 (0.66)	0.38 (0.21)	0.78 (0.72)

Rejection frequencies at the 5% level of significance using critical values from Table 1. In parentheses are the size-adjusted rejection frequencies based on critical values from this table with  $\rho = 1.0$ . There are 2500 replications for each experiment.  $Z_x^*$ ,  $Z_t^*$ , and  $ADF^*$  are the test statistics defined in Eqs. (3.1)–(3.3), respectively, and  $ADF$  is the usual augmented Dickey–Fuller statistic with  $\tau = 1$ . C, C/T, and C/S refer to models 2–4, respectively, and C and C/T for the usual  $ADF$  refer to regressions with a constant and a constant and a trend. For  $ADF^*$  and  $ADF$  the lag length  $K$  is set on the basis of a  $t$ -test (see text).

In Table 3 we first investigate the ability of the tests to detect cointegration in the presence of a level shift with  $\mu_1 = 1$ ,  $\mu_2 = 4$ ,  $\alpha_1 = 2$ , and  $\alpha_2 = 2$ . With only an intercept change we find one prominent difference between the rejection frequencies in Table 2 (with no structural break) and Table 3 (where the intercept shifts): the rejection frequencies for the conventional  $ADF$  test for  $n = 100$  have fallen substantially under breaks. For instance at  $n = 100$  the  $ADF$  test that includes an intercept (C) has fallen from about 80% rejection frequency to roughly 30%, regardless of where the break occurs. In contrast, the rejection frequencies for all three of our cointegration tests that allow for possible shifts are very close to those obtained with time invariant relations. As we might expect the best power results are generally for the test based on the level shift

Table 3  
Structural breaks in the intercept and slope

$$y_{1t} = \mu_t + \alpha_1 y_{2t} + \varepsilon_t, \quad \begin{cases} \mu_t = \mu_1, & \alpha_1 = \alpha_1, & t \leq [\tau T] \\ \mu_t = \mu_2, & \alpha_1 = \alpha_2, & t > [\tau T] \end{cases}$$

$$\varepsilon_t = 0.5\varepsilon_{t-1} + \theta_t, \quad \theta_t \sim \text{NID}(0, 1)$$

$$y_{1t} = y_{2t} + \eta_t, \quad \eta_t = \eta_{t-1} + \omega_t, \quad \omega_t \sim \text{NID}(0, 1)$$

		$\mu_1 = 1, \mu_2 = 4, \alpha_1 = 2, \alpha_2 = 2$				$\mu_1 = 1, \mu_2 = 1, \alpha_1 = 2, \alpha_2 = 4$							
		$n = 100$				$n = 50$				$n = 100$			
		0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
<b>ADF*</b>													
C		0.44 (0.19)	0.48 (0.23)	0.50 (0.23)	0.91 (0.82)	0.92 (0.82)	0.92 (0.85)	0.22 (0.07)	0.31 (0.11)	0.41 (0.17)	0.35 (0.21)	0.51 (0.33)	0.77 (0.63)
C/T		0.39 (0.15)	0.42 (0.17)	0.46 (0.20)	0.85 (0.68)	0.87 (0.74)	0.89 (0.76)	0.21 (0.04)	0.28 (0.07)	0.40 (0.11)	0.33 (0.16)	0.49 (0.30)	0.72 (0.55)
C/S		0.34 (0.18)	0.36 (0.18)	0.38 (0.20)	0.85 (0.76)	0.86 (0.77)	0.88 (0.81)	0.36 (0.16)	0.37 (0.18)	0.36 (0.17)	0.91 (0.82)	0.88 (0.79)	0.85 (0.76)
<b>Z<sub>t</sub>*</b>													
C		0.55 (0.34)	0.56 (0.36)	0.58 (0.38)	0.98 (0.95)	0.98 (0.94)	0.98 (0.94)	0.31 (0.13)	0.35 (0.13)	0.49 (0.21)	0.53 (0.40)	0.63 (0.49)	0.91 (0.81)
C/T		0.50 (0.25)	0.52 (0.29)	0.58 (0.34)	0.94 (0.86)	0.95 (0.88)	0.96 (0.90)	0.31 (0.05)	0.37 (0.08)	0.51 (0.12)	0.49 (0.33)	0.62 (0.47)	0.86 (0.74)
C/S		0.46 (0.29)	0.45 (0.30)	0.48 (0.32)	0.95 (0.89)	0.95 (0.89)	0.96 (0.92)	0.47 (0.27)	0.46 (0.25)	0.46 (0.25)	0.97 (0.93)	0.96 (0.91)	0.96 (0.91)
<b>Z<sub>t</sub><sup>#</sup></b>													
C		0.10 (0.35)	0.11 (0.34)	0.12 (0.37)	0.92 (0.93)	0.92 (0.93)	0.92 (0.94)	0.04 (0.15)	0.05 (0.20)	0.09 (0.30)	0.35 (0.37)	0.45 (0.49)	0.80 (0.83)
C/T		0.02 (0.26)	0.02 (0.29)	0.03 (0.34)	0.75 (0.82)	0.79 (0.85)	0.83 (0.87)	0.01 (0.12)	0.01 (0.16)	0.03 (0.25)	0.23 (0.29)	0.36 (0.44)	0.63 (0.71)
C/S		0.02 (0.31)	0.02 (0.30)	0.02 (0.32)	0.79 (0.88)	0.80 (0.88)	0.83 (0.91)	0.03 (0.33)	0.03 (0.30)	0.02 (0.31)	0.84 (0.92)	0.81 (0.89)	0.80 (0.89)
<b>ADF</b>													
C		0.46 (0.35)	0.15 (0.11)	0.14 (0.09)	0.33 (0.29)	0.29 (0.26)	0.29 (0.26)	0.13 (0.03)	0.17 (0.04)	0.33 (0.10)	0.20 (0.17)	0.24 (0.20)	0.47 (0.43)
C/T		0.36 (0.19)	0.25 (0.12)	0.30 (0.15)	0.43 (0.36)	0.58 (0.50)	0.60 (0.53)	0.13 (0.03)	0.19 (0.06)	0.34 (0.11)	0.19 (0.14)	0.28 (0.21)	0.58 (0.50)

Rejection frequencies at the 5% level of significance using critical values from Table 1 in 2500 replications. In parentheses are the size-adjusted rejection frequencies based on critical values from Table 2 with  $\zeta = 1.0$ . See Table 2 for definitions of symbols.

model 2 (C). Also power appears to be unaffected by the location of the break in the sample.

