Tests for Nonlinear Cointegration*

In Choi† and Pentti Saikkonen‡

First Draft: December, 2004
Revision: September, 2005; September, 2008

Abstract

This paper develops tests for the null hypothesis of cointegration in the nonlinear regression model with $I(1)$ variables. The test statistics we use in this paper are Kwiatkowski, Phillips, Schmidt, and Shin's (1992; KPSS hereafter) tests for the null of stationarity, though using other kinds of tests is also possible. The tests are shown to depend on the limiting distributions of the estimators and parameters of the nonlinear model when they use full-sample residuals from the nonlinear least squares and nonlinear leads-and-lags regressions. This feature makes it difficult to use them in practice. As a remedy, this paper develops tests using subsamples of the regression residuals. For these tests, first, the nonlinear least squares and nonlinear leads-and-lags regressions are run and residuals are calculated. Second, the KPSS tests are applied using subresiduals of size b. As long as $b/T \to 0$ as $T \to \infty$, where T is the sample size, the tests using the subresiduals have limiting distributions that are not affected by the limiting distributions of the full-sample estimators and the parameters of the model. Third, the Bonferroni procedure is used for a selected number of the subresidual-based tests. Monte Carlo simulation shows that the tests work reasonably well in finite samples for polynomial and smooth transition regression models when the block size is chosen by the minimum volatility rule. In particular, the subresidual-based tests using the leads-and-lags regression residuals appear to be promising for empirical work. An empirical example studying the U.S. money demand equation illustrates the use of the tests.

Keywords: nonlinear cointegration, tests for cointegration, subsample, subresiduals, nonlinear least squares regression, leads-and-lags regression.

---

*The first author acknowledges financial support from the RGC Competitive Earmarked Research Grant 2003–2004 under Project No. HKUST6223/03H. The second author thanks the Research Unit of Economic Structures and Growth (RUESG) in the University of Helsinki and the Yrjö Jahnsson Foundation for financial support.

†Department of Economics, Sogang University, #1 Shinsu-dong, Mapo-gu, Seoul, 121-742 Korea. E-mail: inchoi@gmail.com.

‡Department of Mathematics and Statistics, P.O. Box 68 (Gustaf Hällströmin katu 2b), FIN-00014 University of Helsinki, Finland. E-mail: pentti.saikkonen@helsinki.fi

1 Introduction

Recently, research on cointegrated time series has been moving towards nonlinear extensions of the basic theory initiated by Granger (1981) and Engle and Granger (1987). One line of this research has focused on allowing for nonlinear short-run dynamics in error correction models, the goal often being to model potentially nonlinear adjustment mechanisms to deviations from long-run equilibrium relations. Perhaps the best-known example of this approach manifests itself in the concept of threshold cointegration and its smooth versions studied by Balke and Fomby (1997), Hansen and Seo (2002), Bec and Rahbek (2004), and Saikkonen (2008) to mention only a few. Another line of research has attempted to make the cointegrating relations themselves nonlinear. The model used in this context has been a nonlinear cointegrating regression or a nonlinear regression model with integrated regressors. References of this work include Park and Phillips (1999, 2001), Chang, Park, and Phillips (2001), Chang and Park (2003), Saikkonen and Choi (2004; SC hereafter), and Choi and Saikkonen (2004).

In SC, we developed an estimation theory for a fairly general nonlinear cointegrating regression model which covered smooth transition type nonlinearities. In that paper, the existence of a nonlinear cointegrating relation was assumed, whereas Choi and Saikkonen (2004) developed test procedures for testing the linearity of the assumed cointegrating relation. What is still lacking in the literature is a method to test for the existence of a nonlinear cointegrating relation. This paper proposes tests for nonlinear cointegration, thereby making the theory of nonlinear cointegration fully operational for empirical applications.

Previously, two approaches have been used to test for linear cointegration. One takes cointegration as the null hypothesis and noncointegration as the alternative whereas the other approach reverses the roles of the null and alternative hypotheses. In our nonlinear context, the former approach appears more convenient and will be adopted. We will employ Kwiatkowski, Phillips, Schmidt, and Shin's (1992; KPSS hereafter) test statistics that use regression residuals from either nonlinear least squares (NLLS) estimation of the specified nonlinear cointegrating relation or the modified leads-and-lags version developed by SC. It turns out that in both cases the limiting null distributions of the test statistics generally depend on unknown nuisance parameters which are difficult to eliminate. Tabulating these limiting distributions is therefore impractical, and resampling methods such as bootstrapping and subsampling are not likely to work well either.[1]

Our approach to the problem is to employ the KPSS tests using subsamples of the regression residuals. For these tests, the NLLS squares and nonlinear leads-and-lags regressions are run and residuals are calculated. Then, the KPSS tests are applied using subresiduals of size b. As long as $b/T \to 0$ as $T \to \infty$, where T is the sample size, the tests using the subresiduals have limiting distributions that are not affected by the limiting distributions of the full-sample estimators and the parameters of the model. Last, the Bonferroni procedure is used for a selected number of the subresidual-based tests. Under appropriate conditions, this approach yields tests whose asymptotic size can be controlled regardless of whether we estimate the nonlinear cointegrating regression by NLLS or its leads-and-lags modification. Monte Carlo simulation shows that the tests work reasonably well in finite samples for polynomial and smooth transition regression models when the block size is chosen by the minimum volatility rule. This holds, in particular, for the tests based on the leads-and-lags estimation, which outperform those based on the NLLS estimator in terms of finite sample size.

