



ELSEVIER

Journal of Econometrics 86 (1998) 337–368

**JOURNAL OF
Econometrics**

Test for partial parameter instability in regressions with I(1) processes

Biing-Shen Kuo*

Graduate Institute of International Trade, National Chengchi University, Taipei 116, Taiwan

Received 1 December 1993; received in revised form 1 November 1997

Abstract

This paper derives the limiting distribution of LM-type tests for possible departure from constancy in ‘subsets’ of cointegrating coefficients. In particular, models with nonconstancy on intercept or stochastic trend coefficients are considered. It is found that the limiting representations of these subset tests can be characterized as functions of continuous-time martingales depending on the asymptotics of both the whole regressor vector and the regressors whose coefficients are under tests. Critical values are computed using large-sample approximation. Monte Carlo experiments are conducted to investigate the finite sample size and power. The subset tests are found to dominate the joint test when there is partial coefficient variation. © 1998 Elsevier Science B.V. Published by Elsevier Science S.A. All rights reserved.

JEL classification: C12; C15; C22; C52

Keywords: Cointegration; Subset test; Fully modified estimator; Brownian motion

1. Introduction

The problem of regression parameter constancy has a long history in economics. In the context of independent and stationary data, there is a large literature concerning the test statistics for structural change (e.g. Chow, 1960). Recently, more attention has been given empirically to models with non-stationary regressors, starting with the seminal paper of Engle and Granger

*Corresponding author. E-mail: bsku@cc.nccu.edu.tw

(1987). The notion of cointegration delivers rich implications for long-run relationships among integrated variables. This has given rise to the desire to develop parameter constancy tests valid in the context of non-stationary but cointegrating variables.

This article contributes to this new literature. Yet the focal point is testing for the constancy of 'partial' cointegrating coefficients, which pertains to only a subset of the coefficients in the model, using the terminology of Andrews and Fair (1988). This is partly motivated by empirical considerations: economic theory might suggest implications only for the constancy of a subset of the cointegrating slope coefficients in the model. For instance, one might wish to investigate the existence and stability of a long-run money demand function, as demonstrated later as an illustrative example of the proposed tests. A long-run money demand equation could be characterized empirically as a cointegrating relation among real balances, real income, and interest rates, as suggested by Stock and Watson (1993). To analyze the stability of the equation, researchers would be interested in learning not only whether the entire equation is stable, but also whether the individual cointegrating coefficient: the intercept, the income elasticity or the interest semielasticity, varies over the sampling period. The need for tests of this sort appears indisputably clear.

Like the classical Chow test for partial structural change, testing for the stability of partial cointegrating coefficients is justified similarly. In the classical regression model, when the F test rejects, one or more t ratios for individual parameters may be large enough to explain the rejection. As an extension of this procedure to the models with integrated processes, we need analogs of the ' t ratios' of the classical model which allow us to test a particular subset of the cointegrating coefficients. In principle, these subset tests should enable researchers to find the subsets of cointegrating coefficients responsible for the finding of model instability.

To make this extension, a class of the subset tests for constancy of cointegrating coefficients is developed here. It is built on the work of Hansen (1992a) which deals with the joint test statistics. In particular, the models in which the coefficients have nonconstancy of random walk or jump process on intercept or on stochastic trends are considered. The test statistics are derived by the Lagrange multiplier (LM) principle. Their stationary analogs are documented in Nyblom (1989) for random variations and Andrews (1993) for structural breaks. Using the fully modified estimator of Phillips and Hansen (1990), the limiting distributions of the statistics are derived under the null of coefficient constancy. Their asymptotic distributions are non-standard, free of nuisance parameters and characterized as functions of continuous-time martingales. It should be stressed that the asymptotics of these subset tests depend on those of both the whole regressor vector and the regressors of cointegrating coefficients under test. It is essential to know the nature of the regressors for correct application of the tests.

The advantages of the subset tests proposed are, in fact, deeper than their usefulness in applications. Both joint and partial testing procedures can be performed to test for the constancy of a subset of parameters. The obvious issue is which procedure is better. The distributional theory of the test statistics does not provide an answer to this question. However, from Monte Carlo simulations, our subset tests are found to have reasonable power, and dominate the joint test under the alternative that only partial parameters vary to the extent of economic interest. Here some of the information contained in the joint test is not informative but contaminates the estimation. This suggests that it might pay the price of power loss when introducing redundant information into the test.

Intuitively, the test statistics make use of the properties of non-stationary data. Under the alternative of coefficient variations, the partial sums of the residuals behave as those in spurious regression (see Phillips, 1986), and converge to random variables only after renormalization. As functions of the partial sum of residuals, the test statistics hence are consistent. Several recent works also employ this notion in constructing similar test statistics. Kiwatkowski et al. (1992) applied it to testing for stationarity in a univariate time series. Quintos and Phillips (1993) and Shin (1994) work on cointegrating models, and differ from this paper in some ways. Their tests correspond to one of the test statistics here, although signified differently. Their tests are Nyblom-styled and are to detect a martingale specification for parameter variation. Neither paper proposed tests for structural change of unknown timing. On the other hand, Shin (1994) addressed the specific question of testing for the null of *cointegration*. His test is equivalent to the Nyblom-styled test for intercept variation which is one fully modified special case of our tests.

The remainder of the paper is organized as follows. Section 2 discusses the estimation of a cointegration model. The test statistics for coefficient non-constancy are described in Section 3. Section 4 is the core of the paper where the limiting distributions of the tests are derived and their asymptotic critical values are given. The results for gauging the small sample performance of the tests are reported in Section 5, and a brief analysis of the stability of cointegration relation in U.S. money demand is taken up in Section 6. Section 7 is the conclusion. The proof of the main theoretical results are left to the Appendix.

2. The estimation for a cointegration model

2.1. The model

We will stay in a standard cointegrated regression model where the cointegration vectors remain stable over the sampling period.

$$y_t = A_1 + A_2x_{2t} + A_3x_{3t} + u_{1t}, \quad t = 1, \dots, T, \quad (1)$$

where the processes x_{2t} and x_{3t} are generated by

$$\begin{aligned} (x_{2t}, x_{3t}) &= (\pi_1 + \pi_2 k_{2t} + x_{2t}^o, \phi_1 + \phi_2 k_{2t} + x_{3t}^o), \\ (x_{2t}^o, x_{3t}^o) &= (x_{2t-1}^o + u_{2t}, x_{3t-1}^o + u_{3t}), \end{aligned} \tag{2}$$

the error vector

$$u'_t = \begin{pmatrix} u'_{1t} & u'_{2t} & u'_{3t} \\ m_1 & m_2 & m_3 \end{pmatrix},$$

is a sequence of stationary innovation vectors with zero mean, and the sequence k_{2t} , whose components are the power integer of time up to p , is used to describe the time trend. That is, $k_{2t} = (t, t^2, \dots, t^p)'$. Note that k_{2t} is not directly made up of the processes of stochastic trend regressors x_{2t} and x_{3t} . It is also assumed that x_{2t} and x_{3t} , respectively, have m_2 and m_3 elements.

This model is fairly general and encompasses a wide spectrum of variants that cover usual applications. In particular, when $\pi_2 \neq 0$ or $\phi_2 \neq 0$, it allows for the presence of deterministic trends in regressors. This is a common characterization for time series variables. The distribution theory, as will be shown later, can be derived for the model with or without k_{2t} .

2.2. Fully modified estimation

Our tests for partial parameter variation can be derived as LM tests in correctly specified likelihood problems. To construct the suggested tests, the model is estimated under the null of stable cointegrating coefficients.

We begin with defining the following matrices:

$$\begin{aligned} \Omega &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T E(u_j u'_i), \\ A &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^i E(u_j u'_i), \end{aligned} \tag{3}$$

partitioned conformably with u_t ,

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{pmatrix}, \quad A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}.$$

These nuisance parameters are used to describe the nature of temporal and long-run dependence among error series. Ω is usually referred to as long-run covariance matrix.

It is by now well known that the limiting distribution of OLS estimation of Eq. (1) depends on Ω and Λ (see Park and Phillips, 1989) in complicated ways. To overcome this problem, a number of estimators have been developed that can lead to parameter-invariant theory of inference. These include the maximum-likelihood estimator of Johansen (1988) and the leads and lags estimator of Saikkonen (1991) and Stock and Watson (1993). The solution adopted here is the fully modified estimator of Phillips–Hansen. We shall briefly describe this estimation technique.

In the first stage, the two-step Phillips–Hansen correction requires a consistent estimate of Ω and Λ used as an input in the subsequent procedure. Generally u_i is likely to be serially correlated and heteroscedastic. It is desirable to have estimates for Ω and Λ that are robust to these data properties. To that end, the residuals $\hat{u}'_t = (\hat{u}'_{1t}, \hat{u}'_{at})$ are computed, in which ‘ a ’ signifies ‘2’ and ‘3’ taken together, \hat{u}_{1t} is the least-squares residual obtained from regressing (1), and \hat{u}_{at} are the residuals from the regression of Δx_{at} against Δk_{2t} . The covariance matrices Ω and Λ are then estimated from \hat{u}_t through a kernel. Specifically,

$$\hat{\Omega} = \sum_{i=-T}^T w(i/M) \frac{1}{T} \sum_{j=i+1}^T \hat{u}_{j-i} \hat{u}'_j$$

and

$$\hat{\Lambda} = \sum_{i=0}^T w(i/M) \frac{1}{T} \sum_{j=i+1}^T \hat{u}_{j-i} \hat{u}'_j,$$

where $w(\cdot)$ is any kernel that gives a positive-semi-definite estimate such as the Bartlett, Parzen, and quadratic spectral (QS). M is a bandwidth parameter required to be sufficiently large as T grows in order to obtain the consistency. Andrews (1991), using an asymptotic mean-square error criterion, discussed the optimal choice of this number where M is chosen such that $M = O(T^{1/3})$ for the Bartlett kernel and $M = O(T^{1/5})$ for the Parzen and QS kernels, respectively, as $T \rightarrow \infty$ and $M \rightarrow \infty$. Here consistent with our simulation setup and application later, we shall assume that $M^3/T = O(1)$.

With the consistent estimates for Ω and Λ , we now are able to display the fully modified (FM) estimator of cointegration vector in the model above,

$$\hat{A}^+ = (\hat{A}_1^+, \hat{A}_2^+, \hat{A}_3^+) = \left[\sum_{i=1}^T (y_i^+ x'_i - (0, \hat{\Lambda}_{a1}^+)) \right] \left[\sum_{i=1}^T x_i x'_i \right]^{-1},$$

where $x'_i = (1', x'_{at})$ and $y_i^+ = y_i - \hat{\Omega}_{1a} \hat{\Omega}_{aa}^{-1} \hat{u}_{at}$. Accordingly, the residuals associated with the estimates are

$$\hat{u}'_{1t} = y_t^+ - \hat{A}^+ x_t.$$

Related to the LM approach adopted here, the scores corrected for bias in the problem are defined as

$$\hat{s}_t = \left[x_t \hat{u}_{1t}' - \begin{pmatrix} 0 \\ \hat{\lambda}_{a1}^+ \end{pmatrix} \right],$$

satisfying $\sum_{t=1}^T \hat{s}_t = 0$.

To develop the test statistics later, we also need the estimates for $\Omega_{1..a}$, the long-run variance of u_{1t} conditional on u_{at} , which is $\hat{\Omega}_{1..a} = \hat{\Omega}_{11} - \hat{\Omega}_{1a} \hat{\Omega}_{aa}^{-1} \hat{\Omega}_{a1}$.