In the same table we report results when the slope changes but the intercept is fixed over the sample:  $\mu_1 = 1$ ,  $\mu_2 = 1$ ,  $\alpha_1 = 2$ , and  $\alpha_2 = 4$ . As before the best power results are obtained for those tests that allow for a shift in the slope (C/S). Compared to the tests with no structural break (Table 2), the rejection frequencies (both nominal and size-adjusted) for the (C) and (C/T) have fallen considerably when  $\tau = 0.25$  or 0.5. For this experiment, power rises as the breakpoint occurs later in the sample. In a series of experiments we increased  $\sigma_\beta$  relative to  $\sigma_\omega$ . As might be expected power falls as  $\sigma_\beta$  rises relative to  $\sigma_\omega$ . Finally we note that the standard ADF tests reject the null far less frequently compared to tests which account for breaks especially if the slope break is in the earlier part of the sample.

In Table 4 we present some results on the average estimated  $\tau$ 's [the point in the sample where the smallest test statistic is obtained for each test (3.1)–(3.3)] together with their standard error based upon the simulations in Table 3 with the slope change. Consistent with our earlier results we see that the tests C/S estimate the breakpoint more accurately with a smaller standard deviation. The evidence also suggests that breaks that occur in the latter half of the sample are better estimated.

Table 4  
Estimating a breakpoint

$\tau$	$n = 50$			$n = 100$		
	0.25	0.50	0.75	0.25	0.50	0.75
<i>ADF*</i>						
C	0.58 (0.20)	0.66 (0.18)	0.67 (0.21)	0.53 (0.22)	0.66 (0.18)	0.71 (0.20)
C/T	0.56 (0.21)	0.61 (0.20)	0.61 (0.22)	0.53 (0.21)	0.62 (0.19)	0.65 (0.22)
C/S	0.45 (0.18)	0.58 (0.15)	0.66 (0.19)	0.33 (0.13)	0.56 (0.11)	0.70 (0.17)
<i>Z<sub>t</sub>*</i>						
C	0.51 (0.24)	0.64 (0.20)	0.66 (0.23)	0.45 (0.24)	0.64 (0.19)	0.70 (0.21)
C/T	0.54 (0.22)	0.59 (0.21)	0.62 (0.23)	0.50 (0.22)	0.62 (0.20)	0.65 (0.23)
C/S	0.42 (0.18)	0.57 (0.15)	0.64 (0.20)	0.31 (0.12)	0.55 (0.10)	0.70 (0.17)
<i>Z<sub>x</sub>*</i>						
C	0.51 (0.23)	0.63 (0.20)	0.65 (0.24)	0.45 (0.23)	0.64 (0.19)	0.70 (0.21)
C/T	0.54 (0.22)	0.60 (0.21)	0.61 (0.23)	0.50 (0.22)	0.62 (0.19)	0.65 (0.23)
C/S	0.42 (0.18)	0.57 (0.15)	0.62 (0.21)	0.31 (0.12)	0.55 (0.10)	0.70 (0.17)

These are the average estimated  $\tau$ 's which is the point in the sample where the smallest value of the test statistic is obtained. Standard error are in parentheses. The results are for the experiment described in Table 3 for the slope change  $\alpha_1 = 2$  and  $\alpha_2 = 4$ .

Overall the results with  $Z_t^*$  appear to be the best in terms of size and power. However, as we have seen this test is biased away from the null which can lead to large gaps between nominal and size-adjusted power.

## 6. An illustrative example

The question of whether the long-run money-demand equation is stable has received ample attention over the years. Two recent discussions are Lucas (1988) and Stock and Watson (1993). As a practical example, we consider this question in the context of our new tests using annual data (1901–1985) and quarterly data (1960:1–1990:4). The long-run money-demand relation with no structural breaks may be written as

$$\ln(m_t) - \ln(p_t) = \mu + \alpha_1 \ln(y_t) + \alpha_2 r_t + e_t.$$

The annual data are from Lucas (1988) and  $m$  is M1,  $p$  is the implicit price deflator,  $y$  is the real net national product, and  $r$  is the six-month commercial paper rate. The quarterly data are from the CITIBASE tape; the series used are GMPY, FXGM3, GMPY82, FM1, GNNP, GNNP82, and are seasonally adjusted. This specification is identical to Lucas (1988) and Stock and Watson (1991).

With this same data, Gregory, Nason, and Watt (1994) found that some of Hansen's (1992c) tests for structural breaks detected a break and that the conventional *ADF* tests indicated that the null of no cointegration could be rejected for the annual data but not for the quarterly. In Table 5 we report the test statistics for our tests as well as the estimated breakpoint (in parentheses). In addition, we calculate the test statistic for the conventional *ADF* test. For the conventional *ADF* test the lag length for  $K$  (again selected on the basis of a  $t$ -test as outlined in the Section 5) is 1 (annual) and 6 (quarterly) for both the C and C/T tests. The lags selected for *ADF\** tests for the C, C/T, and C/S are (2,2,0) for the annual and (0,0,6) for the quarterly data, respectively.

Examining first the annual data, we find that the null hypothesis of no cointegration is rejected (at the 5% level) by our new tests using the C and C/T type formulations, but not using the C/S formulation. The smallest test statistic occurred roughly at half of the sample (1944), and in Fig. 1 we graph the *ADF*( $\tau$ ) using the annual data for C, C/T, and C/S over the truncated sample. Clearly there is a well-defined single minimum for all three of these tests. Since the conventional *ADF* test rejects the same null, it would be inappropriate to conclude from this piece of information alone that there is a structural break, since a conventional cointegrated system could produce this same set of results. In fact, the point estimates for the second half of the sample (not reported here) are economically meaningless, lending support to the view that there is no structural break. On balance we believe that the evidence suggests