Our approach appears to be similar to Politis, Romano and Wolf's (1999) subsampling in the sense that subsamples of the regression residuals are used. In the subsampling approach, the statistic of interest is computed at subsamples of the data (consecutive sample points in the case of the time series), and the subsampled values of the statistic are used to estimate its finite sample distribution. By contrast, our approach estimates the model using full samples and use the subresiduals to formulate tests and, then, combine these tests along the Bonferroni procedure. Limiting

[1] See Politis, Romano and Wolf (1999) for an introduction to subsampling. Application of subsampling to nonstationary time series include, among others, Romano and Wolf (2001) and Choi (2005).

distributions of the tests involved are obtained by asymptotic analysis.

In order to illustrate the use of our tests, a smooth transition cointegrating relation is tested for the U.S. money demand equation using quarterly data with the sampling period 1959:Q1–2000:Q4. It is reported that the null of cointegration is not rejected at the 5% and 10% levels.

The paper is organized as follows. Section 2 introduces the nonlinear cointegrating regression model under consideration and the necessary assumptions. Section 3 summarizes the main results of SC on parameter estimation in this model. The test procedures of the paper are developed in Section 4 and their finite sample performance is studied in Section 5 by Monte Carlo simulation. The empirical application of the paper is presented in Section 6. Finally, proofs of the theorems are contained in Appendix I and derivation of the cumulative distribution function of $\int_0^1 W^2(s)\,ds$ is reported in Appendix II.

A few words on our notation. Weak convergence is denoted by $\Rightarrow$ and all limits are taken as $T \to \infty$. The largest integer not exceeding x is denoted by $[x]$. Conversely, the smallest integer greater than x is denoted by $[x]^*$. For an arbitrary matrix A, $\|A\| = [tr(A'A)]^{1/2}$ and, when applied to matrices, the inequality signs $>$ and $\geq$ mean the usual ordering of positive definite and semidefinite matrices, respectively.

2 The Model and Preliminary Assumptions

Consider the nonlinear cointegrating regression model

$$y_t = g(x_t, \theta) + u_t, \quad t = 1, 2, \ldots, \tag{1}$$

where x_t $(p \times 1)$ is an $I(1)$ regressor vector, u_t a zero-mean stationary error term, and $g(x_t, \theta)$ a known, smooth function of the process x_t and the parameter vector θ $(k \times 1)$. Since x_t is $I(1)$ and non-cointegrated as assumed in Assumption 3 below, $g(x_t)$ is not $I(0)$. Beyond this, the stochastic nature of $g(x_t, \theta)$ is difficult to generalize and it will depend on the type of function $g(\cdot)$. If x_t is cointegrated with a cointegrating vector θ, $g(x_t, \theta)$ may become $I(0)$ as exemplified by $g(x_t, \theta) = \theta' x_t + (\theta' x_t)^2$. But this case is excluded by Assumption 3.

We call relation (1) cointegrating regression model since a long-run, statistical equilibrium relation exists between y_t and x_t when u_t is $I(0)$. The cointegrating smooth transition model studied in SC is a special case of model (1).

Though model (1) is more general than that of SC, the results of that paper can be applied to model (1) and therefore the assumptions imposed on the present model are adopted from SC.

Assumption 1

$$x_t = x_{t-1} + v_t, \quad t = 1, 2, \ldots, \tag{2}$$

where v_t is a zero-mean stationary process and the initial value x_0 may be any random variable with the property $E\left\|x_0\right\|^4 < \infty$.

Assumption 2 *For some $r > 4$, $w_t = [u_t \; v_t']'$ is a stationary, zero-mean, strong mixing sequence with mixing coefficients of size $-4r/(r-4)$ and $E\left\|w_t\right\|^r < \infty$.*

Assumption 3 *The spectral density matrix $f_{ww}(\lambda)$ is bounded away from zero:*

$$f_{ww}(\lambda) \geq \varepsilon I_{p+1}, \quad \varepsilon > 0. \tag{3}$$

Assumption 4 *(i) The parameter space Θ of θ is a compact subset of $\mathbb{R}^k$ and the true parameter value $\theta_0 \in \Theta^0$ where Θ^0 denotes the interior of Θ.*
(ii) $g(x, \theta)$ is three times continuously differentiable on $\mathbb{R}^p \times \Theta^*$ where Θ^* is an open set containing Θ.

Choosing the real number p in Corollary 14.3 of Davidson (1994) as $2r/(r+2)$, we find that Assumption 2 implies that the serial covariances of the process w_t at lag $|j|$ are of size -2. Thus, the summability condition,

$$\sum_{j=-\infty}^{\infty} |j| \left\|E w_t w_{t+j}'\right\| < \infty, \tag{4}$$

is satisfied implying that the process w_t has a continuous spectral density matrix $f_{ww}(\lambda)$, which we assume to satisfy Assumption 3.

Assumption 2 also implies the multivariate invariance principle

$$T^{-1/2}\sum_{j=1}^{[Ts]} w_j \Rightarrow B(s), \quad 0 \leq s \leq 1, \tag{5}$$

where $B(s)$ is a Brownian motion with covariance matrix $\Omega = 2\pi f_{ww}(0)$ (see the proof of Theorem 3.1 in Hansen, 1992,). We partition $B(s) = \left[B_u(s) \;\; B_v(s)'\right]'$ and

$$\Omega = \begin{bmatrix} \omega_u^2 & \omega_{uv} \\ \omega_{vu} & \Omega_{vv} \end{bmatrix}$$

conformably with the partition of the process w_t.