3. Testing partial parameter variation

Following the standard model, two of the models allowing for partial change in coefficients on either deterministic or stochastic trend regressors are introduced in order. The first model allows the coefficients on the intercept to vary over time.

Model I: parameter change in intercept:

$$y_t = A_{1t} + A_2 x_{2t} + A_3 x_{3t} + u_{1t}, \quad t = 1, \dots, T. \tag{4}$$

Observe that the coefficient A_{1t} is time-varying while those on x_{2t} and x_{3t} are held constant. A parameterization of this sort models changes in the level in the dependent variable.

Model II: parameter changes in stochastic trend regressors:

$$y_t = A_1 + A_{2t} x_{2t} + A_3 x_{3t} + u_{1t}, \quad t = 1, \dots, T, \tag{5}$$

where the coefficients on x_{2t} are time-dependent. This is to entertain the change in the cointegrating slope in the model, such as income elasticity or interest semielasticity in a money demand function.

All the proposed tests have the same null hypothesis that the cointegrating coefficients of interest are stable during the sampling period. That is, $H_0 : A_{h,t} = A_h$. A_h is constant, and $A_{h,t}$ stands for those parameters in Model I or Model II that are considered to have the possibility of departure from constancy (A_{1t} in Model I, or A_{2t} in Model II). Here and elsewhere, we use 'h' that takes values '1' or '2' to signify those subscripts of the time-varying parameters in Model I or Model II, respectively.

In common with the literature, two types of parameter instability which the proposed tests are to detect will be considered here: random-walk variation and a single jump of unknown timing. The former captures the notion of the cointegrating coefficients moving gradually across time, while the latter stresses the feature of their immediate shifts. The difference will be embodied in the alternative and the formulation of the tests suggested below.

The first test we consider is to detect a single-structure break of unknown timing. In this event, $A_{h,t}$ follows a single-structural change at time i

$$A_{h,t} = \begin{cases} A_h, & t \leq i, \\ A_h + \Delta, & t > i, \end{cases}$$

where Δ denotes a jump of non-zero size, and $1 < i < T$.

Naturally, the alternative of the test is $H_a: \Delta \neq 0$ and i is unobservable. It is suggested as

$$Sup F_h = \sup_{t/T \in \mathcal{F}} F_{h,Tt},$$

where

$$F_{h,Tt} = \text{vec}(S_{h,Tt})' [\hat{\Omega}_{1,a} \otimes V_{h,Tt}]^{-1} \text{vec}(S_{h,Tt}),$$

$$S_{h,Tt} = \sum_{j=1}^t \hat{s}_{h,j},$$

$$V_{h,Tt} = M_{h,Tt} - M_{h,Tt}^* M_{TT}^{-1} M_{h,Tt}^{*'},$$

where $M_{h,Tt} = \sum_{j=1}^t x_{h,j} x'_{h,j}$, $M_{TT} = \sum_{j=1}^T x_j x'_j$, $M_{h,Tt}^* = \sum_{j=1}^t x_{h,j} x'_{h,j}$, and \mathcal{F} denotes some compact subset of $[0, 1]$. x_{ht} corresponds to the vector of regressors whose coefficients are under test. $\hat{s}_{h,t}$ is the restricted version of the partial sums of sample scores \hat{s}_t . More precisely, $\hat{s}_{1,t} = \hat{u}_{1t}^{+'}$, and $\hat{s}_{2,t} = x_{2t} \hat{u}_{1t}^{+'} - \hat{\Lambda}_{21}^+$. The restriction under test is that when the coefficients are partitioned into two subsets, the subset of cointegrating coefficients denoted by h is stable over time. $V_{h,Tt}$ is a natural estimator of the variance of the partial sums $S_{h,Tt}$. Note that it is also tied down at the upper end of $[0, 1]$ just like V_{Tt} (the estimator of that of the partial sums S_{Tt}).

There is another way to think of *SupF* test in terms of regression. It is the maximum of a sequence of LM tests for $B = 0$ in the regression defined as either $y_t = A_1 + B d_{1t}(i) + A_2 x_{2t} + A_3 x_{3t} + u_{1t}$ or $y_t = A_1 + B d_{2t}(i) x_{2t} + A_2 x_{2t} + A_3 x_{3t} + u_{1t}$, where $d_{ht}(i)$ are the dummy variables indexed by the break point $i (= 1, \dots, T)$.

The notion of *SupF* test can run back as far as Quandt (1960). It can be found that the timing of the shift is not conveyed to the test, which evolves from the *F* test for a structural break. The information regarding the occurrence of a break is normally incorporated into the conventional *F* test in an explicit way. In practice, note that the region \mathcal{F} cannot contain the endpoints (0 and 1) that would result in a divergent sequence when the test statistics are calculated. A reasonable suggestion by Andrews (1993) is to select $\mathcal{F} = [0.15, 0.85]$.

To detect the parameter instability of random-walk type, another two test statistics are proposed. The time-varying parameters in Models I and II are

specified as

$$A_{h,Tt} = A_{h,Tt-1} + \varepsilon_t,$$

where $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$, and $E(\varepsilon_t \varepsilon_t') = \delta^2 G_{h,t}$. $G_{h,t}$ is the covariance structure of the martingale process $A_{h,Tt}$, and \mathcal{F}_t is the sigma field generated by the process ε_t . Therefore, when the parameters are considered constant under the null, it amounts to confining the variance of the martingale process to zero ($\delta^2 = 0$). The alternative hypothesis for the tests is then $H_a: \delta^2 \neq 0$.

Associated with the martingale specification, one test is given by

$$Mean F_h = \frac{1}{T^*} \sum_{t|T \in \mathcal{T}} F_{h,Tt},$$

where $T^* = \sum_{t|T \in \mathcal{T}} 1$, and the other test is

$$L_{h,c} = \frac{1}{T} tr \left\{ M_{h,TT}^{-1} \sum_{t=1}^T S_{h,Tt} \hat{\Omega}_{1,a}^{-1} S'_{h,Tt} \right\}.$$

The covariance structures of the parameter process $G_{h,t}$ for these tests, however, are not the same. It is assumed that for $Mean F$, $G_{h,t} = [\hat{\Omega}_{1,a} \otimes V_{h,Tt}]^{-1}$, but $[\hat{\Omega}_{1,a} \otimes M_{h,TT}]^{-1}$ for L_c . In general, the choice for $G_{h,t}$ should be guided by a prior belief about the covariance structure of the martingale process, A_{Tt} . Thus, the choice for $G_{h,t}$ is not unique. Here they are so chosen for mathematical convenience in a way to provide a tractable limiting distribution that is free of nuisance parameters. This practice can find its precedent in Nyblom (1989) and Hansen (1992a). Sometimes applied researchers may wish to entertain multiple structural breaks as an alternative. The alternatives of $Mean F$ and L_c tests can in principle be generalized to include a process with multiple jumps such as those discussed in Nyblom (1989). To be able to do so, some information about the jump process on which the limit distribution of the test under consideration will depend has to be known. However, such a test would not be free of parameters relevant to this information; hence, in practice, it is difficult to apply.

Both these tests have the optimality property that the power is maximized against the alternatives close to the null hypothesis. $Mean F_h$ is the limit of the exponential LM statistics considered in Andrews and Ploberger (1994), while $L_{h,c}$ is an asymptotic approximation to the locally most powerful test for constant coefficients studied in Nyblom (1989). Note that the partial sum of the residuals, $S_{h,Tt}$, is the basic element in constructing the tests. Under the alternative that $\delta^2 \neq 0$, as shown later, it diverges at a rate faster than that under the null ($= O_p(T^{1/2})$). As a result, the consistency of these tests can be ensured by taking advantage of this non-stationarity in the data.

The method of trimming the region is also required for $Mean F$, since it suffers the same non-convergence problem as $Sup F$. On the contrary, L_c , formulated

independently of *Sup F* and *Mean F*, is able to escape from the non-convergence problem, and is also computationally easier.

It cannot be expected that all three proposed tests come to an identical conclusion when applied to a specific data set. This is because each test serves to discover different alternatives, even though they have power in similar directions. Tailoring a test for a particular application therefore hinges on the purpose of the test.

4. The limiting distributions of the test statistics

The asymptotic distributions of the proposed tests given below rely on the theory of weak convergence on the space *D*. This approach can characterize all the limiting distributions of the tests as functionals of Brownian motions under mild regularity conditions. The set of assumptions required for the distributional theory is stated below. A few words on the notations used throughout: ‘ \Rightarrow ’ stands for weak convergence of the associated probability measures, and ‘*BM*(Ω)’ denotes a Brownian motion with covariance matrix Ω .

Assumptions: Let $\{\alpha_m\}$ be the strong mixing coefficients for $\{u_t\}$, for some $q > \beta > \frac{5}{2}$,

1. $E(u_t) = 0$;
2. $\{\alpha_m\}$ are of size $-q\beta/[2(q - \beta)]$;
3. $\sup_t E|u_t|^q < \infty$;
4. Ω as defined in Eq. (3) exists with finite elements;
5. $\Omega_{aa} > 0$ and $\Omega_{1.a} > 0$;
6. $k'_t = (1, k'_{2t}) = (1, t, t^2, \dots, t^p)$;
7. $\text{rank}(\pi_2) < m_2$;
8. $M^3/T = o(1)$.

To get better insights into the weak convergence results for the tests, we will sketch the proof procedures that are spelled out in the appendix. Observe that the proposed tests are functions of partial sample score sums $S_{h,Tt}$, which in turn are functions of a variety of partial regressor moment matrices. The regressor vector consists of both deterministic and stochastic trends converging at different rates. To accommodate the different rates of convergence, the weight matrix that could appropriately standardize the regressors needs to be constructed. Specifically, we rewrite the partial sums process of sample scores as functionals on $[0, 1]$,

$$S_{hT}(\tau) = S_{h,T[\tau]} = \sum_{t=1}^{[T\tau]} \hat{s}_{h,t}$$

where ‘ $[\cdot]$ ’ denotes ‘integer part’ and $\tau \in [0, 1]$.

We also apply the same routine to the process $F_{h,T\tau}$,

$$F_{hT}(\tau) = F_{h,T[T\tau]} = \text{vec}(S_{hT}(\tau))' [\hat{\Omega}_{1,a} \otimes V_{hT}(\tau)]^{-1} \text{vec}(S_{hT}(\tau)),$$

where $V_{hT}(\tau) = V_{h,T[T\tau]} = M_{hT}(\tau) - M_{hT}^*(\tau)M_T(1)^{-1}M_{hT}^*(\tau)'$, $M_{hT}(\tau) = M_{h,T[T\tau]} = \sum_{j=1}^{[T\tau]} X_{h,j}X_{h,j}'$, $M_T(\tau) = M_{T[T\tau]} = \sum_{j=1}^{[T\tau]} X_jX_j'$, and $M_{hT}^*(\tau) = M_{h,T[T\tau]}^* = \sum_{j=1}^{[T\tau]} X_{h,j}X_j'$.

Let Δ_{hT} denote the well-defined weight matrix for the process $S_{hT}(\tau)$ whose structure is formulated in the appendix, $k(\tau) = [1', k_2'(\tau)]' = (1, \tau, \tau^2, \dots, \tau^p)'$ partitioned conformably with k_τ , and $\rho(\pi_2)$ denote the indicator function for event π_2 (takes the value of 0 if $\pi_2 = 0$, otherwise takes 1). Then, by applying this weight matrix, the asymptotic distributions of $S_{hT}(\tau)$ and $F_{hT}(\tau)$ are concluded with the following theorem.