Table 5  
Testing for regime shifts in U.S. money demand

$$\ln(m_t) - \ln(p_t) = \mu + \alpha_1 \ln(y_t) + \alpha_2 r_t + e_t$$

	Annual data, 1901–1985		Quarter data, 1960: 1–1990: 4	
	Test stat.	Breakpoint	Test stat.	Breakpoint
<i>ADF*</i>				
C	– 5.43**	(0.49)	– 3.71	(0.44)
C/T	– 5.66**	(0.50)	– 4.34	(0.84)
C/S	– 5.01	(0.49)	– 4.71	(0.55)
<i>Z<sub>t</sub>*</i>				
C	– 5.33**	(0.50)	– 3.57	(0.39)
C/T	– 5.85**	(0.49)	– 4.66	(0.85)
C/S	– 5.38*	(0.49)	– 5.82**	(0.52)
<i>Z<sub>α</sub>*</i>				
C	– 46.39*	(0.50)	– 22.94	(0.85)
C/T	– 52.52*	(0.49)	– 36.38	(0.85)
C/S	– 44.36	(0.50)	– 53.57*	(0.52)
<i>ADF</i>				
C	– 4.55**	(–)	– 2.91	(–)
C/T	– 4.37**	(–)	– 2.71	(–)

These are the test statistics where an \* and \*\* indicates significance at the 10% and 5% levels, respectively. Beside these in parentheses are the estimated  $\tau$ 's (the point in the sample where the smallest value of the test statistic is obtained). The annual data are from Lucas (1988) and  $m$  is M1,  $p$  is the implicit price deflator,  $y$  is the real net national product, and  $r$  is the six-month commercial paper rate. The quarterly data are from the CITIBASE tape; the series used are GMPY, FXGM3, GMPY82, FM1, GNNP, GNNP82, and are seasonally adjusted.

that there is some sort of long-run cointegrating relationship between these variables.

Turning to the quarterly data, we first notice that the conventional *ADF* tests fail to reject the null of no cointegration. Thus some applied researchers might conclude that there is not sufficient evidence to pursue the possibility of a long-run relationship. The tests which allow for only a level shift (the C and C/T tests) also fail to reject the null. We find, however, that the null is rejected at the 5% level for the  $Z_t^*$  and 10% for  $Z_\alpha^*$  tests by the most general alternative (the C/S tests) which allows for both the intercept and the slope coefficient to shift. The same test for *ADF\** does not reject at this level. For this data set and model, allowing for the possibility of a regime shift does raise important questions

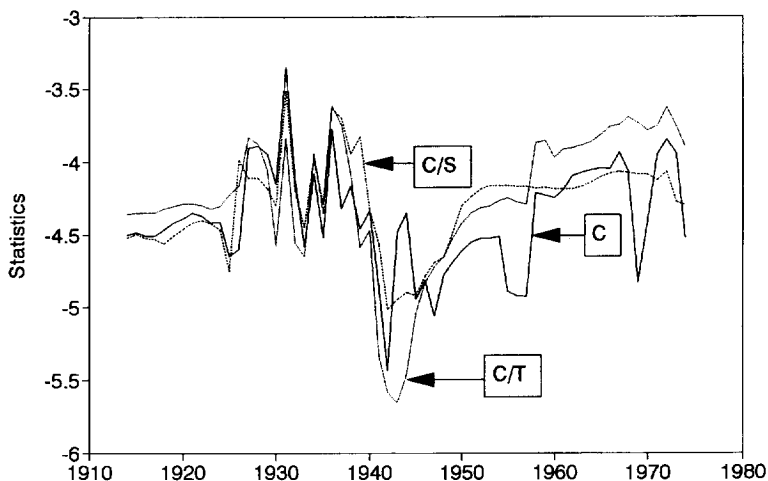


Fig. 1. Regime shifts with  $ADF^*$ , U.S. annual money demand (1901–1985)

regarding the long-run relationship between the series. For the two tests that rejected the null of no cointegration, the breakpoint for the C/S tests is at the middle of the sample with  $\hat{\tau} = 0.52$  (1975:2).

## 7. Final remarks

The concept of cointegration, originally suggested by Granger (1981) and formulated by Engle and Granger (1987), is that over the long run a special time-invariant linear combination of nonstationary variables may be stationary. As empirical work has progressed, it is becoming clear some applied economists are interested in allowing the cointegrating relationships to change over time. In order for the concept of cointegration to retain significant empirical content, empirical work will have to restrict the types of structural change permitted. The most basic type of structural change is the one-time regime shift model. Other models are of course possible, but may require more careful analysis.

If structural change is to be entertained in cointegrated models, applied economists need appropriate test statistics to determine if there is any evidence for such a model. The standard testing procedure is to set up the null of no cointegration against the alternative of cointegration, so rejection is considered evidence in favor of the model. In this paper we extend this family of test statistics by setting the alternative hypothesis to be cointegration while allowing for a one-time regime shift of unknown timing. Rejection of the null hypothesis, therefore, provides evidence in favor of this specification.

It is important to note, however, that this type of hypothesis test does not provide much evidence concerning the question of whether or not there was a regime shift, since the alternative hypothesis contains as a special case the standard model of cointegration with no regime shift. The naive alternative of examining the individual or joint significance of the dummy variables is dangerously flawed for two reasons. First, OLS estimation is not efficient, and test statistics do not generally have asymptotic standard distributions under the hypothesis of cointegration. Second, under the hypothesis of no regime shift, the date of the break is not identified, so even efficient test statistics have non-standard distributions. Appropriate statistics and their asymptotic distributions for testing the hypothesis of no regime shift against the alternative of a regime shift are given in Hansen (1992a) and Quintos and Phillips (1993). The test statistics of the present paper complement these tests, and both types of test statistics are potentially useful.

As illustrated in our analysis of the money-demand relationship in the previous section, we believe that empirical investigations will be best served by using a number of complementary statistical tests. One difficult task for the applied researcher is to juggle these separate pieces of the puzzle, but we can offer a few suggestions. The standard *ADF* statistic and our *ADF\** statistics both test the null of no cointegration, so rejection by either statistic implies that there is some long-run relationship in the data. If the standard *ADF* statistic does not reject, but the *ADF\** does, this implies that structural change in the cointegrating vector may be important. If both the *ADF* and the *ADF\** reject, no inference that structural change has occurred is warranted from this piece of information alone, since the *ADF\** statistic is powerful against conventional cointegration. In this event, the tests of Hansen (1992a) are useful to determine whether the cointegrating relationship has been subject to a regime shift. Unfortunately at this stage we have little guidance as to how to control for Type I error under such procedures which suggests that additional Monte Carlo work would be worthwhile.