Assumption 3, specialized to the case $\lambda = 0$, implies that the components of the $I(1)$ process x_t are not cointegrated. In addition, it is required for the estimation theory of Section 3 that (3) also holds for other values of λ.

Assumption 4 is the usual assumption required for deriving consistency and asymptotic distributions of nonlinear estimators. In the cointegrating smooth transition model of SC, the assumption of a compact parameter space can be relaxed for some parameters, but it is retained for the sake of generality of the model used here.

3 Nonlinear Cointegrating Regressions

We will use triangular array asymptotics[2] in order to study the NLLS and leads-and-lags estimators for model (1). In the triangular asymptotics, the actual sample size is fixed at T_0, say, and the model is embedded in a sequence of models depending on a sample size T which tends to infinity. The embedding is obtained by replacing the $I(1)$ regressor x_t by $(T_0/T)^{1/2}\, x_t$. This makes the regressand dependent on T and, when $T = T_0$, the original model is obtained. If T_0 is large, the triangular array asymptotics can be expected to give reasonable approximations for the finite sample distributions of the estimators and test statistics. Indeed, SC show that it provides reasonable approximations to the finite sample distributions of estimators

[2] In the econometrics literature, Andrews and McDermott (1995) use triangular array asymptotics for nonlinear econometric models with deterministically trending variables. Related references can also be found in that paper.

and tests for the cointegrating smooth transition model. By contrast, the conventional asymptotics fail to identify some parameters unless these are assumed to depend on sample sizes.[3] An example illustrating the failure of the conventional asymptotics is provided in SC. It may also be noted that, to the best of our knowledge, results on the NLLS and leads-and-lags estimators based on conventional asymptotics are not available for model (1) when the error term u_t is allowed to be serially correlated and correlated with the regressor x_t.

Discussions in this section will be brief because proofs of the main results use methods similar to those in SC. The reader is referred to SC for more details on the required assumptions, limiting results, and proofs.

3.1 Nonlinear least squares regression

This subsection considers the triangular array asymptotics of the NLLS estimator for model (1). It will be shown below that the NLLS estimator has a limiting distribution that, in general, depends on nuisance parameters. Using such an estimator for residual-based tests of the null of cointegration induces nuisance-parameter dependency in the limiting distributions, making them unsuitable for empirical applications (see Choi and Ahn, 1995). Still, there are two reasons we study the NLLS estimator. First, it will be used as an initial estimator for the leads-and-lags estimator in the next subsection. Second, the tests we will propose are free of nuisance parameters in the limit even when the NLLS estimator is used.

In order to use the triangular array asymptotics, embed model (1) in a sequence of models

$$y_{tT} = g\left(x_{tT}, \theta\right) + u_t, \quad t = 1, \ldots, T,$$

where $x_{tT} = \left(T_0/T\right)^{1/2} x_t$. Of course, we always choose $T = T_0$ in practice, so that the transformation x_{tT} is not required. The transformation is made only for the development of the asymptotic analysis.

[3] Such a model is used by Jin (2004).

The NLLS estimator of the parameter θ is obtained by minimizing the function

$$Q_T(\theta) = \sum_{t=1}^{T} (y_{tT} - g(x_{tT}, \theta))^2$$

with respect to θ over the parameter space Θ. Since the objective function is continuous on Θ for each $(y_{1T}, ..., y_{TT}, x_{1T}, ..., x_{TT})$ and the parameter space is compact by Assumption 4 (i), the NLLS estimator, denoted by $\tilde{\theta}_T$, exists and is Borel measurable (see Lemma 3.4 of Pötscher and Prucha, 1997).

We need to introduce additional assumptions for consistency of the NLLS estimator. For these, use the multivariate invariance principle, (5), to conclude that

$$x_{tT} \Rightarrow T_0^{1/2} B_v(s) \stackrel{def}{=} B_v^0(s) \text{ as } T \to \infty.$$

This fact and a standard application of the continuous mapping theorem show that, for every $\theta \in \Theta$,

$$T^{-1} \sum_{t=1}^{T} g(x_{tT}, \theta)^2 \Rightarrow \int_0^1 g\left(B_v^0(s), \theta\right)^2 ds.$$

The following assumption, together with our previous assumptions, ensures that the NLLS estimator is consistent. This assumption guarantees that the limit of the objective function is minimized (a.s.) at the true parameter vector θ_0. This amounts to the usual identification condition for nonlinear econometric models.

Assumption 5 *For some $s \in [0,1]$ and all $\theta \neq \theta_0$,*

$$g\left(B_v^0(s), \theta\right) \neq g(B_v^0(s), \theta_0) \text{ (a.s.).}$$

In order to obtain the limiting distribution of the NLLS estimator $\tilde{\theta}_T$, the following assumption is also needed.

Assumption 6

$$\int_0^1 K\left(B_v^0(s), \theta_0\right) K\left(B_v^0(s), \theta_0\right)' ds > 0 \text{ (a.s.),}$$

where $K(x, \theta_0) = \frac{\partial g(x,\theta)}{\partial \theta}|_{\theta=\theta_0}$.

Theorem A.1 in the Appendix reports the limiting distribution of the NLLS estimator $\tilde{\theta}_T$. It is shown that the limiting distribution depends on nuisance parameters in a complicated manner, rendering it inefficient and, in general, unsuitable for hypothesis testing. Still, it can be used for testing nonlinear cointegration as will be seen in the next section. In the special case where the processes v_t and u_t are totally uncorrelated, the limiting distribution becomes mixed normal.