Theorem 1. Under the null hypothesis,

1. $\frac{1}{\sqrt{T}}\Delta_{hT}S_{hT}(\tau) \Rightarrow S_h^*(\tau)\Omega_{1,a}^{1/2}$,
2. $F_{hT}(\tau) \Rightarrow F_h(\tau) \equiv \chi^2(v_h)$ for $\tau \in \mathcal{T}$,

where $h = 1, 2$; $S_h^*(\tau) = S_h(\tau) - M_h^*(\tau)M(1)^{-1}S(1)$; $F_h(\tau) = \text{tr}[S_h^*(\tau)'V_h(\tau)^{-1}S_h^*(\tau)]$; $S(\tau) = \int_0^\tau X dW_1'$; $S_h(\tau) = \int_0^\tau X_h dW_1'$; $M(\tau) = \int_0^\tau XX'$; $M_h(\tau) = \int_0^\tau X_hX_h'$; $M_h^*(\tau) = \int_0^\tau X_hX'$; $V_h(\tau) = M_h(\tau) - M_h^*(\tau)M(1)^{-1}M_h^*(\tau)'$; $X(\tau) = [1', X_a'(\tau)]' = [1', W_a'(\tau)]'$ if $\rho(\pi_2) = 0$, or otherwise $= [1', k_2'(\tau), W_a'(\tau)]'$; $v_1 = m_1$, $v_2 = m_1m_2$; and $[W_1(\tau), W_a(\tau)] \equiv [BM(I_{m_1}), BM(I_{m_2+m_3-\rho(\pi_2)p})]$, in which $W_1(\tau)$ is independent of X .

This theorem generalizes Theorem 2 of Hansen (1992a) and Lemma 5 of Shin (1994). The former gives the asymptotics of the processes $S_T(\tau)$ and $F_T(\tau)$, $S^*(\tau)$ and $F(\tau)$, under the null that the whole cointegration vector is constant. The latter shows the univariate version of the asymptotics of $S_{hT}(\tau)$, using the linear leads and lags estimator of Saikkonen (1991) and Stock and Watson (1993). It should be emphasized that Theorem 1.2 holds pointwise for any fixed τ . When $h = 1$, the result can have a much simplified version because now $X_1(\tau) = 1$ for any τ .

The functional limit to the partial sums process of sample scores $S_{hT}(\tau)$, $S_h^*(\tau)$, is a tied-down version of $S_h(\tau)$. In consequence, the covariance function of S_h^* vanishes at the upper end of $[0, 1]$. To see this, note that conditional on $\mathcal{F}_x = \sigma(X(\tau): 0 \leq r \leq 1)$ (the sigma field generated by $X(\tau)$), the covariance functions of $S^*(\tau)$ and $S_h^*(\tau)$ are expressed as

$$E[\text{vec}(S^*(\tau_1))\text{vec}(S^*(\tau_2))' | \mathcal{F}_x] = \Omega_{1,a} \otimes [M(\tau_1) - M(\tau_1)M(1)^{-1}M'(\tau_2)],$$

$$E[\text{vec}(S_h^*(\tau_1))\text{vec}(S_h^*(\tau_2))' | \mathcal{F}_x] = \Omega_{1,a} \otimes [M_h(\tau_1) - M_h^*(\tau_1)M(1)^{-1}M_h^*(\tau_2)],$$

given $\tau_1 \leq \tau_2$. The covariance function for S_h^* is zero if $\tau_1 = \tau_2 = 1$ as it does for S^* , since S_h^* is a subset of S^* . Note that when $\tau_1 = \tau_2$, the covariance function turns into the variance functions denoted by $V_h(\tau)$.

In the theorem, the degrees of freedom of the chi-square distribution represent the dimension of the regressors whose coefficients are under test. The sum of the degrees of freedom of $\chi^2(v_1)$ and $\chi^2(v_2)$ is equal to that of the functional limit of $F_T(\tau)$, $F(\tau)$. $F_T(\tau)$, using the full vector of sample scores, is also pointwise asymptotically chi-square distributed (see Theorem 2 Hansen (1992a)). This is always true because the degrees of freedom of any chi-square distribution mentioned are determined by the number of coefficients under test. However, in general $\chi^2(v_1)$ and $\chi^2(v_2)$ will not be independent. It can be shown by observing that with probability one, the off-diagonal submatrix of the conditional covariance matrix of $F(\tau)$, $V(\tau)$ ($= M(\tau) - M(\tau)M(1)^{-1}M(\tau)$), is non-zero.

Theorem 1 also shows the difference in the asymptotics of the whole trend regressor vector, X , when including the deterministic trends in regressors. For a fixed number of regressors ($= m_2 + m_3 + 1$), more variations would be introduced into X when $\pi_2 \neq 0$ than when $\pi_2 = 0$ by increasing the number of the stochastic trends (from $m_2 + m_3 - p$ to $m_2 + m_3$) across the regressors. This is because the behavior of stochastic trends is dominated by that of deterministic trends in the limit. With the convergence results above, we can now lay out descriptions for the limiting distributions of the test statistics mounted before.

Theorem 2. Under the null hypothesis,

1. $Sup F_h \Rightarrow \sup_{\tau \in \mathcal{T}} F_h(\tau)$,
2. $Mean F_h \Rightarrow \int_{\mathcal{T}} F_h(\tau) d\tau$,
3. $L_{h,c} \Rightarrow \int_0^1 tr\{S_h^*(\tau)M_h(1)^{-1}S_h^*(\tau)\}$,

where $h = 1, 2$.

The result is related to some recent works. The tests proposed by Quintos and Phillips (1993) and Shin (1994) correspond to the L_c test here. This test is an extension of the test for stationarity suggested by Kiwatkowski et al. (1992) to the context of cointegration. The tests of Shin (1994) are useful in testing for the null of *cointegration*, but might not be suitable for testing for the constancy of cointegrating coefficients. His test is equivalent to the test for intercept nonconstancy which is a special case of our subset tests. While Quintos and Phillips (1993) dealt with the case allowing for the presence of deterministic trends in $I(1)$ regressors, Shin (1994) did not. The theorem here makes such an extension to $Sup F_h$ and $Mean F_h$ tests.

The covariance structure of $vec(S_h^*(\tau))$, $\Omega_{1..a} \otimes V_h(\tau)$, has an influence on the asymptotic distributions of the tests. Due to this influence, the estimation for the covariance is expected to play a role concerning the finite-sample behavior of the tests. As displayed in the next section, the power performance of the tests appears to be affected partly by the estimation of the covariance parameters.

More importantly, Theorem 2 shows that the test statistics depend on the trend nature of the regressors and the regressors whose coefficients are under test. This is an echo of the previous finding that the trend property of the regressors makes a difference in their joint asymptotics. Thus the theorem substantiates that it is important to know the nature of the trend in X (i.e. m_2 and p) to infer properly from the test.

The limiting distributions of the test statistics are free of nuisance parameters and not standard. They do not have any practical merit until their critical values are calculated. Because the limiting representations of the suggested tests only depend on a few parameters, the critical values can be approximated well by large sample Monte Carlo simulations. Using samples of size 1000 in 15 000 replications, the critical values for the models of primary interest in applications are provided in Table 1 at three different levels (1%, 5% and 10%). Only the case for a single regressand is under consideration ($m_1 = 1$), while the controlled regressors indexed by p and m_2 are allowed to vary ($p = 0, 1$; and $m_2 = 1, 2, 3$).

An immediate application of these subset tests is to test for cointegration as Shin (1994) did. To understand this, rewrite Model I as $y_t = A_{10} + A_2x_{2t} + A_3x_{3t} + v_t$ where $v_t = u_t + \sum_{i=1}^t A_{1i}$, an I(1) process. As noted in Hansen (1992a), this is simply stating that y_t and x_{2t} are not cointegrated. In other words, *Mean F* and *L_c*, tests for intercept random variation, are also tests of H_0 : *cointegration* against H_a : *no cointegration*. Being a specification test for cointegration, *Sup F* would also possess power against the alternative of no cointegration.

It is also important to establish the consistency of the suggested tests. This is stated below.

Theorem 3. Under the alternative hypothesis,

1. $Sup F_h = O_p\left(\frac{T}{M}\right)$,
2. $Mean F_h = O_p\left(\frac{T}{M}\right)$,
3. $L_{h,c} = O_p\left(\frac{T}{M}\right)$,

where $h = 1, 2$.

The result indicates that our subset tests, regardless of type, diverge at a rate T/M under the alternative. Quintos and Phillips (1993) and Shin (1994) also proved the same rate for their Nyblom-styled tests. It should be noted that the rate, as commonly found for tests involving semiparametric corrections, depends on the bandwidth number, M . Though the result is termed in an asymptotic sense, selecting an appropriate lag number proves very important for the tests to have an adequate performance in small samples. This is the subject in the next section.

Table 1
Asymptotic critical values for L_c , Mean F , and Sup F

Usage:

1. The table is for a single equation ($m_1 = 1$). The critical values are computed by Monte Carlo simulations using samples of size 1000 in 15 000 replications.
2. m and m_2 stand for the number of stochastic trends in the whole regressor vector and in the regressors whose coefficients are under test respectively. p is the maximum power integer of the time trend in regressors. 'Model' is to indicate which model (I or II) is under consideration. $\rho(\pi_2)$ assumes 0 in the absence of deterministic trend in regressors whose coefficients are under test (i.e. $\pi_2 = 0$), otherwise 1.
3. Reject the null of constancy of the cointegrating coefficients of concern if the computed value of the statistics of interest is *greater* than the appropriate critical value. For example, given $(m,p) = (1,0)$, $m_2 = 1$, and $\rho(\pi_2) = 0$, to test if $H_0:A_1$ is constant (Model I) with L_c , one should reject at 5% level if the computed statistic is greater than 0.30.

(m,p)	Model	m_2	$\rho(\pi_2)$	(1) L_c			(2) Mean F			(3) Sup F		
				1%	5%	10%	1%	5%	10%	1%	5%	10%
(1,0)	I	0	0	0.55	0.31	0.23	4.20	2.60	2.01	13.07	9.58	7.94
(1,0)	II	1	0	0.49	0.30	0.22	4.19	2.63	2.00	12.98	9.47	7.94
(2,0)	I	0	0	0.37	0.22	0.17	3.73	2.44	1.89	13.48	10.09	8.53
(2,0)	II	1	0	0.42	0.24	0.18	3.95	2.52	1.95	13.26	9.78	8.23
(2,0)	II	2	0	0.70	0.46	0.36	6.27	4.38	3.59	16.08	12.45	10.07
(2,1)	I	0	0	0.18	0.12	0.098	3.29	2.22	1.80	14.16	10.75	9.21
(2,1)	II	1	0	0.34	0.20	0.15	3.84	2.47	1.93	13.52	9.98	8.46
(2,1)	II	1	1	0.24	0.16	0.12	3.09	2.14	1.74	14.16	10.71	9.23
(2,1)	II	2	1	0.65	0.40	0.31	6.18	4.35	3.55	16.96	13.25	11.56
(3,1)	I	0	1	0.15	0.10	0.08	3.11	2.14	1.75	14.50	11.05	9.55
(3,1)	II	1	0	0.30	0.17	0.13	3.68	2.42	1.88	13.82	10.21	8.71
(3,1)	II	1	1	0.21	0.13	0.10	2.98	2.09	1.70	14.70	11.02	9.54
(3,1)	II	2	0	0.58	0.36	0.28	6.27	4.31	3.48	16.60	12.78	11.17
(3,1)	II	2	1	0.53	0.34	0.27	5.97	4.20	3.46	17.44	13.28	11.74
(3,1)	II	3	1	0.83	0.56	0.44	8.36	5.95	5.04	19.43	15.33	13.68

5. Monte Carlo experiments

In this section, we conduct a small scale of Monte Carlo experiments to evaluate the finite-sample performance of the suggested tests. Other than size and power, we are also concerned with the issue of to what extent the presence of a non-constancy in a certain subset of coefficients contaminates the size of subset tests for the coefficients not in this subset. Ideally, it is preferred that when only this subset of coefficients varies, the power of the corresponding subset test is close to its ideal power function; meanwhile, the size of other subset tests for

those coefficients not in this subset is minimized. It is expected, however, that the analogs to the ideal situation might not be observed because of erratic behaviors of finite-sample sequences.