The analysis of this paper has been confined to the question of developing residual-based tests for cointegration in the presence of a regime shift. We have not addressed the issues of efficient testing or efficient estimation, and leave these subjects for future research. For example, it would be interesting and useful to develop an analogous set of test statistics using the likelihood ratio approach advocated by Johansen (1988, 1991). Some progress along these lines has been recently initiated by H. Hansen and Johansen (1993).

### Appendix: Proof of the theorem

We will rigorously prove the Theorem for a simplified setting in which the test statistic is constructed from residuals from the following regression model:

$$y_{1t} = \alpha_1^\top y_{2t} + \alpha_2^\top y_{2t} \varphi_{rt} + e_t \quad (\text{A.1})$$

which is the same as model 4, except that there is no intercept or level shift. See our working paper version, Gregory and Hansen (1992), for a complete proof for model 4. To simplify the presentation we assume that  $y_0 = 0$ .

A.1. Fundamental convergence results

Throughout the proof,  $\Rightarrow$  denotes weak convergence of the associated probability measures with respect to the uniform metric over  $\tau \in [0,1]$  or  $\tau \in T$ , where appropriate, and  $\{\cdot\}$  denotes the indicator function. Let  $S_t = \sum_{i=1}^t u_i$  be the cumulative sum of the innovations, and partition  $S_t = (S_{1t}, S_{2t}^\top)^\top$  in conformity with  $u_t$ . Set  $S_{2\tau} = S_{2t} \varphi_{\tau}$  where  $\varphi_{\tau} = \{t > [n\tau]\}$ , and set  $X_{\tau} = (S_t^\top, S_{2\tau}^\top)^\top$ . It will be convenient to also define the subvector  $X_{2\tau} = (S_{2t}^\top, S_{2\tau}^\top)^\top$ , as these are the regressors in our model (A.1).

The starting point for the asymptotic analysis is the multivariate invariance principle

$$n^{-1/2} S_{[n\tau]} \Rightarrow B(\tau), \tag{A.2}$$

where  $B(\tau)$  is a vector Brownian motion with covariance matrix  $\Omega$ . This was shown by Herrndorf (1984) (see Phillips and Durlauf, 1986, for the extension to the vector case). Partition  $B(r) = (B_1 \ B_2^\top)^\top$  in conformity with  $S_t$ , define  $B_{2\tau}(r) = B_2(r) \varphi_{\tau}(r)$  and  $X_{\tau}(r) = (B(r)^\top, B_{2\tau}(r)^\top)^\top$  where  $\varphi_{\tau}(r) = \{r \geq \tau\}$ , and define the subvector  $X_{2\tau}(r) = (B_2(r)^\top, B_{2\tau}(r)^\top)^\top$ .

Our proof method involves explicitly writing out all test statistics as functions of partial sums of functions of the data, in contrast to the proof method of Zivot and Andrews (1992) who write the test statistics as explicit functions of the indicator function  $\phi_{\tau}$ . Our method is more cumbersome notationally, but is more convenient in terms of proving uniformity over the breakpoint  $\tau$ . The key to our approach rests on the joint weak convergence of

$$\frac{1}{n^2} \sum_{t=[n\tau]}^n S_t S_t^\top \Rightarrow \int_{\tau}^1 BB^\top \tag{A.3}$$

and

$$\frac{1}{n} \sum_{t=[n\tau]}^n S_t u_{t+1}^\top \Rightarrow \int_{\tau}^1 B dB^\top + (1 - \tau)A, \tag{A.4}$$

where the weak convergence in (A.3) and (A.4) is with respect to the uniform metric over  $\tau \in [0, 1]$ , and  $A = \lim_n (1/n) \sum_{t=1}^n E(S_t u_{t+1}^\top)$ . (A.3) follows from (A.2) and the continuous mapping theorem (CMT, see Billingsley, 1968, Thm. 5.1) since  $\int_{\tau}^1 BB^\top$  is continuous with respect to  $\tau$ . (A.4) is proved under our assumptions by Hansen (1992b, Thm. 4.1). Note that the latter result is stronger than the pointwise convergence theorems of Chan and Wei (1988) and Phillips (1988). To ensure that there is no confusion with mere pointwise convergence, for the

remainder of the proof we will refer to results such as (A.3) and (A.4) as holding ‘uniformly over  $\tau$ ’. The remainder of the proof is largely a series of repeated applications of the CMT to (A.3), (A.4), and functions of these sample moments.

A.2. Least squares coefficient process

(A.4) implies

$$\frac{1}{n^2} \sum_{t=1}^n X_{t\tau} X_{t\tau}^\top = \begin{bmatrix} \frac{1}{n^2} \sum_{t=1}^n S_t S_t^\top & \frac{1}{n^2} \sum_{t=[n\tau]}^n S_t S_{2t}^\top \\ \frac{1}{n^2} \sum_{t=[n\tau]}^n S_{2t} S_t^\top & \frac{1}{n^2} \sum_{t=[n\tau]}^n S_{2t} S_{2t}^\top \end{bmatrix} \tag{A.5}$$

$$\Rightarrow \begin{bmatrix} \int_0^1 B B^\top & \int_0^\tau B B_2^\top \\ \int_0^1 B_2 B^\top & \int_0^\tau B_2 B_2^\top \end{bmatrix} = \int_0^1 X_\tau X_\tau^\top,$$

uniformly over  $\tau \in [0,1]$ .

Define  $\hat{\alpha}_\tau = (\hat{\alpha}_1^\top, \hat{\alpha}_2^\top)^\top$  as the least-squares estimator of (A.1) for each  $\tau$ , and set  $\hat{\eta}_\tau = (1, -\hat{\alpha}_\tau^\top)^\top$ . It follows from (A.5) and the CMT that

$$\hat{\eta}_\tau = \begin{pmatrix} 1 \\ -\left(\frac{1}{n^2} \sum_{t=1}^n X_{2t\tau} X_{2t\tau}^\top\right)^{-1} \left(\frac{1}{n^2} \sum_{t=1}^n X_{2t\tau} S_{1t}\right) \end{pmatrix} \tag{A.6}$$

$$\Rightarrow \begin{pmatrix} 1 \\ -\left(\int_0^1 X_{2\tau} X_{2\tau}^\top\right)^{-1} \int_0^1 X_{2\tau} B_1 \end{pmatrix} \equiv \eta_\tau,$$

uniformly over  $\tau \in T$ . Note that we need to restrict  $\tau$  to a compact subset of  $(0,1)$  since  $X_{2t\tau}$  is multicollinear when  $\tau = 0$  or  $\tau = 1$ .