3.2 Leads-and-lags regression

This subsection considers the leads-and-lags estimation procedure for model (1). The method is used in Saikkonen (1991) for linear cointegrating regressions and extended to the cointegrating smooth transition model in SC. The leads-and-lags estimation makes use of the fact that, under the assumptions of Section 2, the error term u_t can be expressed as

$$u_t = \sum_{j=-\infty}^{\infty} \pi_j' v_{t-j} + e_t, \tag{6}$$

where e_t is a zero-mean stationary process such that $Ee_t v_{t-j}' = 0$ for all $j = 0, \pm 1, \ldots,$ and

$$\sum_{j=-\infty}^{\infty} (1 + |j|) \|\pi_j\| < \infty.$$

As is well known, the long-run variance of the process e_t can be expressed as $\omega_e^2 = \omega_u^2 - \omega_{uv}\Omega_{vv}^{-1}\omega_{vu}$.

Using equations (2) and (6), we can write model (1) as

$$y_t = g(x_t, \theta) + \sum_{j=-K}^{K} \pi_j' \Delta x_{t-j} + e_{Kt}, \quad t = K+2, K+3, \ldots, \tag{7}$$

where Δ signifies the difference operator and

$$e_{Kt} = e_t + \sum_{|j|>K} \pi_j' v_{t-j}.$$

As in the previous subsection, embed the auxiliary model (7) with $t = K + 2, \ldots, T - K$ in the sequence of models defined by

$$y_{tT} = g(x_{tT}, \theta) + V_t'\pi + e_{Kt}, \quad t = K+2, \ldots, T-K, \tag{8}$$

where $x_{tT} = (T_0/T)^{1/2}\, x_t$, $V_t = \left[\Delta x'_{t-K} \ \ldots \ \Delta x'_{t+K}\right]'$ and $\pi = \left[\pi'_{-K} \ \ldots \ \pi'_{K}\right]'$. Using the NLLS estimator $\tilde{\theta}_T$ as an initial estimator, we will consider the two-step estimator[4] defined by

$$\begin{bmatrix} \hat{\theta}_T^{(1)} \\ \hat{\pi}_T^{(1)} \end{bmatrix} = \begin{bmatrix} \tilde{\theta}_T \\ 0 \end{bmatrix} + \left(\sum_{t=K+2}^{T-K} \tilde{p}_{tT} \tilde{p}'_{tT} \right)^{-1} \sum_{t=K+2}^{T-K} \tilde{p}'_{tT}\, \widetilde{u}_{tT},$$

where $\widetilde{u}_{tT} = y_{tT} - g(x_{tT}, \tilde{\theta}_T)$ and $\tilde{p}_{tT} = \left[\tilde{K}\left(x_{tT}, \tilde{\theta}_T\right)' \ V'_t \right]'$ with $\tilde{K}\left(x_{tT}, \tilde{\theta}_T\right) = \frac{\partial g(x_{tT},\theta)}{\partial \theta}_{|\theta=\tilde{\theta}_T}$.

Theorem A.2 in the Appendix reports the asymptotic properties of the two-step estimators $\hat{\theta}_T^{(1)}$ and $\hat{\pi}_T^{(1)}$. It is shown that the the limiting distribution of $\hat{\theta}_T^{(1)}$ is mixed normal and that $\hat{\pi}_T^{(1)}$ is consistent. These results will be used for testing nonlinear cointegration in the next section.

4 Tests for Nonlinear Cointegration

Model (1) implies the existence of a nonlinear cointegrating relationship between the processes y_t and x_t. In order to test for the existence of this relationship, one has to test for the stationarity of the error process u_t. As mentioned in the introduction, this testing problem has been approached in two different ways in the case of linear cointegration. One approach attempts to test the null hypothesis of cointegration against noncointegration, whereas the other approach reverses the roles of the null and alternative hypotheses. In the present nonlinear context, the former approach appears more convenient. In the latter approach, one would need to establish the asymptotic properties of the estimators, $\tilde{\theta}_T$, $\hat{\theta}_T^{(1)}$ and $\hat{\pi}_T^{(1)}$, when the error term u_t is $I(1)$. This would mean solving the problem of spurious regression in a nonlinear context, which seems difficult. Thus, we consider testing the null hypothesis that u_t is stationary against the alternative that it is an $I(1)$ process.

[4] We find that the asymptotic properties of proper NLLS estimators of the parameters in (8) are difficult to obtain.

4.1 Cointegration tests using full residuals

This subsection studies tests for the null of cointegration that use regression residuals. Residuals from the NLLS estimation of model (1) are written as

$$\tilde{u}_t = y_{tT} - g(x_{tT}, \tilde{\theta}_T).$$

Residuals $\{\tilde{u}_t\}_{t=1}^{T}$ will be called full residuals. Using the full residuals, we formulate the test statistic

$$C_{NLLS} = T^{-2}\tilde{\omega}_u^{-2} \sum_{t=1}^{T} \left(\sum_{j=1}^{t} \tilde{u}_j \right)^2, \tag{9}$$

where $\tilde{\omega}_u^2$ is a consistent estimator of the long-run variance ω_u^2 based on the full residuals. This test statistic has the same functional form as the KPSS test for stationarity and tests the null hypothesis that $\{u_t\}$ is $I(0)$. Large values of the test statistic provide evidence for the alternative. Choi and Ahn (1995) introduce other kinds of test statistics that can also be used, but our focus will be on test statistic (9) and the one to be introduced below.