The experiments are based on the following simple model:

$$y_t = \mu_t + \beta_{2t}x_{1t} - \beta_{3t}x_{2t} + \varepsilon_t, \quad \varepsilon_t = \rho\varepsilon_{t-1} + \theta_t,$$

$$x_{1t} = x_{1t-1} + \eta_{1t}, \quad \eta_{1t} = \alpha_1\eta_{1t-1} + \omega_{1t},$$

$$x_{2t} = x_{2t-1} + \eta_{2t}, \quad \eta_{2t} = \alpha_2\eta_{2t-1} + \omega_{2t},$$

where y_t, x_{1t} and x_{2t} are scalars, and $\{\theta_t, \omega_{1t}, \omega_{2t}\}' \sim nid(0, \Sigma)$. The experimental design is similar to that of Gregory et al. (1996). For both size and power simulations, there are four settings of interest featured by different degrees of temporal dependence as the parameters (ρ, α_1, α_2 , and Σ) vary.

Each experiment starts with the setting at which the errors are exogenous and serially uncorrelated (Case A), then moves to the settings at which they are serially correlated (Cases B and C), and winds up at the setting at which the regressors are endogenous with the serial correlation (Case D). In each setting, the cointegration error, ε_t , is renormalized to yield a constant unit long-run variance by scaling down the variance of autoregressive error, θ_t . The settings are detailed in the tables where the simulations are also given. Under each test type suggested, five subset test statistics are applied to various collections of cointegrating coefficients. The null of each individual test under consideration is

Test	(1)	(2)	(3)	(4)	(5)
Coefficient under Test	$(\mu_t, \beta_{1t}, \beta_{2t})$	μ_t	β_{1t}	β_{2t}	(β_{1t}, β_{2t})
Null:	$(\mu_0, \beta_{10}, \beta_{20})$	μ_0	β_{10}	β_{20}	(β_{10}, β_{20})

for which $(\mu_0, \beta_{10}, \beta_{20})$ denotes the constant initials of $(\mu_t, \beta_{1t}, \beta_{2t})$ having the value of $(1, 3, -1)$. The criterion is the rejection frequencies at 5% significance level using critical values from Table 1 (for the subset tests) and Tables 1–3 indexed by $m_2 = 2$ and $p = 0$ in Hansen (1992a) (for the joint test) in 5000 replications. The sample size is 100, roughly corresponding to typical macroeconomic time series and our application. The covariance is estimated using a Bartlett kernel with M set at 4 as recommended by Andrews (1991) (Table 1). Later we will discuss the effect of choosing other bandwidth numbers.

5.1. Size and power

In Table 2, we first display the result of a size comparison. The general impression from simulations is that regardless of the degree of dependence, the size for most of the subset tests seems to be close to the asymptotic values. It suggests that there is little need to size correct the power calculations for these

Table 2
Size of the tests

$$y_t = 1 + 3x_{1t} - x_{2t} + \varepsilon_t, \varepsilon_t = \rho\varepsilon_{t-1} + \theta_t,$$

$$x_{1t} = x_{1t-1} + \eta_{1t}, \eta_{1t} = \alpha_1\eta_{1t-1} + \omega_{1t},$$

$$x_{2t} = x_{2t-1} + \eta_{2t}, \eta_{2t} = \alpha_2\eta_{2t-1} + \omega_{2t},$$

$$\{\theta_t, \omega_{1t}, \omega_{2t}\}' \sim NID(0, \Sigma).$$

	ρ	α_1	α_2	Σ
Case A	0.0	0.0	0.0	I_3
Case B	0.5	0.0	0.0	$\sigma_{11} = 0.25$
Case C	0.5	0.0	0.0	$\sigma_{11} = 0.25 \sigma_{12} = 0.3 \sigma_{13} = 0.4 \sigma_{23} = 0.5$
Case D	0.5	0.6	0.7	$\sigma_{11} = 0.25 \sigma_{12} = 0.3 \sigma_{13} = 0.4 \sigma_{23} = 0.5$

	Case A	Case B	Case C	Case D
<i>L_c</i>				
(1)	0.05	0.08	0.08	0.07
(2)	0.03	0.06	0.07	0.05
(3)	0.05	0.07	0.07	0.06
(4)	0.04	0.07	0.07	0.05
(5)	0.05	0.07	0.07	0.06
<i>Mean F</i>				
(1)	0.04	0.05	0.05	0.07
(2)	0.02	0.06	0.06	0.05
(3)	0.04	0.07	0.06	0.07
(4)	0.04	0.06	0.07	0.07
(5)	0.04	0.07	0.07	0.07
<i>Sup F</i>				
(1)	0.03	0.01	0.01	0.02
(2)	0.01	0.02	0.02	0.03
(3)	0.04	0.03	0.03	0.04
(4)	0.04	0.03	0.04	0.05
(5)	0.05	0.03	0.03	0.05

Note:

1. Rejection frequencies are calculated at the 5% level of significance using critical values from Tables 1–3 in Hansen (1992a) (indexed by $(m_2, p) = (2, 0)$) for the joint test, and Table 1 in the text (indexed by $(m, p) = (2, 0)$) for the subset tests in 5000 replications.
2. The numbers in the leftmost column refer to the cointegrating coefficients under test: (1) the whole cointegration vector; (2) the intercept; (3) the coefficient on x_{1t} ; (4) the coefficient on x_{2t} ; and (5) the coefficients on x_{1t} and x_{2t} .

subset tests when sample size is moderate. However, there are a number of instances of overrejection for the subset tests of *Mean F* and *Sup F* type. In contrast, underrejection is more likely to occur for the subset tests of *Sup F* type.

Table 3
Power of the tests: structural break

$$y_t = \mu_t + 3x_{1t} - \beta_{2t}x_{2t} + \varepsilon_t,$$

$$\left\{ \begin{matrix} \mu_t = 1, & t \leq [\tau T] \\ \mu_t = 2, & t > [\tau T] \end{matrix} \right\} \text{ or } \left\{ \begin{matrix} \beta_{2t} = -1, & t \leq [\tau T] \\ \beta_{2t} = -1.4, & t > [\tau T] \end{matrix} \right\}, \varepsilon_t = \rho\varepsilon_{t-1} + \theta_t,$$

$$x_{1t} = x_{1t-1} + \eta_{1t}, \eta_{1t} = \alpha_1\eta_{1t-1} + \omega_{1t},$$

$$x_{2t} = x_{2t-1} + \eta_{2t}, \eta_{2t} = \alpha_2\eta_{2t-1} + \omega_{2t},$$

$$\{\theta_t, \omega_{1t}, \omega_{2t}\}' \sim NID(0, \Sigma).$$

	ρ	α_1	α_2	Σ
Case A	0.0	0.0	0.0	I_3
Case B	0.5	0.0	0.0	I_3
Case C	0.5	0.0	0.0	$\sigma_{11} = 0.25 \sigma_{12} = 0.3 \sigma_{13} = 0.4 \sigma_{23} = 0.5$
Case D	0.5	0.6	0.7	$\sigma_{11} = 0.25 \sigma_{12} = 0.3 \sigma_{13} = 0.4 \sigma_{23} = 0.5$

τ	Case A			β_{2t}			Case B			β_{2t}		
	μ_t						μ_t					
	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
L_c												
(1)	0.26	0.24	0.26	0.29	0.56	0.62	0.36	0.30	0.36	0.35	0.63	0.68
(2)	0.28	0.34	0.30	0.20	0.37	0.45	0.37	0.42	0.38	0.24	0.41	0.50
(3)	0.13	0.21	0.26	0.14	0.28	0.43	0.19	0.26	0.33	0.16	0.32	0.45
(4)	0.13	0.20	0.26	0.31	0.68	0.72	0.18	0.26	0.33	0.37	0.73	0.76
(5)	0.16	0.21	0.25	0.32	0.60	0.65	0.22	0.26	0.34	0.38	0.65	0.69
Mean F												
(1)	0.20	0.18	0.21	0.31	0.61	0.62	0.29	0.25	0.30	0.39	0.69	0.67
(2)	0.31	0.37	0.33	0.23	0.48	0.57	0.43	0.49	0.46	0.28	0.54	0.62
(3)	0.16	0.20	0.23	0.20	0.33	0.44	0.23	0.27	0.31	0.24	0.39	0.46
(4)	0.17	0.20	0.23	0.56	0.86	0.79	0.23	0.27	0.31	0.66	0.91	0.85
(5)	0.18	0.19	0.21	0.45	0.74	0.66	0.25	0.25	0.29	0.55	0.81	0.73
Sup F												
(1)	0.06	0.05	0.06	0.37	0.59	0.63	0.08	0.07	0.08	0.48	0.72	0.71
(2)	0.22	0.23	0.20	0.22	0.42	0.60	0.33	0.34	0.32	0.28	0.47	0.63
(3)	0.13	0.13	0.15	0.23	0.29	0.43	0.16	0.17	0.22	0.27	0.33	0.45
(4)	0.12	0.12	0.15	0.68	0.87	0.85	0.15	0.17	0.22	0.77	0.93	0.90
(5)	0.11	0.10	0.11	0.57	0.77	0.76	0.12	0.12	0.16	0.68	0.86	0.83
	Case C			β_{2t}			Case D			β_{2t}		
	μ_t						μ_t					
L_c												
(1)	0.35	0.28	0.33	0.32	0.60	0.65	0.32	0.25	0.30	0.40	0.67	0.72
(2)	0.35	0.39	0.36	0.22	0.39	0.49	0.30	0.33	0.30	0.26	0.45	0.55
(3)	0.17	0.25	0.31	0.18	0.38	0.49	0.15	0.22	0.27	0.23	0.43	0.54
(4)	0.17	0.24	0.31	0.34	0.71	0.74	0.14	0.21	0.27	0.35	0.73	0.77
(5)	0.21	0.25	0.31	0.35	0.62	0.62	0.19	0.22	0.27	0.41	0.68	0.72

Table 3 continued

τ	Case A			Case B								
	μ_t			β_{2t}			μ_t			β_{2t}		
	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
<i>Mean F</i>												
(1)	0.28	0.23	0.27	0.35	0.65	0.66	0.31	0.24	0.29	0.55	0.79	0.81
(2)	<u>0.41</u>	<u>0.45</u>	<u>0.42</u>	0.25	0.52	0.61	<u>0.39</u>	<u>0.41</u>	<u>0.39</u>	0.35	0.67	0.73
(3)	<u>0.21</u>	<u>0.26</u>	<u>0.29</u>	0.28	0.46	0.50	<u>0.23</u>	<u>0.25</u>	<u>0.28</u>	0.45	0.58	0.61
(4)	0.21	0.26	0.30	<u>0.61</u>	<u>0.89</u>	<u>0.81</u>	0.22	0.25	0.29	<u>0.81</u>	<u>0.94</u>	<u>0.90</u>
(5)	0.22	0.24	0.27	0.50	<u>0.77</u>	<u>0.70</u>	0.26	0.25	0.29	<u>0.73</u>	<u>0.87</u>	<u>0.81</u>
<i>Sup F</i>												
(1)	0.07	0.06	0.07	0.39	0.63	0.64	0.07	0.07	0.07	0.88	0.94	0.94
(2)	<u>0.29</u>	<u>0.30</u>	<u>0.27</u>	0.24	0.44	0.61	<u>0.26</u>	<u>0.26</u>	<u>0.26</u>	0.47	0.62	0.79
(3)	<u>0.13</u>	<u>0.16</u>	<u>0.20</u>	0.31	0.41	0.49	<u>0.17</u>	<u>0.16</u>	<u>0.19</u>	0.58	0.57	0.64
(4)	0.13	0.15	0.19	<u>0.71</u>	<u>0.88</u>	<u>0.85</u>	0.16	0.17	0.21	<u>0.97</u>	<u>0.99</u>	<u>0.99</u>
(5)	0.10	0.11	0.13	0.60	0.79	0.76	0.14	0.12	0.14	0.95	0.98	0.98

Note:

1. See note to Table 2.
2. The numbers underlined are empirical powers of the specific subset tests for changes in the coefficients considered in the experimentation.