A.3. Convergence to the stochastic integral process

Partition  $\Lambda$  and  $\Omega$  in conformity with  $S_i$ :

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

and set  $\Lambda_2 = (\Lambda_{21} \ \Lambda_{22})$  and  $\Lambda_{\cdot 2} = (\Lambda_{12}^\top \ \Lambda_{22}^\top)^\top$ . (A.5) implies that

$$\frac{1}{n} \sum_{t=1}^n X_{t\tau} \Delta S_{t+1}^\top = \begin{bmatrix} \frac{1}{n} \sum_{t=1}^n S_t u_{t+1}^\top \\ \frac{1}{n} \sum_{t=[n\tau]}^n S_{2t} u_{t+1}^\top \end{bmatrix} \tag{A.7}$$

$$\Rightarrow \begin{bmatrix} \int_0^1 B dB^\top + \Lambda \\ \int_\tau^1 B_2 dB^\top + (1-\tau) \Lambda_2 \end{bmatrix} = \int_0^1 X_\tau dB^\top + \begin{bmatrix} \Lambda \\ (1-\tau) \Lambda_2 \end{bmatrix},$$

uniformly over  $\tau \in [0, 1]$ .

The process  $S_{2t\tau}$  has differences

$$\Delta S_{2t\tau} = \Delta S_{2t} \varphi_{t\tau} + S_{2t-1} \Delta \varphi_{t\tau}, \tag{A.8}$$

where  $\Delta \varphi_{t\tau} = \varphi_{t\tau} - \varphi_{t-1\tau} = \{t = [n\tau]\}$ . Note that  $\varphi_{t\tau} \Delta \varphi_{t+1\tau} = 0$ . To derive the large-sample counterpart we need to define the differential  $d\varphi_\tau$ . Since  $\varphi_\tau(r)$  is a step function with a jump at  $\tau$ ,  $d\varphi_\tau$  is naturally defined as the dirac function with the property that for all functions  $f(\cdot)$  with left-limits and all  $a < \tau < b$ ,  $\int_a^b f d\varphi_\tau = \lim_{z \uparrow \tau} f(z)$ . Note that  $\varphi_\tau d\varphi_\tau = 0$  since the left-limit of  $\varphi_\tau$  at  $\tau$  is 0. We then define the differential  $dB_{2\tau}(r) = \varphi_\tau(r) dB_2(r) + B_2(r) d\varphi_\tau(r)$ , giving the relationships

$$\int_0^1 B dB_{2\tau}^\top = \int_\tau^1 B dB_2^\top + B(\tau) B_2(\tau)^\top \tag{A.9}$$

and

$$\int_0^1 B_{2\tau} dB_{2\tau}^\top = \int_\tau^1 B_{2\tau} dB_2^\top. \tag{A.10}$$

Using (A.8), (A.9), and (A.10),

$$\frac{1}{n} \sum_{t=1}^n X_{t\tau} \Delta S_{2t+1\tau}^\top = \begin{bmatrix} \frac{1}{n} \sum_{t=[n\tau]-1}^n S_t u_{2t+1}^\top + \frac{1}{n} S_{[n\tau]-1} S_{2[n\tau]-1}^\top \\ \frac{1}{n} \sum_{t=[n\tau]-1}^n S_{2t} u_{2t+1}^\top \end{bmatrix} \tag{A.11}$$

$$\Rightarrow \begin{bmatrix} \int_\tau^1 B dB_2^\top + (1-\tau) \Lambda_2 + B(\tau) B_2(\tau)^\top \\ \int_\tau^1 B_2 dB_2^\top + (1-\tau) \Lambda_{22} \end{bmatrix} = \int_0^1 X_\tau dB_{2\tau}^\top + (1-\tau) \begin{bmatrix} \Lambda_{\cdot 2} \\ \Lambda_{22} \end{bmatrix},$$

where the convergence is uniform over  $\tau \in [0, 1]$ . Setting  $dX_\tau^\top = (dB^\top, dB_{2\tau}^\top)^\top$ , then (A.7) and (A.11) yield

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n X_{t\tau}^\top \Delta X_{t+1\tau}^\top &= \left( \frac{1}{n} \sum_{t=1}^n X_{t\tau} \Delta S_{t+1}^\top, \frac{1}{n} \sum_{t=1}^n X_{t\tau} \Delta S_{2t+1\tau}^\top \right) \\ &\Rightarrow \\ \int_0^1 X_\tau dX_\tau^\top + A_\tau, \end{aligned} \tag{A.12}$$

uniformly over  $\tau$ , where

$$A_\tau = \begin{bmatrix} A & (1-\tau)A_{.2} \\ (1-\tau)A_{.2} & (1-\tau)A_{22} \end{bmatrix}.$$

A.4. Serial correlation coefficient estimate

Note that  $\hat{\epsilon}_{t\tau} = \hat{\eta}_\tau^\top X_{t\tau}$  so by (A.2), (A.3), and the CMT,

$$\begin{aligned} \frac{1}{n^2} \sum_{t=1}^n \hat{\epsilon}_{t\tau}^2 &= \hat{\eta}_\tau^\top \frac{1}{n^2} \sum_{t=1}^n X_{t\tau} X_{t\tau}^\top \hat{\eta}_\tau \\ &\Rightarrow \\ \eta_\tau^\top \int_0^1 X_\tau X_\tau^\top \eta_\tau &= \int_0^1 X_\tau^{*2}, \end{aligned} \tag{A.13}$$