Since $e_t \sim I(0)$ and $e_{Kt} \approx e_t$ for large K under the null of cointegration, we may also base our test on the residuals of the estimated version of model (8)

$$\hat{e}_{Kt} = y_{tT} - g(x_{tT}, \hat{\theta}_T^{(1)}) - V_t'\hat{\pi}_T^{(1)}, \quad t = K+2, \ldots, T-K.$$

Again $\{\hat{e}_{Kt}\}_{t=K+2}^{T-K}$ will be called full residuals. The test statistic is defined by

$$C_{LL} = N^{-2}\tilde{\omega}_e^{-2} \sum_{t=K+2}^{T-K} \left(\sum_{j=K+2}^{t} \hat{e}_{Kj} \right)^2, \tag{10}$$

where $\tilde{\omega}_e^2$ is a consistent estimator of the long-run variance ω_e^2 based on $\{\hat{e}_{Kt}\}_{t=K+2}^{T-K}$. See Andrews (1991), for example, for the long-run variance estimation. As discussed in SC, the consistency of conventional long-run variance estimators is guaranteed by Theorem A.2. This test statistic has also been considered by Shin (1994) for the null of linear cointegration.

The limiting distributions of the C_{NLLS} and C_{LL} test statistics under the null hypothesis are given in Lemma A.5 in the Appendix. It shows that, in general, the limiting distributions of test statistics C_{NLLS} and C_{LL} depend on unknown parameters. In the special case of a linear regression, the dependences cancel for the latter test statistic whose limiting distribution reduces to Shin's (1994) special case of Theorem 2. This also happens in the case of a polynomial regression model. However, even in these cases, the limiting distributions of test statistic C_{NLLS} depends on nuisance parameters unless the regressors are strictly exogenous (i.e., $\omega_{uv} = 0$).

The above discussion implies that tabulations of the limiting distributions of test statistics C_{NLLS} and C_{LL} are impractical when the regression model is nonlinear. A workable test procedure will be introduced in next subsection.

4.2 Cointegration tests using subresiduals

This subsection introduces a new test procedure for nonlinear cointegration that uses subresiduals and the Bonferroni procedure. To this end, consider test statistics

$$C_{NLLS}^{b,\mathbf{i}} = b^{-2}\tilde{\omega}_{i,u}^{-2} \sum_{t=\mathbf{i}}^{\mathbf{i}+b-1} \left(\sum_{j=\mathbf{i}}^{t} \tilde{u}_j \right)^2 \tag{11}$$

and

$$C_{LL}^{b,\mathbf{i}} = (b-2K-1)^{-2}\tilde{\omega}_{i,e}^{-2} \sum_{t=\mathbf{i}+K+2}^{\mathbf{i}+b-K} \left(\sum_{j=\mathbf{i}+K+2}^{t} \hat{e}_{Kj} \right)^2, \tag{12}$$

which have the same functional forms as (9) and (10), respectively, but use subresiduals $\{\tilde{u}_t\}_{t=\mathbf{i}}^{\mathbf{i}+b-1}$ and $\{\hat{e}_{Kt}\}_{t=\mathbf{i}}^{\mathbf{i}+b-1}$. The index $\mathbf{i}$ denotes the starting point of the subresiduals and b the size of subresiduals, which we call the block size.

When the block size grows with the sample size but at a slower rate, the regressors have no effects on the limiting distributions of test statistics (11) and (12). Therefore, the limiting distributions are different from those in Lemma A.5 in the Appendix, as shown in the following theorem.

Theorem 1 *(i) Suppose that the assumptions of Theorem A.1 hold and that $\tilde{\omega}_{i,u}^2$ is a consistent estimator of ω_u^2. If $b \to \infty$ and $b/T \to 0$ as $T \to \infty$, for any i with $1 \le i \le T-b$,*

$$C_{NLLS}^{b,\mathbf{i}} \Rightarrow \int_0^1 W^2(s)\,ds,$$

where $W(s)$ is a standard Brownian motion.

(ii) Suppose that the assumptions of Theorem A.2 hold and that $\tilde{\omega}_{i,e}^2$ is a consistent estimator of ω_e^2. If $b \to \infty$ and $b/T \to 0$ as $T \to \infty$, for any i with $1 \le i \le T - b$,

$$C_{LL}^{b,\mathbf{i}} \Rightarrow \int_0^1 W^2(s)\, ds.$$

The results of this theorem can also be proved by using conventional asymptotics provided one can show that the NLLS estimator and the leads-and-lags estimator are consistent with an appropriate order and the first and second partial derivatives $\partial g(x_t, \theta)/\partial\theta$ and $\partial g(x_t, \theta)/\partial\theta\partial\theta'$ satisfy appropriate boundedness conditions. This can be seen from the proof of Theorem 1 where some additional remarks on this point are given.

As shown in Appendix II, the cumulative distribution function of $\int_0^1 W^2(s)\, ds$ is given as

$$cdf(z) = \sqrt{2} \sum_{n=0}^{\infty} \frac{\Gamma(n+1/2)}{n!\Gamma(1/2)} (-1)^n \left[1 - Erf\left(\frac{u}{2\sqrt{z}}\right)\right],\ z \ge 0, \tag{13}$$

where $Erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-\beta^2) d\beta$ and $u = \frac{\sqrt{2}}{2} + 2n\sqrt{2}$. In practice, the series can be truncated at small n. The experimental results reported in Section 5 use critical values of the distribution obtained by truncating the series at $n = 10$.