Two forms of variation, jump or random-walk, are studied for the powers of the tests. In Table 3, we illustrate the power of the subset tests under two experiments: a structural break occurs in the intercept (μ_t), or in the slope on the second stochastic trend (β_{2t}). Under the alternative the coefficient processes subject to break are specified as

$$\text{Coefficient} = \begin{cases} \mu_t & \beta_{2t} \\ \mu_0 & \beta_{20} & t \leq [T\tau] \\ \mu_0 + \Delta_1 & \beta_{20} + \Delta_2 & t > [T\tau], \end{cases}$$

where $\Delta = (\Delta_1, \Delta_2)$ is the jump of size $(1, -0.4)$, and τ is the break point which takes place at 0.25, 0.50 and 0.75, roughly representing the beginning, middle and end of the sample. By varying τ , we will examine whether the power of the tests is independent of the break point, given that our suggested tests assume arbitrary breaks.

Our proposed tests perform well. In the presence of a slope jump, the tests appropriate for this context are those labelled (1)-joint test, (4)-test for β_{2t} , (5)-test for β_{1t}, β_{2t} . But the tests having the best power to discover a slope jump are Test (4) of *Sup F* and *Mean F* types. These are the specific subset tests mainly designed to detect non constancy in this form. Test (2) of each type, relative to others, also performs the best in the presence of an intercept jump. Indeed, for

$\tau = 0.50$, when a slope jumps, Test (4) rejects in 86–99% of the trials, and when an intercept jumps, Test (2) rejects in 23–49% of the trials. We note that Test (1) of *Sup F* type, a joint test, appears to have trouble dealing with a jump in intercept, with a rejection frequency of only 5–8%. This suggests a merit of using our subset tests when only part of the cointegration coefficients is not constant.

In addition, the power of the subset tests is little affected by the degree of dependence. In particular, the rejection frequency of the tests remains more or less unchanged within Cases B–D. But there is a power loss by approximately 5–10%, moving from the case of mild dependence (Case B) to the case of no dependence (Case A). This is the adverse effect of selecting too large a bandwidth number. Also the subset tests tend to reject more often when the intercept or slope jumps at the middle or end of the sample.

Table 4 reports the capability of detecting a random variation by the subset tests. The non-constancy here takes the form of a random walk,

$$\begin{aligned} \text{Coefficient} \quad & \mu_t \quad \beta_{2t} \\ = \quad & \mu_{t-1} + \delta_{1t} \quad \beta_{2t-1} + \delta_{2t}, \end{aligned}$$

in which $\delta_{it} \sim \text{nid}(1, v^2)$ for $i = 1, 2$. We set $v = 0.2, 0.05$ for current experiments.

The subset tests applicable to detect a random intercept or slope also perform reasonably well. Typically the specific tests of *Mean F* and *L_c* types have greater power than others. The rejection rate is 38–50% in the presence of a random intercept, and 54–67% in the presence of a random slope. Again, similar to the results with a structural shift, the degree of dependence does not much impair the power of the tests.

Unfortunately, there is a problem with size contamination. For instance, when a slope jump or a random slope is present, Test (2) has a rejection rate more than half of that of Test (4). These are the rejections under the null of a constant intercept, thus indicating a sizable distortion. It implies that the subset test for some coefficient variations, because of the presence of other coefficient variations, has a tendency to spuriously reject. Although when employing these subset tests for inference one needs to exercise care, spurious rejection is not as bad as its name implies. It is noted that, in the presence of a change in a certain coefficient, the rejection rates are not often lower than 25% across subset tests. Thus, a reasonable view of this phenomenon is that spurious rejection, with the good power property of the subset tests, is potentially directing toward the presence of some sort of coefficient variations in the data. In other words, when applying our tests to various subsets of cointegrating coefficients, rejection by most of the statistics, possibly as a result of spurious rejection, simply suggests that the changes in a certain part of cointegrating coefficients may be significant, implying a misspecification in the regression. We

Table 4
Power of the tests: random variation

$$y_t = \mu_t + 3x_{1t} - \beta_t x_{2t} + \varepsilon_t, \varepsilon_t = \rho \varepsilon_{t-1} + \theta_t,$$

$$\mu_t = \mu_{t-1} + \delta_{0t}, \delta_{0t} \sim NID(0, 0.04), \delta_{0t} = 1, \text{ or}$$

$$\beta_{2t} = \beta_{2t-1} + \delta_{2t}, \delta_{2t} \sim NID(0, 0.0025), \delta_{20} = -1$$

$$x_{1t} = x_{1t-1} + \eta_{1t}, \eta_{1t} = \alpha_1 \eta_{1t-1} + \omega_{1t},$$

$$x_{2t} = x_{2t-1} + \eta_{2t}, \eta_{2t} = \alpha_2 \eta_{2t-1} + \omega_{2t},$$

$$\{\theta_t, \omega_{1t}, \omega_{2t}\}' \sim NID(0, \Sigma).$$

	ρ	α_1	α_2	Σ
Case A	0.0	0.0	0.0	I_3
Case B	0.5	0.0	0.0	$\sigma_{11} = 0.25$
Case C	0.5	0.0	0.0	$\sigma_{11} = 0.25 \sigma_{12} = 0.3 \sigma_{13} = 0.4 \sigma_{23} = 0.5$
Case D	0.5	0.6	0.7	$\sigma_{11} = 0.25 \sigma_{12} = 0.3 \sigma_{13} = 0.4 \sigma_{23} = 0.5$

	Case A		Case B		Case C		Case D	
	μ_t	β_{2t}	μ_t	β_{2t}	μ_t	β_{2t}	μ_t	β_{2t}
<i>L_c</i>								
(1)	0.37	0.46	0.43	0.50	0.41	0.49	0.39	0.54
(2)	<u>0.39</u>	0.31	<u>0.45</u>	0.34	<u>0.42</u>	0.33	<u>0.38</u>	0.38
(3)	0.28	0.29	0.31	0.31	0.29	0.34	0.27	0.40
(4)	0.28	<u>0.54</u>	0.31	<u>0.59</u>	0.30	<u>0.57</u>	0.27	<u>0.60</u>
(5)	0.29	<u>0.48</u>	0.34	<u>0.52</u>	0.31	<u>0.50</u>	0.31	<u>0.56</u>
<i>Mean F</i>								
(1)	0.31	0.45	0.37	0.50	0.36	0.47	0.40	0.60
(2)	<u>0.43</u>	0.36	<u>0.50</u>	0.41	<u>0.48</u>	0.40	<u>0.47</u>	0.51
(3)	0.28	0.28	0.32	0.31	0.31	0.35	0.33	0.45
(4)	0.29	<u>0.56</u>	0.32	<u>0.62</u>	0.31	<u>0.59</u>	0.32	<u>0.67</u>
(5)	0.29	<u>0.49</u>	0.33	<u>0.54</u>	0.32	<u>0.52</u>	0.37	<u>0.62</u>
<i>Sup F</i>								
(1)	0.10	0.24	0.11	0.27	0.10	0.23	0.12	0.36
(2)	<u>0.28</u>	0.24	<u>0.32</u>	0.27	<u>0.29</u>	0.26	<u>0.29</u>	0.35
(3)	0.17	0.19	0.20	0.22	0.18	0.26	0.21	0.35
(4)	0.18	<u>0.48</u>	0.20	<u>0.52</u>	0.18	<u>0.49</u>	0.20	<u>0.60</u>
(5)	0.14	<u>0.36</u>	0.15	<u>0.40</u>	0.14	<u>0.36</u>	0.16	<u>0.49</u>

Note: See note to Table 3.

conclude that spurious rejection has a bearing on the merits of using our tests as a specification test in the regression with nonstationary processes.

5.2. Robustness check

The simulation results just presented depend on the choice of a bandwidth parameter. The optimal choice of M for the experiments under Case A is

actually 0 rather than 5. We thus have found a power loss to the tests due to using too large a lag window. On the other hand, selecting too small a number risks a size distortion when autocorrelation is not well accounted for in the data. It is natural to ask how robust our results are to alternative bandwidth numbers. As an attempt to answer this question, we examined the size and power of the tests with another two bandwidth values, 0 and 12, following Kiwatkowski et al. (1992) and other recent simulations. The latter is considered an extreme choice and roughly gives an upper bound for the power loss. As expected, this produces a power loss but an insignificant size distortion for the tests. More precisely, in the presence of an intercept or a slope jump, it reduces the power of our subset tests by approximately 20–35%. In the presence of a random intercept or slope, the loss is approximately 30%. While having an unfavorable effect on the power of the tests, a too-large lag substantially mitigates the problem with spurious rejection. We found that the subset tests for a certain coefficient now very often reject at only 10% or less of the trials when another coefficient is not constant. This is as anticipated from Theorem 3 where we can see that the stochastic order of the tests under the alternative will be decreased by an increase in M . In contrast, at $M = 0$, when each coefficient is constant, the tests usually are found to have a rejection rate of more than 30% within each setting, except Case A. This indicates a severe overrejection. The Monte Carlo evidence here suggests that it would be valuable to use more than one of the consistent long-run covariance estimators when applying the tests.

Also, we experimented with the lag $M = 4$ and the smaller sample size $T = 50$. All of the tests now have a much worse power performance; nevertheless, their sizes are approximately correct. Presumably, this is a reflection of the consistency of the tests.

6. An application

Stock and Watson (1993) and Gregory et al. (1994) recently studied whether there is a stable money demand relation spanning pre- and postwar periods. The latter focused on testing for constancy of the whole cointegration vector by applying the joint test of Hansen (1992a). Our previous simulations suggest that the power of a joint test might not be good when only a subset of cointegrating coefficients varies over time. Stock and Watson (1993), on the other hand, employed a classical Chow test for the null of no break in the income elasticity or in the interest semielasticity. This test, however, is crippled by the need to specify an ad hoc timing for the break point. As an illustrative example, we use the suggested subset tests to avoid these drawbacks in this empirical exercise. Consistent with those found by Stock and Watson (1993), the notion of a stable money demand appears to be strengthened by the evidence presented here.

Table 5

Testing for the stability in U.S. money demand (Annual data) $\ln(m_t) - \ln(p_t) = \mu + \beta_1 \ln(y_t) + \beta_2 r_t + e_t$.