uniformly over  $\tau \in T$ , where  $X_\tau^*(r) = B_1(r) - (\int_0^1 B_1 X_{2\tau}^\top) (\int_0^1 X_{2\tau} X_{2\tau}^\top)^{-1} X_{2\tau}(r)$  is the stochastic process in  $(r, \tau)$  obtained by projecting  $B_1(r)$  orthogonal to  $X_{2\tau}(r)$ . Define  $\sigma = [\Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21}]^{1/2}$ , and set  $W_1(r) = \sigma^{-1} [B_1 - \Omega_{12} \Omega_{22}^{-1} B_2(r)]$  and  $W_2(r) = \Omega_{22}^{-1/2} B_2(r)$ , so that  $W_1 \equiv BM(1)$  is independent of  $W_2 \equiv BM(I_m)$ . By standard projection arguments,  $X_\tau^*(r) = \sigma W_\tau(r)$ , where  $W_\tau(r) = W_1(r) - (\int_0^1 W_1 W_{2\tau}^\top) (\int_0^1 W_{2\tau} W_{2\tau}^\top)^{-1} W_{2\tau}(r)$ , and  $W_{2\tau}(r) = [W_2^\top(r), W_2^\top(r) \varphi_\tau(r)]^\top$ . Thus (A.13) becomes

$$\frac{1}{n^2} \sum_{t=1}^n \hat{\epsilon}_{t\tau}^2 \Rightarrow \sigma^2 \int_0^1 W_\tau^2, \tag{A.14}$$

uniformly over  $\tau \in T$ .

Similarly using (A.12) and the CMT,

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \hat{\epsilon}_{t\tau} \Delta \hat{\epsilon}_{t+1\tau} &= \hat{\eta}_\tau^\top \frac{1}{n} \sum_{t=1}^n X_{t\tau} \Delta X_{t+1\tau}^\top \hat{\eta}_\tau \\ &\Rightarrow \\ \eta_\tau^\top \left( \int_0^1 X_\tau dX_\tau^\top + A_\tau \right) \eta_\tau &= \sigma^2 \int_0^1 W_\tau dW_\tau + \eta_\tau^\top A_\tau \eta_\tau, \end{aligned} \tag{A.15}$$

uniformly over  $\tau \in T$ . (A.14), (A.15), and the CMT yield

$$n(\hat{\rho}_\tau - 1) = \frac{\frac{1}{n} \sum_{t=1}^n \hat{e}_{t\tau} \Delta \hat{e}_{t+1\tau}}{\frac{1}{n^2} \sum_{t=1}^n \hat{e}_{t\tau}^2} \Rightarrow \frac{\sigma^2 \int_0^1 W_\tau dW_\tau + \eta_\tau^\top A_\tau \eta_\tau}{\sigma^2 \int_0^1 W_\tau^2}, \tag{A.16}$$

uniformly over  $\tau \in T$ . The limit in (A.16) exists for  $\tau \in T$  since  $\int_0^1 W_\tau^2 > 0$  a.s. by Lemma A.2 of Phillips and Hansen (1990).

A.5. Bias correction

The Phillips’ statistics are constructed using the statistic  $\hat{\lambda}_\tau = \sum_{j=1}^M w(j/M) \times (1/n) \sum_t \hat{v}_{t-j\tau} \hat{v}_{t\tau}$  where  $\hat{v}_{t\tau} = \hat{e}_{t\tau} - \hat{\rho}_\tau \hat{e}_{t-1\tau} = \Delta \hat{e}_{t\tau} - (\hat{\rho}_\tau - 1) \hat{e}_{t-1\tau}$ . For brevity, denote  $w(j/M)$  by  $w_j$ . Now by the triangle inequality and Holder’s inequality,

$$\begin{aligned} \left| \sum_{j=1}^M w_j \frac{1}{n} \sum_t \Delta \hat{e}_{t-j\tau} (\hat{v}_{t\tau} - \Delta \hat{e}_{t\tau}) \right| &\leq M \max_{j \leq M} \left| \frac{1}{n} \sum_t \Delta \hat{e}_{t-j\tau} \hat{e}_{t-1\tau} \right| |(\hat{\rho}_\tau - 1)| \\ &\leq \frac{M}{\sqrt{n}} \left| \frac{1}{n} \sum_t \Delta \hat{e}_{t\tau}^2 \right|^{1/2} \left| \frac{1}{n^2} \sum_t \hat{e}_{t-1\tau}^2 \right|^{1/2} |n(\hat{\rho}_\tau - 1)| \\ &\rightarrow_p 0, \end{aligned}$$

uniformly in  $\tau$ , since  $M/\sqrt{n} \rightarrow 0$  and  $\sup_\tau |n(\hat{\rho}_\tau - 1)| = O_p(1)$  by (A.16). Similarly, we can replace  $\Delta \hat{e}_{t-j\tau}$  with  $\hat{v}_{t-j\tau}$  to show that

$$\hat{\lambda}_\tau = \sum_{j=1}^M w_j \frac{1}{n} \sum_t \Delta \hat{e}_{t-j\tau} \Delta \hat{e}_{t\tau} + o_p(1), \tag{A.17}$$

uniformly in  $\tau$ . Now note that

$$\sum_{j=1}^M w_j \frac{1}{n} \sum_t \Delta X_{t-j\tau} \Delta X_{t\tau}^\top = \begin{bmatrix} \sum_{j=1}^M w_j \frac{1}{n} \sum_t u_{t-j} u_t^\top & \sum_{j=1}^M w_j \frac{1}{n} \sum_t u_{t-j} \Delta S_{2t\tau}^\top \\ \sum_{j=1}^M w_j \frac{1}{n} \sum_t \Delta S_{2t-j\tau} u_t^\top & \sum_{j=1}^M w_j \frac{1}{n} \sum_t \Delta S_{2t-j\tau} \Delta S_{2t\tau}^\top \end{bmatrix} \tag{A.18}$$

In Hansen (1992c, Thm. 1), it was shown under our assumptions that

$$\sum_{j=1}^M w_j \frac{1}{n} \sum_{t=1+j}^n u_{t-j} u_t^\top \rightarrow_p A,$$



which gives the asymptotic behavior of the upper-left element of (A.18). A slight modification of the proof in Hansen (1992c) yields the stronger result

$$\sum_{j=1}^M w_j \frac{1}{n} \sum_{t=1+j}^{[n\tau]} u_{t-j} u_t^\top \rightarrow_p \tau A, \tag{A.19}$$

uniformly in  $\tau \in [0, 1]$ . This is possible since Theorem 1 in Hansen (1992c) is based on a moment inequality which can be strengthened to a maximal inequality via Lemma 2 of Hansen (1991).