Though Theorem 1 shows that test statistics $C_{NLLS}^{b,\mathbf{i}}$ and $C_{LL}^{b,\mathbf{i}}$ are free of nuisance parameters in the limit, the tests are likely to have low power compared to those employing full residuals. Thus, we consider using the tests along with the Bonferroni procedure. For this, select M tests $C_{NLLS}^{b,\mathbf{i}_1}, ..., C_{NLLS}^{b,\mathbf{i}_M}$ and define

$$C_{NLLS}^{b,\max} = \max\left(C_{NLLS}^{b,\mathbf{i}_1}, ..., C_{NLLS}^{b,\mathbf{i}_M}\right).$$

These M tests have the same block size, but use different starting points $\mathbf{i}_1, ..., \mathbf{i}_M$. Due to Theorem 1 and the Bonferroni inequality,

$$\lim_{T\to\infty} P[C_{NLLS}^{b,\max} \le c_{\alpha/M}] \ge 1 - \alpha,$$

where $c_{\alpha/M}$ is the $\frac{\alpha}{M}$-level critical value from the distribution of $\int_0^1 W^2(s)\, ds$ (i.e., $P[\int_0^1 W^2(s)\, ds \ge c_{\alpha/M}] = \frac{\alpha}{M}$). The same steps can be followed using $C_{LL}^{b,\mathbf{i}_1}, ..., C_{LL}^{b,\mathbf{i}_M}$.

This testing procedure implies that the α-level critical values for test statistics $C_{NLLS}^{b,\max}$ and $C_{LL}^{b,\max}$ are taken from the distribution of $\int_0^1 W^2(s)\,ds$ using the level $\frac{\alpha}{M}$.

There is a host of issues to resolve before implementing the above tests in practice. We examine these issues next.

4.2.1 Choosing subresidual-based tests

For a given block size b, we suggest choosing M, the number of subresidual-based tests used in the Bonferroni procedure, and $\mathbf{i}_1, ..., \mathbf{i}_M$, the starting points of the subresiduals needed for test statistic $C_{NLLS}^{b,\max}$, as follows.

Step 1: Let $M = [T/b]^*$, where $[x]^*$ denotes the smallest integer greater than or equal to x.

Step 2: Let $\mathbf{i}_1 = 1, \mathbf{i}_2 = T - b + 1, \mathbf{i}_3 = b + 1, \mathbf{i}_4 = T - 2b + 1, ...$

These steps guarantee that the whole sample is used to calculate $C_{NLLS}^{b,\max}$, while trying to minimize M. If M is too large, the test procedure will have quite low power. Some sample points may be used more than once for the test, but this does not pose any problem. For the $C_{LL}^{b,\max}$, replace T with the effective sample size $T - 2K - 1$ in the above steps. No doubt, the above methods are arbitrary, and other methods may be used. But experimental results in next section will show that these work well in finite samples.

4.2.2 Choosing block size

We will use the minimum volatility rule in order to choose the block size. This method has often been used in the literature of subsampling (see, e.g., Romano and Wolf, 2001). The algorithm for the minimum volatility method is summarized as follows.

Step 1: Choose integers b_{small} and b_{big} with a restriction $b_{small} < b_{big}$.

Step 2: For each integer b_i in the interval $[b_{small}, b_{big}]$, calculate test statistics $C_{NLLS}^{b_{i-m},\max}, ..., C_{NLLS}^{b_{i+m},\max}$. Here, m is a small positive integer.

Step 3: Calculate the standard deviation of $C_{NLLS}^{b_{i-m},\max}, ..., C_{NLLS}^{b_{i+m},\max}$ obtained in Step 2 and denote it as SC_i.

Step 4: Choose the block size that gives the minimum of SC_i over $b_i = b_{small}, ..., b_{big}$.

The same steps are taken for test statistic $C_{LL}^{b,\max}$. In the simulation of the next section, we used the rule $b_{small} = [T^{0.7}]$ and $b_{big} = [T^{0.9}]$ with $[x]$ denoting the integer part of x. When $T = 150$, for example, b_{small} is 33 and b_{big} 90. No doubt, other rules can also be used instead of ours. In addition, we chose $m = 2$ as in Romano and Wolf (2001). These rules gave reasonably good simulation results as will be seen in the next section.

5 Simulation

This section reports simulation results for the cointegration tests we have proposed. Two nonlinear models are considered in this section: polynomial and smooth transition regression models. In addition, linear regression model is also considered in order to compare the subresidual test with conventional cointegration tests.[5]

5.1 Polynomial regression model

Data were generated by using

$$\begin{aligned} y_t &= \theta_0 + \theta_1 x_t + \theta_2 x_t^2 + u_t,\ (t = -29, ..., T) \qquad (14) \\ \theta_0 &= 0;\ \theta_1 = \theta_2 = 1; \end{aligned}$$

with

$$\begin{aligned} u_t &= \alpha u_{t-1} + \varepsilon_t; \qquad (15) \\ \begin{pmatrix} \Delta x_t \\ \varepsilon_t \end{pmatrix} &\sim iid\ N\left(0, \begin{bmatrix} 1 & \lambda \\ \lambda & 1 \end{bmatrix}\right) \\ \lambda &= 0.5. \end{aligned}$$

[5] We thank Bruce Hansen for suggesting this.

In this data generation, the case $|\alpha| < 1$ corresponds to nonlinear cointegration and the case $\alpha = 1$ to nonlinear noncointegration. We used $\alpha = 0.5, 0.8, 0.95$ for the null of nonlinear cointegration. In addition, the regressors and errors are correlated when parameter λ takes nonzero values. This section reports results using $\lambda = 0.5$, but $\lambda = 0.2$ did not give any qualitatively different results except that the test using the NLLS residuals improved relative to that using the leads-and-lags regression residuals.