	1901–1985		1901–1945		1946–1985	
	Estimate	Std err.	Estimate	Std err.	Estimate	Std err.
μ	1.822	0.128	2.127	0.351	5.669	0.300
β_1	0.982	0.028	0.906	0.075	0.208	0.059
β_2	-0.099	0.007	-0.103	0.014	-0.017	0.006
	Test		Test		Test	5% c.v.
L_c						
(1)	0.500		0.398		0.201	0.78
(2)	0.058		0.083		0.084	0.12
(3)	0.055		0.087		0.090	0.16
(4)	0.117		0.111		0.126	0.20
(5)	0.421*		0.277		0.185	0.40
<i>Mean F</i>						
(1)	6.920		12.961*		3.003	7.69
(2)	1.285		1.220		1.684	2.22
(3)	1.301		1.319		1.671	2.14
(4)	2.820*		1.337		1.777	2.47
(5)	5.237*		7.032*		2.576	4.35
<i>Sup F</i>						
(1)	11.677		53.727*		6.008	17.3
(2)	6.961		5.065		5.153	10.75
(3)	6.848		5.313		5.062	10.71
(4)	6.833		5.573		5.310	9.98
(5)	9.598		32.378*		6.094	13.25

Note:

1. The test statistics are obtained using a Bartlett kernel with bandwidth set at 5.
2. The annual data are from Lucas (1988). m is $M1$, p the implicit price deflator, y real net national product and r the six-month commercial paper rate.
3. The numbers in the leftmost column refer to the cointegrating coefficients under test: (1) the whole cointegration vector; (2) the intercept; (3) the income elasticity; (4) the interest semielasticity; and (5) both the income elasticity and the interest semielasticity.
4. The starred test statistics are significant at 5% level.
5. Based on the specification that $\ln(y_t)$ is an $I(1)$ with drift and r_t is an $I(1)$ without drift, the critical values of test statistics are from Tables 1–3 in Hansen (1992a) (indexed by $(m_2, p) = (1, 1)$) for the joint test, and Table 1 in the text (indexed by $(m, p) = (2, 1)$) for the subset tests at 5% level.

The long-run money demand relation with constant coefficients is specified as

$$\ln(m_t) - \ln(p_t) = \mu + \beta_1 \ln(y_t) + \beta_2 r_t + e_t,$$

where m is $M1$, p is the implicit price deflator, y is real-net national product and r is the six-month commercial paper rate. Following Stock and Watson, $\ln(y_t)$ is characterized as an $I(1)$ with drift, and r_t an $I(1)$ without drift. In Table 5, we

report the estimations under the null of parameter constancy, and the test statistics for the constancy of different subsets of cointegrating coefficients, using annual data (1901–1985) from Lucas (1988). The data set is also used in Stock and Watson where results for the fully modified estimator were obtained using a Bartlett kernel with five lags. Here we adopt this procedure for a comparison.

Using the full sample, the fully modified regression gives estimates close to those reported in the literature. For example, the income elasticity is not significantly different from 1. With most of the p -values greater than 0.05, the test statistics seem to support the null that each individual coefficient is constant, and thus the money demand has a stable cointegrating relation. Recall that the test for intercept instability is also a test for the null of cointegration. Rejection of the alternative of intercept instability also suggests that there be a cointegration relation among real balance, real income, and interest rate, over the entire period.

To permit a further comparison with Stock and Watson (1993), prewar and postwar estimates are computed as well. Apparently, the postwar estimates differ from the prewar ones in having a smaller income elasticity, while the latter exhibit comparable magnitude with the full-sample estimates. This difference raises the question of whether there has been a regime shift in the long-run money demand relation. Inspection of the test statistics for each subsample, however, yields no sufficient evidence against the hypothesis of no breaks in money demand. Specifically, the subset tests of any type fail to reject the null of constancy of each single cointegrating coefficient for each subsample, with the exception that, for the prewar period, the null is rejected at 5% for the subset tests of *Mean F* and *Sup F* types that allow for both elasticities to vary. It should be noted that intercept and income elasticity for the postwar data are not estimated as precisely as those for the full-sample, with twice bigger standard errors. This piece of information, along with non-rejection of the cointegration null by the subset test for the intercept for each subsample, is indicative of the possibility that the prewar series cointegrate the postwar ones with a common long-run vector. Overall, the evidence is not much at odds with the existence of a stable long-run money demand.

7. Concluding remarks

This paper began with two goals. The first was to develop the limiting distributions of three test statistics for detecting the non-constancy of the subset of cointegrating coefficients. Under mild regularity conditions, it has been shown that the limiting representations of the proposed tests can be characterized as functions of continuous martingales depending on the nature of the regressor process.

The second, and more practical objective of the paper was to address the question of how good it is for the finite-sample performance of the tests. The answer to this is related to the previous theoretical findings. The asymptotics of the tests share an essential ingredient in their constructions: the partial score sums process. This generality, although a desirable property in one way, is a drawback in the other. The large-sample approximation to the exact distribution sometimes may not be quite accurate. This is particularly true for the subset tests in the context where an inappropriate bandwidth number is chosen. On the other hand, the size contamination problem appears unavoidable in some cases. This is another facet of the problem of slow convergence to the limiting distribution just mentioned. Caution is required when drawing inferences from the tests. Despite these caveats, our Monte Carlo results did lend support to the finite-sample performance of the proposed subset tests. In particular, for small sample size, the subset tests are found to dominate the joint test with good power against the alternative that only partial parameters are time-varying.

Our Monte Carlo experiments open a question that bears further investigations. The subset tests may experience a power loss or size distortion when the lag window is not optimally selected along the lines of Andrews (1991). This problem would happen to those developed by Quintos and Phillips (1993) and Shin (1994) as well. The estimation of the long-run covariance matrix has proven to be a key factor in affecting size and power of the subset tests. If the test is to perform adequately, it is helpful to have a means of estimation independent of the bandwidth number for the long-run covariance. The recently developed spectral estimation procedure by Den Haan and Levin (1996) and Perron and Ng (1996) should be appealing solutions to this problem.

Lastly, rejections by the proposed tests here suggest a misspecification in the model or a time-varying cointegrating relationship between variables. It is a reasonable first attempt to search for the available omitted variables. Comparing post-sample performances, which is complementary to tests for constancy, at this stage may help developments of a satisfactory specification. This implies that some practical problems ought to be taken up. Should the possibility of time-varying cointegration need to be considered, allowing for a structural break or a random variation is likely to be a sensible and parsimonious modeling strategy. Yet some technical issues need to be studied. For example, Bai (1996) and Bai et al. (1997) began to address estimation of a break point in cointegrating regressions. Still, much work in that direction remains to be done.

8. For further reading

Andrews (1990), Andrews and Monohan (1992), Gregory and Hansen (1996), Hansen (1992c), Johansen (1991), Phillips (1987), Phillips (1991b), Phillips and Durlauf (1986).

Acknowledgements

I am deeply indebted to an editor, an associate editor, three anonymous referees and Masao Ogaki for helpful comments and suggestions, and especially to Bruce Hansen for providing enormous advice while a previous version was written under his supervision as a part of my doctoral thesis at the University of Rochester. I also wish to thank Tim Lane and the associate editor, who went beyond the call of duty, for offering numerous stylistic and grammatical recommendations which greatly enhance the readability of the paper. The financial support from the Chiang Ching-Kuo Foundation through the dissertation fellowship is also gratefully acknowledged.

Appendix A. Mathematical proofs

We will prove Theorems 1–3 for the case allowing for parameter change in stochastic trend regressors ($h = 2$) rigorously. The proof of other case ($h = 1$) can be followed similarly with minor modifications, and hence is omitted here.

A.1. Weight matrix and moment matrix

Define the cumulative process for the innovation vector, u_t ,

$$Y_t = \sum_{i=1}^t u_i.$$

Our assumptions are sufficient for the following multivariate invariance principle:

$$T^{-1/2} Y_{[T\tau]} \Rightarrow B(\tau) \equiv BM(\Omega) \quad (\text{A.1})$$

which has been shown to hold by Herrndorf (1984).

The weight matrices are defined for the purpose of reconciling with the various regressors converging at different rates. For the sequence of deterministic trend k_t , the weight matrix is $\delta_T = \text{diag}(1, T^{-1}, T^{-2}, \dots, T^{-p})$. Also, let $k(\tau) = (1, \tau, \tau^2, \dots, \tau^p)'$, thus we have $\delta_T k_{[T\tau]} \Rightarrow k(\tau)$ as $T \rightarrow \infty$ uniformly in τ . For k_{2t} , simply let $\delta_{2T} = \text{diag}(T^{-1}, T^{-2}, \dots, T^{-p})$.

A little complication would be involved in designing the weight matrix for x_{2t} . Recall that x_{2t} in Eq. (2) is driven by a constant term, k_{2t} and x_{2t}^0 . Because k_{1t} is included in the level regression, k_{2t} and x_{2t}^0 will remain after least squares project x_{2t} onto k_{1t} orthogonally. Due to that the behavior of stochastic trend is dominated by that of deterministic trend asymptotically, separating the effect of

the former from the latter is necessary. For this, if $\pi_2 \neq 0$, define

$$\Delta_{2T} = \begin{bmatrix} \delta_{2T}(\pi_2' \pi_2)^{-1} \pi_2' \\ 1/\sqrt{T}(\pi_2^{*'} \Omega_{22} \pi_2^*)^{-1/2} \pi_2^{*'} \end{bmatrix},$$

where π_2^* , of dimension $m_2 \times (m_2 - p)$, is constructed in the null space of π_2 . Therefore,

$$\Delta_{2T} x_{2t} = \begin{bmatrix} \delta_{2T}(\pi_2' \pi_2)^{-1} \pi_2'(\pi_1 + x_{2t}^o) + \delta_{2T} k_{2t} \\ 1/\sqrt{T}(\pi_2^{*'} \Omega_{22} \pi_2^*)^{-1/2} \pi_2^{*'} x_{2t}^o \end{bmatrix}.$$

Note that since $x_{2[T\tau]}^o = O_p(T^{1/2})$, $\delta_{2T}(\pi_2' \pi_2)^{-1} \pi_2'(\pi_1 + x_{2[T\tau]}^o) = o_p(1)$. Then together with Eq. (A.1)

$$\Delta_{2T} x_{2[T\tau]} \Rightarrow \begin{bmatrix} k_2(\tau) \\ W_2(\tau) \end{bmatrix} \begin{matrix} p \\ m_2 - p \end{matrix} = X_2(\tau), \tag{A.2}$$

where $W_2(\tau) = (\pi_2^{*'} \Omega_{22} \pi_2^*)^{-1/2} \pi_2^{*'} B_2(\tau) \equiv BM(I_{m_2 - p})$. If $\pi_2 = 0$, let $\Delta_{2T} = (1/\sqrt{T})\Omega_{22}^{-1/2}$. In such case, $\Delta_{2T} x_{2[T\tau]} \Rightarrow W_2(\tau)$, where $W_2(\tau) = \Omega_{22}^{-1/2} B_2(\tau) \equiv BM(I_{m_2})$.