Now using (A.8), we can obtain the asymptotic behavior of the lower-left element of (A.18):

$$\begin{aligned} \sum_{j=1}^M w_j \frac{1}{n} \sum_t \Delta S_{2t-j\tau} u_t^\top &= \sum_{j=1}^M w_j \frac{1}{n} \sum_t (\varphi_{t-j\tau} u_{2t-j\tau} u_t^\top + \Delta \varphi_{t-j\tau} S_{2t-j-1} u_t^\top) \\ &= \sum_{j=1}^M w_j \frac{1}{n} \sum_{t \geq [n\tau]+j} u_{2t-j\tau} u_t^\top + \sum_{j=1}^M w_j \frac{1}{n} S_{2[n\tau]-1} u_{[n\tau]+j}^\top \\ &\rightarrow_p (1 - \tau) A_2, \end{aligned} \tag{A.20}$$

uniformly in  $\tau \in [0, 1]$  by (A.19) and the fact that

$$\left| \sum_{j=1}^M w_j \frac{1}{n} S_{2[n\tau]-1} u_{[n\tau]+j}^\top \right| \leq \frac{M}{\sqrt{n}} \max_{t \leq n} \left| \frac{1}{\sqrt{n}} S_{2t} \right| \max_{t \leq n} \left| \frac{1}{M} \sum_{j=1}^M w_j u_{t+j} \right| \leq o_p(1),$$

where the final inequality follows from Markov’s inequality and

$$E \max_{t \leq n} \left| \frac{1}{M} \sum_{j=1}^M w_j u_{t+j} \right|^2 = O(1),$$

which is an application of Corollary 3 of Hansen (1991).

Similar analysis for the upper-right and lower-right elements of (A.18), combined with (A.19) and (A.20), yield the complete result

$$\sum_{j=1}^M w_j \frac{1}{n} \sum_t \Delta X_{t-j\tau} \Delta X_{t\tau}^\top \longrightarrow_p A_\tau, \tag{A.21}$$

uniformly in  $\tau \in [0, 1]$ . Finally, (A.17), (A.21), the fact that  $\Delta \hat{e}_{t\tau} = \hat{\eta}_\tau^\top \Delta X_{t\tau}$ , and (A.3) yield

$$\hat{\lambda}_\tau = \hat{\eta}_\tau^\top \sum_{j=1}^M w_j \frac{1}{n} \sum_t \Delta X_{t-j\tau} \Delta X_{t\tau}^\top \hat{\eta}_\tau + o_p(1) \Rightarrow \eta_\tau^\top A_\tau \eta_\tau, \tag{A.22}$$

uniformly over  $\tau \in T$ .

This allows us to find the limit distribution of  $Z_\alpha(\tau)$ . By (A.14), (A.16), (A.22), and the CMT,

$$\begin{aligned}
 Z_\alpha(\tau) &= \frac{\frac{1}{n^2} \sum_{t=1}^n \hat{\epsilon}_{1t} \Delta \hat{\epsilon}_{t+1\tau} - \hat{\lambda}_\tau}{\frac{1}{n^2} \sum_{t=1}^n \hat{\epsilon}_{1t}^2} \\
 &\Rightarrow \\
 &\frac{\sigma^2 \int_0^1 W_\tau dW_\tau + \eta_\tau^\top A_\tau \eta_\tau - \eta_\tau^\top A_\tau \eta_\tau}{\sigma^2 \int_0^1 W_\tau^2} = \frac{\int_0^1 W_\tau dW_\tau}{\int_0^1 W_\tau^2}, \tag{A.23}
 \end{aligned}$$

uniformly in  $\tau \in T$ . Since the supremum mapping is continuous in the uniform metric, (A.23) and the CMT imply that

$$Z_\alpha^* = \inf_{\tau \in T} Z_\alpha(\tau) \Rightarrow \inf_{\tau \in T} \frac{\int_0^1 W_\tau dW_\tau}{\int_0^1 W_\tau^2},$$

as stated in the theorem.

A.6. Long-run variance estimate

We can define a matrix  $\Omega_\tau$ , analogous to  $A_\tau$ :

$$\Omega_\tau = \begin{pmatrix} \sigma^2 & \Omega_{12} & (1-\tau)\Omega_{12} \\ \Omega_{21} & \Omega_{22} & (1-\tau)\Omega_{22} \\ (1-\tau)\Omega_{21} & (1-\tau)\Omega_{22} & (1-\tau)\Omega_{22} \end{pmatrix}.$$

We can factor this as  $\Omega_\tau = \Omega_e^\top D_\tau^* \Omega_e$  where

$$D_\tau^* = \begin{pmatrix} 1 & 0 \\ 0 & D_\tau \end{pmatrix}, \quad D_\tau = \begin{pmatrix} I_m & (1-\tau)I_m \\ (1-\tau)I_m & (1-\tau)I_m \end{pmatrix},$$

and

$$\Omega_e = \begin{pmatrix} \sigma & 0 \\ \underline{\Omega}_{21} & \underline{\Omega}_{22}^{1/2} \end{pmatrix},$$

with

$$\underline{\Omega}_{21} = \begin{pmatrix} \Omega_{22}^{-1/2} \Omega_{21} \\ 0 \end{pmatrix} \quad \text{and} \quad \underline{\Omega}_{22} = \begin{pmatrix} \Omega_{22} & 0 \\ 0 & \Omega_{22} \end{pmatrix}.$$

Standard manipulations show that

$$\underline{\Omega}_{22}^{1/2} \alpha_\tau = \left( \int_0^1 W_{2\tau} W_{2\tau}^\top \right)^{-1} \int_0^1 W_{2\tau} W_1 \sigma + \underline{\Omega}_{21} = \kappa_\tau + \underline{\Omega}_{21},$$

so

$$\sigma^{-1} \Omega_e \eta_\tau = \sigma^{-1} \begin{pmatrix} \sigma & 0 \\ \underline{\Omega}_{21} & \underline{\Omega}_{22}^{1/2} \end{pmatrix} \begin{pmatrix} 1 \\ -\alpha_\tau \end{pmatrix} = \begin{pmatrix} 1 \\ -\kappa_\tau \end{pmatrix}.$$