The first thirty observations (corresponding to $t = -29, ..., 0$) were cut in calculating the test statistics in order to minimize the impact of the initial values x_{-29} and u_{-29}. The initial values were taken to be zero. The tests require selecting the lag length and spectral window for the long-run variance estimators $\tilde{\omega}_{i,u}^2$ and $\tilde{\omega}_{i,e}^2$. We used $[4(b/100)^{0.25}]$ and $[12(b/100)^{0.25}]$ for the lag length as suggested by KPSS. In addition, the Quadratic Spectral window was used for the long-run variance estimation, mainly because it is recommended for use by Andrews (1991). But it has been known that the choice of a window does not affect the finite sample size and power of the KPSS tests in any significant way. The minimum volatility rule was used for choosing block sizes.

Empirical size and power of the $C_{NLLS}^{b,\max}$ and $C_{LL}^{b,\max}$ tests are reported in Table 1. Here and elsewhere empirical size and power refer to the rejection frequencies in percentage under the null and alternative hypotheses, respectively. The number of replications was 3,000.

The results in Table 1 can be summarized as follows.

- The subresidual-based tests show empirical size less than the corresponding test level when $\alpha = 0.5, 0.8$. But they show size distortions when $\alpha = 0.95$ with the lag length $[4(b/100)^{0.25}]$. When the lag length $[12(b/100)^{0.25}]$ is used, the size distortions with $\alpha = 0.95$ disappear. Size distortions of the KPSS test when the null is close to the alternative are not new. They are also reported elsewhere (e.g., Caner and Kilian, 2001).

- The empirical power is not high when $T = 150$, but it improves as the sample size increases.

- The $C_{LL}^{b,\max}$ test tends to reject less often than the $C_{NLLS}^{b,\max}$ test.

- The empirical size and power tend to decrease as K increases.

- The empirical size and power are lower with the lag length $[12(b/100)^{0.25}]$ than with $[4(b/100)^{0.25}]$.

Table 1 reports reasonably satisfactory properties of the subresidual-based tests. But they tend to underreject under the null hypothesis most likely due to the Bonferroni inequality being used. This also translates into lower power as we have seen In Table 1.

5.2 Smooth transition regression model

Data for this subsection were generated by

$$
\begin{aligned}
y_t &= \theta_0 + \theta_1 x_t + \theta_2 g(x_t) x_t + u_t, \ (t = -29, ..., T); \\
g(x_t) &= \frac{1}{1 + \exp(-\theta_3(x_t - \theta_4))}; \\
\theta_0 &= 0; \theta_1 = \theta_2 = 1; \theta_3 = 1; \theta_4 = 5
\end{aligned}
$$

and the data generating process (15) in the previous subsection. The data $\{x_t\}$ were generated such that θ_4 is located in between the 15th and 85th percentiles of $\{x_t\}$. When θ_4 is near the endpoints of the sample, extremely poor estimates of parameter θ_4 are sometimes produced which affects other parameter estimates to the extent that evaluating the finite sample performance of the subresidual-based cointegration test at different sample sizes becomes meaningless. Other aspects of our simulation are the same as in the previous subsection except that the numbers of replications were set at 500, 300, 300 for $T = 150, 300, 600,$ respectively.

Empirical size and power of the subresidual-based tests for the smooth transition regression model are reported in Table 2. We find that the results in Table 2 are qualitatively no different from those in Table 1 except that the empirical power and the empirical size under $\alpha = 0.95$ become lower in Table 2. The lower empirical power is expected because there are more parameters to estimate for the smooth transition model that for the polynomial regression model. The results in Tables 1 and 2 suggests that the subresidual-based tests along with the block size choice rules we have used work well for the models we are experimenting with and that they are also likely to work reasonably well for other nonlinear models.

5.3 Linear regression model

The subresidual-based tests considered so far overcome the nuisance parameter problem in the limiting distributions, but are expected to deliver lower power than conventional tests if these conventional tests are allowed to be used. In order to gauge the power loss, we apply the subresidual-test to a linear model and compare its performance with that of the conventional $KPSS$ cointegration test (cf. Shin, 1994). The data were generated in the same way for model (14) except that x_t^2 is not present.

Empirical size and power of the $C_{NLLS}^{b,\max}$ and $C_{LL}^{b,\max}$ tests are reported in Table 3, and those of Shin's (1994) $KPSS$ cointegration test in Table 4. In Table 4, we also consider the $KPSS$ cointegration test using the OLS residuals for completeness but it should be borne in mind that using the OLS residuals for the $KPSS$ test invites the problem of nuisance parameters in limiting distributions. The number of replications for Tables 3 and 4 was 3,000.

We summarize the results in Tables 3 and 4 as follows.

- The size and power properties of the subresidual-based tests reported in Table 3 are similar to those in Tables 1 and 2.

- Comparing Tables 3 and 4, we find that the conventional $KPSS$ cointegration test is more powerful than the subresidual-based tests at the cost of higher size distortions in the vicinity of the null hypothesis. The lower power of the subresidual-based tests stems most likely from the Bonferroni inequality being used for them.

- It is hard to judge which test procedure should be preferred because the conventional $KPSS$ cointegration test is more powerful only at the cost of higher size distortions. If one is quite concerned about rejecting the null hypothesis falsely, the subresidual-based tests should be used and vice versa.

6 An empirical example

This section reports an application of our test procedure to the U.S. money demand equation. We use the U.S. quarterly data with the sampling period 1959:Q1–2000:Q4. The data we use are M1 for money, GDP for income, GDP deflator for price level and the 90-day Treasury bill rate for a short-term interest rate. These data were taken from International Financial Statistics. Note that the M1 and GDP series are seasonally adjusted. Three variables—real M1, real GDP and the Treasury bill rate—were used for the smooth transition modelling of the money demand equation. Natural logs were taken of the real M1 and real GDP. Natural log is also taken of the Treasury bill rate for the smooth transition regression because this provides better fits than using the rate itself.