Next, if $\psi_2' = [\pi_2', \phi_2'] \neq 0$, first define

$$\Delta_{aT} = \begin{bmatrix} \delta_{2T}(\psi_2' \psi_2)^{-1} \psi_2' \\ 1/\sqrt{T}(\psi_2^{*'} \Omega_{aa} \psi_2^*)^{-1/2} \psi_2^{*'} \end{bmatrix},$$

where ψ_2^* , of dimension $(m_2 + m_3) \times (m_2 + m_3 - p)$, is constructed in the null space of ψ_2 . Then the weight matrix for x_t is

$$\Delta_T = \begin{bmatrix} 1 & 0 \\ 0 & \Delta_{aT} \end{bmatrix}. \tag{A.3}$$

Now, that Δ_T is a $(m_2 + m_3) \times (m_2 + m_3)$ square matrix,

$$\Delta_T x_{[T\tau]} \Rightarrow \begin{bmatrix} 1 \\ k_2(\tau) \\ W_a(\tau) \end{bmatrix} \begin{matrix} 1 \\ p \\ m_2 + m_3 - p \end{matrix} = X(\tau). \tag{A.4}$$

Otherwise, define $\Delta_{aT} = (1/\sqrt{T})\Omega_{aa}^{-1/2}$, so $\Delta_T x_{[T\tau]} \Rightarrow [1', W_a'(\tau)]'$, where $W_a(\tau) \equiv BM(I_{m_2 + m_3})$.

By Eq. (A.2), Eq. (A.4), and the continuous mapping theorem (CMT, see Billingsley, 1968, Theorem 5.1), we then have

$$\begin{aligned} \frac{1}{T} \Delta_T M_T(\tau) \Delta_T' &\Rightarrow M(\tau) = \int_0^\tau X X', \\ \frac{1}{T} \Delta_{2T} M_{2T}(\tau) \Delta_{2T}' &\Rightarrow M_2(\tau) = \int_0^\tau X_2 X_2', \\ \frac{1}{T} \Delta_{2T} M_{2T}^*(\tau) \Delta_{2T}' &\Rightarrow M_2^*(\tau) = \int_0^\tau X_2 X' \end{aligned}$$

and

$$\frac{1}{T} \Delta_{2T} V_{2T}(\tau) \Delta'_{2T} \Rightarrow V_2(\tau) = M_2(\tau) - M_2^*(\tau) M(1)^{-1} M_2^*(\tau)'$$

Note that $\int_0^\tau X X' > 0$ and $\int_0^\tau X_2 X_2' > 0$ for all $\tau > 0$ almost surely (Phillips and Hansen, 1990, Lemma A.2).

A.2. Some preliminary results.

A number of large sample behaviors of various partial sums processes will be studied in this section. Note first that our assumptions also lead to

$$\frac{1}{T} \sum_{t=1}^{[T\tau]} Y_t u'_{t+1} \Rightarrow \int_0^\tau B dB' + \tau A, \tag{A.5}$$

as well as

$$\hat{\Omega} \rightarrow_p \Omega, \quad \hat{\Lambda} \rightarrow_p A. \tag{A.6}$$

The proof of convergence to matrix stochastic integral in Eq. (A.5) can be found in Hansen (1992d). The consistency of covariance parameters estimation was documented by Hansen (1992b).

Next, define $u_{1t}^+ = u_{1t} - \Omega_{1a} \Omega_{aa}^{-1} u_{at}$. Then if $\pi_2 \neq 0$, by Eqs. (A.2) and (A.5)

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{[T\tau]} u_{1t}^+ x'_{2t} \Delta'_{2T} &= \left[\frac{1}{T} \sum_{t=1}^{[T\tau]} u_{1t}^+ k'_{2t} \delta_{2t} + o_p(1), \frac{1}{T} \sum_{t=1}^{[T\tau]} u_{1t}^+ x'_{2t} \pi_2^* (\pi_2^* \Omega_{22} \pi_2^*)^{-1} \right] \Rightarrow \\ &\left[\left(\int_0^\tau dB_{1.a} k'_{2t} \left(\int_0^\tau dB_{1.a} B'_{2t} + \tau A_{21}^+ \right) \pi_2^* (\pi_2^* \Omega_{22} \pi_2^*)^{-1} \right) \right] = \\ &\left[\left(\int_0^\tau dB_{1.a} k'_{2t} \left(\int_0^\tau dB_{1.a} W'_{2t} + \tau A_{21}^* \right) \right) \right] = \int_0^\tau dB_{1.a} X'_2 + \tau(0, A_{21}^*), \end{aligned} \tag{A.7}$$

where $A_{21}^* = (\pi_2^* \Omega_{22} \pi_2^*)^{-1/2} \pi_2^* A_{21}^+$, and $B_{1.a} \equiv BM(\Omega_{1.a})$. And, by (A.6)

$$\sqrt{T} \hat{\Lambda}'_{21} \Delta'_{2T} = \hat{\Lambda}'_{21} \sqrt{T} \left[\pi_2 (\pi_2' \pi_2)^{-1} \delta'_{2T}, \frac{1}{T} \pi_2^* (\pi_2^* \Omega_{22} \pi_2^*)^{-1/2} \right] \rightarrow_p (0, A_{21}^*), \tag{A.8}$$

where $\hat{\Lambda}_{21}^+$ is the first element in $\hat{\Lambda}_{a1}^+$. If $\pi_2 = 0$, the above procedure follows with $\pi_2^* (\pi_2^* \Omega_{22} \pi_2^*)^{-1/2}$ replaced by $\Omega_{22}^{-1/2}$. Last, also note that

$$\frac{1}{\sqrt{T}} \Delta_{2T} \sum_{t=1}^{[T\tau]} [x_{2t} (u'_{at} - \hat{u}'_{at})] = \frac{1}{T} \Delta_{2T} \sum_{t=1}^{[T\tau]} x_{2t} \Delta k'_{2t} \delta'_{2T} \delta_{2T}^{-1} (\hat{\psi}_2 - \psi_2)' \sqrt{T} = o_p(1), \tag{A.9}$$

where $\hat{\psi}$ is the consistent least-square estimator of ψ as stated in Section 2. A final result is due to Theorem 1(e) of Hansen (1992a), which proves the limiting distribution of least-squares estimation of cointegration vector corrected for bias under the null:

$$\sqrt{T}(\hat{A}^+ - A)\Delta_T^{-1} \Rightarrow \int_0^1 dB_{1,a} X' \left(\int_0^1 X X' \right)^{-1}, \tag{A.10}$$

in which $B_{1,a}(\tau)$ is independent of $X(\tau)$.

A.3. Proof of Theorem 1

The suggested tests are functions of the partial sums process $S_{2,T}(\tau)$. We need to investigate their asymptotic distribution before deriving the distributional theory of the tests.

First, rewrite

$$\begin{aligned} \frac{1}{\sqrt{T}}\Delta_{2T}S_{2T}(\tau) &= \frac{1}{\sqrt{T}}\Delta_{2T}\sum_{t=1}^{[T\tau]} \hat{s}_{2,t} = \frac{1}{\sqrt{T}}\Delta_{2T}\sum_{t=1}^{[T\tau]} (x_{2t}\hat{u}_{1t}' - \hat{A}_{21}^+) \\ &= \frac{1}{\sqrt{T}}\Delta_{2T}\sum_{t=1}^{[T\tau]} x_{2t}u_{1t}' - \frac{1}{T}\Delta_{2T}\left(\sum_{t=1}^{[T\tau]} x_{2t}x_t'\right)\Delta_T'\Delta_T^{-1}(\hat{A}^+ - A)\sqrt{T} \\ &\quad - \frac{[T\tau]}{T}\sqrt{T}\Delta_{2T}\hat{A}_{21}^+ + \frac{1}{T}\Delta_{2T}\sum_{t=1}^{[T\tau]} [x_{2t}(u'_{at} - \hat{u}'_{at})]\hat{\Omega}_{aa}^{-1}\hat{\Omega}_{a1} \\ &\quad - \frac{1}{\sqrt{T}}\Delta_{2T}\left(\sum_{t=1}^{[T\tau]} x_{2t}u'_{at}\right)\hat{\Omega}_{aa}^{-1}\hat{\Omega}_{a1}. \end{aligned}$$

Putting together the above preliminary results, the CMT and joint convergence, we finally have

$$\begin{aligned} \frac{1}{T}\Delta_{2T}S_{2T}(\tau) &\Rightarrow \int_0^\tau X_2 dB'_{1,a} + \tau \begin{pmatrix} 0 \\ \Lambda_{21}^* \end{pmatrix} \\ &\quad - \int_0^\tau X_2 X' \left(\int_0^1 X X' \right)^{-1} \int_0^1 X dB'_{1,a} - \tau \begin{pmatrix} 0 \\ \Lambda_{21}^* \end{pmatrix} \\ &= S_2(\tau)\Omega_{1,a}^{1/2} - M_2^*(\tau)M(1)^{-1}S(1)\Omega_{1,a}^{1/2} = S_2^*(\tau)\Omega_{1,a}^{1/2}, \tag{A.11} \end{aligned}$$

where $S_2^*(\tau) = S_2(\tau) - M_2^*(\tau)M(1)^{-1}S(1)$, $S(\tau) = \int_0^\tau X dW'_1$, $S_2(\tau) = \int_0^\tau X_2 dW'_1$, and $W_1 = \Omega_{1,a}^{1/2}B_{1,a} \equiv BM(I_{m_1})$, independent of X .

By (A.11) and the CMT

$$\begin{aligned}
 F_{2T}(\tau) &= \text{tr}\{S_{2T}(\tau)'V_{2T}(\tau)^{-1}S_{2T}(\tau)\hat{\Omega}_{1.a}^{-1}\} \\
 &= \text{tr}\left\{\frac{1}{\sqrt{T}}S_{2T}(\tau)'A'_{2T}\left(\frac{1}{T}A'_{2T}V_{2T}(\tau)A_{2T}\right)^{-1}\frac{1}{\sqrt{T}}A_{2T}S_{2T}(\tau)\hat{\Omega}_{1.a}^{-1}\right\} \\
 &\Rightarrow \text{tr}\{\Omega_{1.a}^{1/2}S_{2T}^*(\tau)'V_{2T}(\tau)^{-1}S_{2T}^*(\tau)\Omega_{1.a}^{1/2}\Omega_{1.a}^{-1}\} = \text{tr}\{S_{2T}^*(\tau)'V_{2T}(\tau)^{-1}S_{2T}^*(\tau)\} \\
 &= F_2(\tau) \equiv \chi^2(v_2),
 \end{aligned} \tag{A.12}$$

where $F_2(\tau) = \text{tr}(S_{2T}^*(\tau)'V_{2T}(\tau)^{-1}S_{2T}^*(\tau))$. The equivalence relation in the last equation comes directly from the observation that conditional on \mathcal{F}_x , $\text{vec}(S_{2T}^*(\tau)) \equiv N(0, \Omega_{1.a} \otimes V_{2T}(\tau))$ for any $\tau \in \mathcal{T}$.

A.4. Proof of Theorem 2

As an application of the CMT and Eq. (A.12), both part (1) and part (2) hold naturally. By the moment matrix convergence results ($M_2(1)$), (A.6), and (A.11), part (3) follows as a result.