Analysis similar to that which led to (A.22) can show that

$$\hat{\sigma}_\tau^2 \Rightarrow \eta_\tau^\top \Omega_\tau \eta_\tau = (1 - \kappa_\tau^\top) \begin{pmatrix} 1 & 0 \\ 0 & D_\tau \end{pmatrix} \begin{pmatrix} 1 \\ -\kappa_\tau \end{pmatrix} = 1 + \kappa_\tau^\top D_\tau \kappa_\tau, \tag{A.24}$$

uniformly in  $\tau \in T$ . We can derive from (A.23), (A.24), and the CMT the limit distribution of the  $Z_t$  process

$$Z_t(\tau) = Z_\alpha(\tau) \left( \frac{1}{\hat{\sigma}_\tau^2 n^2} \sum_{t=1}^n \hat{e}_{t\tau}^2 \right)^{1/2} \Rightarrow \frac{\int_0^1 W_\tau dW_\tau}{\left( \int_0^1 W_\tau^2 \right)^{1/2} \left( \kappa_\tau^\top D_\tau \kappa_\tau \right)^{1/2}}, \tag{A.25}$$

uniformly in  $\tau \in T$ .

Once again, the continuity of the supremum mapping in the uniform metric allows the CMT to be applied to (A.25) to yield our final result:

$$Z_\tau^* = \inf_{\tau \in T} Z_t(\tau) \Rightarrow \inf_{\tau \in T} \frac{\int_0^1 W_\tau dW_\tau}{\left( \int_0^1 W_\tau^2 \right)^{1/2} \left( \kappa_\tau^\top D_\tau \kappa_\tau \right)^{1/2}}.$$

### References

Andrews, D.W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.

Andrews, D.W.K. and J.C. Monahan, 1992, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica* 60, 953–966.

Banerjee, A., J.J. Dolado, D.F. Hendry, and G.W. Smith, 1986, Exploring equilibrium relationships in econometrics through static models: Some Monte Carlo evidence, *Oxford Bulletin of Economics and Statistics* 48, 253–277.

Banerjee, A., R. Lumsdaine, and J. Stock, 1992, Recursive and sequential tests of the unit-root and trend-break hypotheses: Theory and international evidence, *Journal of Business and Economic Statistics* 10, 271–287.

Billingsley, P., 1968, *Convergence of probability measures* (Wiley, New York, NY).

Chan, N.H. and C.Z. Wei, 1988, Limiting distributions of least-squares estimates of unstable autoregressive processes, *Annals of Statistics* 16, 367–401.

- Christiano, L.J., 1992, Searching for a break in GNP, *Journal of Business and Economic Statistics* 10, 237–250.
- Engle, R.F. and C.W.J. Granger, 1987, Cointegration and error correction: Representation, estimation, and testing, *Econometrica* 55, 251–276.
- Granger, C.W.J., 1981, Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, 121–130.
- Gregory, A.W., 1994, Testing for cointegration in linear quadratic models, *Journal of Business and Economic Statistics* 2, 347–360.
- Gregory, A.W., J.M. Nason, and D. Watt, 1994, Testing for structural breaks in cointegrated relationships, *Journal of Econometrics*, forthcoming.
- Hansen, B.E., 1991, Strong laws for dependent heterogeneous processes, *Econometric Theory* 7, 213–221.
- Hansen, B.E., 1992a, Tests for parameter instability in regressions with  $I(1)$  processes, *Journal of Business and Economic Statistics* 10, 321–335.
- Hansen, B.E., 1992b, Convergence to stochastic integrals for dependent heterogeneous processes, *Econometric Theory* 8, 489–500.
- Hansen, B.E., 1992c, Consistent covariance matrix estimation for dependent heterogeneous processes, *Econometrica* 60, 967–972.
- Hansen, H. and S. Johansen, 1993, Recursive estimation in cointegrated VAR models, Unpublished manuscript (University of Copenhagen, Copenhagen).
- Herrndorf, N., 1984, A functional central limit theorem for weakly dependent sequences of random variables, *Annals of Probability* 12, 829–839.
- Johansen, S., 1988, Statistical analysis of cointegrating vectors, *Journal of Economic Dynamics and Control* 12, 231–254.
- Johansen, S., 1991, Estimation and hypothesis testing of cointegration vectors in the presence of linear trend, *Econometrica* 59, 1551–1580.
- Lucas, R.E., 1988, Money demand in the United States: A quantitative review, *Carnegie–Rochester Conference Series on Public Policy* 29, 137–168.
- MacKinnon, J.G., 1991, Critical values for cointegration tests, in: R.F. Engle and C.W.J. Granger, eds., *Long-run economic relationships: Readings in cointegration* (Oxford Press, Oxford) 267–276.
- Perron, P., 1989, The great crash, the oil price shock, and the unit root hypothesis, *Econometrica* 57, 1361–1401.
- Perron, P. and T.J. Vogelsang, 1992, Nonstationarity and level shifts with an application to purchasing power parity, *Journal of Business and Economic Statistics* 10, 301–320.
- Phillips, P.C.B., 1987, Time series regression with a unit root, *Econometrica*, 55, 277–301.
- Phillips, P.C.B., 1988, Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations, *Econometric Theory* 4, 528–533.
- Phillips, P.C.B. and S.N. Durlauf, 1986, Multiple time series regression with integrated processes, *Review of Economic Studies* 53, 473–495.
- Phillips, P.C.B. and B.E. Hansen, 1990, Statistical inference in instrumental variables regression with  $I(1)$  processes, *Review of Economic Studies* 57, 99–125.
- Phillips, P.C.B. and S. Ouliaris, 1990, Asymptotic properties of residual based tests for cointegration, *Econometrica*, 58, 165–193.
- Quintos, C.E. and P.C.B. Phillips, 1993, Parameter constancy in cointegrating regressions, *Empirical Economics* 18, 675–703.
- Stock, J.H. and M.W. Watson, 1993, A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* 61, 783–820.
- Zivot, E. and D.W.K. Andrews, 1992, Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis, *Journal of Business and Economic Statistics* 10, 251–270.