A.5. Proof of Theorem 3

For ease of exposition, we only consider the standard case that $m = 1$ and $\psi = 0$ because the general case follows with identical arguments. We first investigate the stochastic order for a number of processes under the alternative that $\delta^2 \neq 0$. Let A_2 be the initial fixed value of $A_{2,t}$. Thus, $y_t = A_1 + A_2x_{2t} + A_3x_{3t} + v_t$, where $v_t = u_{1t} + \sum_{i=1}^r \varepsilon_i x_{2t}$. Note that $T^{-1} \sum_{i=1}^r \varepsilon_i x_{2t} = O_p(1)$, $\sqrt{T}(\hat{\psi} - \psi)'A'_{aT} = O_p(1)$, and

$$T^{-1}(\hat{A} - A)'A'_{T}^{-1} = T^{-1}\left(\sum_{i=1}^T v_i x'_i\right)\left(\sum_{i=1}^T x_i x'_i\right)^{-1} A'_{T}^{-1} = O_p(1),$$

where \hat{A} and $\hat{\psi}$ are least-squares estimates from Eqs. (1) and (2). Under the alternative,

$$\begin{aligned}
 \frac{1}{T^{1/2}M}\hat{\Omega}_{1a} &= \frac{1}{T^{1/2}M} \sum_{i=-T}^T \left\{w(i/M)\frac{1}{T} \sum_{j=i+1}^T \hat{u}_{1j-i}\hat{u}_{aj}\right. \\
 &= \frac{1}{M} \sum_{i=-T}^T \left\{w(i/M)\left[\left(\frac{1}{T^{3/2}} \sum_{j=i+1}^T v_{j-i}u'_{aj}\right)\right.\right. \\
 &\quad \left.\left.- \frac{1}{T}(\hat{A} - A)'A'_{T}^{-1}\left(A'_{T}\frac{1}{\sqrt{T}} \sum_{j=i+1}^T x_{j-i}u'_{aj}\right)\right]\right\} \\
 &\quad \left.- \left(\frac{1}{T^2} \sum_{j=i+1}^T v_{j-i}x'_{aj}A'_{aT}\right)A'_{aT}^{-1}(\hat{\psi} - \psi)\sqrt{T}\right\}
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{T}(\hat{A} - A)'A_T^{-1} \left(A_T' \frac{1}{T} \sum_{j=i+1}^T x_{j-i} x'_{aj} A_T' \right) \\
 & \times \left. A_{aT}^{-1}(\hat{\psi} - \psi)\sqrt{T} \right] \Big\} = O_p(1)(O_p(1) + O_p(1)O_p(1)) \\
 & + O_p(1)O_p(1) + O_p(1)O_p(1)O_p(1) = O_p(1).
 \end{aligned}$$

The order is established by a combination of appendix of Phillips (1991a), Hansen (1992d) (Theorem 4.2), and previous results on the least-square estimates. Since $1/T^{1/2}M \rightarrow 0$, it follows that

$$\hat{\Omega}_{1a} = O_p(T^{1/2}M).$$

By identical arguments, we can show that

$$\hat{\Omega}_{11} = O_p(T^2M), \hat{\Lambda}_{a1} = O_p(T^{1/2}M), \hat{\Lambda}_{a1}^+ = O_p(T^{1/2}M),$$

and

$$\hat{\Omega}_{1.a} = \hat{\Omega}_{11} - \hat{\Omega}_{1a}\hat{\Omega}_{aa}^{-1}\hat{\Omega}_{a1} = O_p(T^2M),$$

where it should be noted that $\hat{\Omega}_{aa}, \hat{\Lambda}_{aa} = O_p(1)$ under both the null and the alternative.

Next, under the alternative,

$$\begin{aligned}
 \sqrt{T}(\hat{A}^+ - A)A_T^{-1} & = \left[T^{3/2} \left(\frac{1}{T^2} \sum_{i=1}^T v_i x_i' A_T' \right) \right. \\
 & \quad \left. - (\hat{\Omega}_{1a}\hat{\Omega}_{aa}^{-1}) \left(\frac{1}{\sqrt{T}} \sum_{i=1}^T u_{ai} x_i' A_T' \right) + (\hat{\Omega}_{1a}\hat{\Omega}_{aa}^{-1}) \right] \\
 & \times \left[(\sqrt{T}(\hat{\pi} - \pi)'A_{2T}^{-1}) \left(A_{2T}' \frac{1}{T} \sum_{i=1}^T x_{2i} x_i' A_T' \right) \right. \\
 & \quad \left. - \sqrt{T}(0, \hat{\Lambda}_{a1}^+)A_T' \left[A_T' \frac{1}{T} \sum_{i=1}^T x_i x_i' A_T' \right]^{-1} \right] \\
 & = (O_p(T^{3/2}) + O_p(T^{1/2}M) + O_p(T^{1/2}M) \\
 & \quad + O_p(T^{1/2}M))O_p(1) = O_p(T^{3/2})
 \end{aligned}$$

and thus

$$\begin{aligned}
 \frac{1}{\sqrt{T}}A_{2T}S_{2,rt} & = \frac{1}{\sqrt{T}}A_{2T} \sum_{i=1}^t \hat{s}_{2,i} \\
 & = T^{3/2} \left(\frac{1}{T^2}A_{2T} \sum_{i=1}^t x_{2i} v_i' \right) - (\hat{\Omega}_{1a}\hat{\Omega}_{aa}^{-1}) \left(\frac{1}{\sqrt{T}}A_{2T} \sum_{i=1}^t x_{2i} u_i' \right)
 \end{aligned}$$

$$\begin{aligned}
 & - (\hat{\Omega}_{1.a} \hat{\Omega}_{aa}^{-1}) \left(\frac{1}{T} \Delta_{2T} \sum_{i=1}^t x_{2i} x'_{2i} \Delta'_{2T} \right) (\Delta'_{2T}{}^{-1} (\hat{\pi} - \pi) \sqrt{T}) \\
 & - \left(\Delta_{2T} \frac{1}{T} \sum_{i=1}^t x_{2i} x'_i \Delta'_T \right) (\Delta'_T{}^{-1} (\hat{A}^+ - A) \sqrt{T}) \\
 & - \frac{1}{\sqrt{T}} \Delta_{2T} \sum_{i=1}^t \hat{A}_{21}^+ = O_p(T^{3/2}) + O_p(T^{1/2}M) \\
 & + O_p(T^{1/2}M) + O_p(T^{3/2}) + O_p(T^{1/2}M) = O_p(T^{3/2}).
 \end{aligned}$$

With the above results,

$$\begin{aligned}
 F_{2,Tt} &= tr \left\{ \frac{1}{\sqrt{T}} S'_{2,Tt} \Delta'_{2T} \left(\frac{1}{T} \Delta'_{2T} V_{2,Tt} \Delta_{2T} \right)^{-1} \frac{1}{\sqrt{T}} \Delta_{2T} S_{2,Tt} \hat{\Omega}_{1.a}^{-1} \right\} \\
 &= O_p(T^{3/2}) O_p(1) O_p(T^{3/2}) O_p \left(\frac{1}{T^2 M} \right) = O_p \left(\frac{T}{M} \right).
 \end{aligned}$$

Thus,

$$Mean F_2 = \frac{1}{T^*} \sum_t F_{2,Tt} = \frac{T^*}{T^*} O_p \left(\frac{T}{M} \right) = O_p \left(\frac{T}{M} \right)$$

as claimed. It also can be found that

$$\begin{aligned}
 L_{2,c} &= \frac{1}{T} \sum_t tr \left\{ \frac{1}{\sqrt{T}} S'_{2,Tt} \Delta'_{2T} \left(\frac{1}{T} \Delta'_{2T} M_{2,Tt} \Delta_{2T} \right)^{-1} \frac{1}{\sqrt{T}} \Delta_{2T} S_{2,Tt} \hat{\Omega}_{1.a}^{-1} \right\} \\
 &= \frac{T}{T} tr \left\{ O_p(T^{3/2}) O_p(1) O_p(T^{3/2}) O_p \left(\frac{1}{T^2 M} \right) \right\} = O_p \left(\frac{T}{M} \right).
 \end{aligned}$$

Turn to $Sup F_2$. Under the alternative that $A_{2,t}$ has a non-zero jump of size Δ at time $i, v_t = u_{1t} + x_{2t} 1(t \geq i) \Delta$ where $1(\cdot)$ is an indicator function. In this case, $T^{-1/2} v_t = O_p(1)$. Following the above arguments, we can show that

$$\hat{\Omega}_{1.a} = O_p(TM).$$

But for $t \geq i$,

$$\frac{1}{T} \Delta_{2T} S_{2,Tt} = O_p(T) \quad \text{and} \quad F_{2,Tt} = O_p \left(\frac{T}{M} \right),$$

while for $t < i$,

$$\frac{1}{T}A_{2T}S_{2,Tt} = O_p(1) \quad \text{and} \quad F_{2,Tt} = O_p\left(\frac{1}{TM}\right).$$

It follows that

$$\sup_t F_2 = \sup_t F_{2,Tt} = O_p\left(\frac{T}{M}\right).$$

References

- Andrews, D.W.K., 1990. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.
- Andrews, D.W.K., Fair, R.C., 1988. Inference in nonlinear econometric models with structural change. *Review of Economic Studies* 55, 615–640.
- Andrews, D.W.K., Monohan, J.C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.
- Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1384–1414.
- Bai, J., 1996. A note on spurious break and regime shift in cointegrating relationship. Working paper no. 96-13, Economics Department, MIT.
- Bai, J., Lumsdaine, R.L., Stock, J.H., 1997. Testing for and dating common breaks in multiple time series. *Review of Economic Studies*, forthcoming.
- Billingsley, P., 1968. *Convergence of probability measures*, John Wiley, New York.
- Chow, G.C., 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28, 591–605.
- Den Haan, W.J., Levin, A., 1996. Inference from parametric and non-parametric covariance matrix estimation procedures. NBER technical working paper 195.
- Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55, 251–276.
- Gregory, A.W., Hansen, B.E., 1996. Residual-based tests for cointegration with regime shifts. *Journal of Econometrics* 70, 99–126.
- Gregory, A.W., Nason, J.M., Watt, D.G., 1994. Testing for structural breaks in cointegrated relationships. Queen's University, unpublished manuscript.
- Gregory, A.W., Nason, J.M., Watt, D.G., 1996. Testing for structural breaks in cointegrated relationships. *Journal of Econometrics* 71, 321–341.
- 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. Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* 60, 967–972.
- Hansen, B.E., 1992c. Efficient estimation and testing of cointegrating vector in the presence of deterministic trends. *Journal of Econometrics* 53, 87–121.
- Hansen, B.E., 1992d. Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.
- Herndorf, 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 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.
- Kiwatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y.C., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root?. *Journal of Econometrics* 54, 159–178.
- Lucas, R.E., 1988. Money demand in the United States: a quantitative review. *Carnegie-Rochester Conference Series on Public Policy* 29, 137–168.
- Nyblom, J., 1989. Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 84, 856–864.
- Park, J.Y., Phillips, P.C.B., 1989. Statistical inference in regressions with integrated processes, Part I. *Econometric Theory* 4, 468–497.
- Perron, P., Ng, S., 1996. An autoregressive spectral density estimator at frequency zero for nonstationary tests. University of Montreal, manuscript.
- Phillips, P.C.B., 1986. Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–340.
- Phillips, P.C.B., 1987. Time series regression with a unit root. *Econometrica* 55, 277–301.
- Phillips, P.C.B., 1991a. Spectral regression cointegrated time series, In: Barnett, W. (Ed.), *Non-parametric and Semiparametric Methods in Economics and Statistics*. Cambridge University Press, Cambridge.
- Phillips, P.C.B., 1991b. Optimal inference in cointegrated system. *Econometrica* 59, 283–306.
- Phillips, P.C.B., Durlauf, S.N., 1986. Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495.
- Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with $I(1)$ processes. *Review of Economic Studies* 57, 99–125.
- Quintos, C.E., Phillips, P.C.B., 1993. Parameter constancy in cointegrating regressions. *Empirical Economics* 18, 675–703.
- Quandt, R., 1960. Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association* 55, 324–330.
- Saikkonen, P., 1991. Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory* 7, 1–21.
- Shin, Y.C., 1994. A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91–115.
- Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.