The model we use is

$$\begin{aligned} y_t &= \theta_0 + \theta_1 x_{1t} + \theta_2 x_{2t} + \theta_3 g\left(x_{2t}\right) x_{2t} + u_t; \qquad (16) \\ g(x_{2t}) &= \frac{1}{1+\exp(-\theta_4(x_{2t}-\theta_5))}, \end{aligned}$$

where $y = \ln(\text{M1}) - \ln(\text{GDP deflator}), x_1 = \ln(\text{GDP}) - \ln(\text{GDP deflator}), x_2 = \ln($T-bill rate). This model indicates a nonlinear relationship between money and interest rate when the coefficients θ_3 and θ_4 are both nonzero. If both θ_2 and θ_3 take negative values, the money demand is more responsive to the interest rate when the rate is higher. If the interest rate is higher, the opportunity cost of holding money also becomes higher, inducing the public to be more sensitive to the interest rate when they make decisions on holding money. This reasoning renders economic sense to the negative values of the coefficients θ_2 and θ_3. Using the real income as a transition variable in equation (16) is also a sensible option, because the money demand function may display an asymmetric behavior over various phases of the business cycle (see Neftci, 1984). However, we did not pursue this here for brevity.

Choi and Saikkonen's (2004) tests for the null of linearity applied to the data show evidence of nonlinearity when the interest rate is used as a transition variable,[6] though this does not indicate that the smooth transition model is most appropriate. In the linear cointegrating regression for the money demand equation (see, e.g., Hoffman and Rasche, 1991; Hoffman, Rasche and Tieslau, 1995; Lütkepohl, Teräsvirta and Wolters, 1999; Stock and Watson, 1993 and Chapter 10 of Patterson, 2000), model (16) without the nonlinear term has been used. Long-term interest rates have also been used instead of the short-terms rate, but this does not usually give noticeable differences.

Table 5 reports asymptotic p-values of the $C_{LL}^{b,\max}$ and $C_{NLLS}^{b,\max}$ tests. The p-values were calculated using the distribution function of $\int_0^1 W^2(s)\,ds$ given in (13). If the p-value is smaller than the adjusted nominal level—the nominal level divided by the number of subresidual-based tests used for the Bonferroni procedure (M)—the null hypothesis of cointegration is rejected. The p-values were calculated using the cumulative distribution function given right after Theorem 1. Results in Table 5 indicate that the null is not rejected at the 5% and 10% levels no matter which test is used. Stable linear cointegrating relations among the variables we have studied are reported elsewhere (e.g., Patterson, 2000; Stock and Watson, 1993; Hoffman and Rasche, 1991), but the nonlinear cointegrating relation reported here is new.

Table 6 reports the estimation results for the smooth transition money demand equation. Both the NLLS and leads-and-lags regressions provide the qualitatively same results: the money demand is positively related to the real income and negatively to the interest rate. Notably, the coefficient for the nonlinear term is negative, which implies that the money demand is more responsive to the interest rate when it is higher, but the value of the coefficient for the linear interest rate term (θ_2) is significantly different from zero, which excludes the possibility of a liquidity trap. Standard errors are also reported in the table. For the coefficients inducing nonlinearity (θ_3 and θ_4), these standard errors cannot be used in the conventional manner to check whether these coefficients take a zero value. This is because, under the relevant null

[6] Details are omitted here.

hypotheses, some parameters of the model are not identified.

This section considers cointegration tests using the smooth transition model for the U.S. money demand. However, more research is needed to determine if other nonlinear models are more appropriate than the one used here. Even if the smooth transition model continues to be used, the search for better models is still in order. Pursuing all these possibilities is beyond the scope of this paper. The test procedure developed in this paper will certainly facilitate that endeavor.

Appendix I: Proofs

The two theorems below report the asymptotic distributions of the NLLS and leads-and-lags estimators.

Theorem A.1 *Suppose that Assumptions 1–6 hold. Then,*

$$T^{1/2}\left(\tilde{\theta}_T-\theta_0\right) \Rightarrow \left(\int_0^1 K\left(B_v^0(s),\theta_0\right)K\left(B_v^0(s),\theta_0\right)' ds\right)^{-1} \times\left(\int_0^1 K\left(B_v^0(s),\theta_0\right)dB_u(s)+\int_0^1 K_1\left(B_v^0(s),\theta_0\right)ds\kappa_{vu}\right) \stackrel{def}{=} \psi\left(B_v^0,\theta_0,\kappa_{vu}\right),$$

where $K_1(x,\theta_0)=\frac{\partial K(x,\theta)}{\partial x'}_{|\theta=\theta_0}$ *and* $\kappa_{vu}=\sum_{j=0}^{\infty}Ev_0u_j$.

Proof: This can readily be obtained by adapting the proofs of Theorems 1 and 2 of SC. Details are thus omitted.

Note also that Theorem A.1 shows that the estimator $\tilde{\theta}_T$ is consistent of order $O_p\left(T^{-1/2}\right)$ and not of order $O_p\left(T^{-1}\right)$ in the case of a linear cointegrating regression. This difference is due to the triangular array asymptotics in which the regressand is made bounded.

In the following theorem, we denote by π_0 the true value of π and, for convenience, let $N=T-2K-1$ signify the effective number of observations used to form the estimators $\hat{\theta}_T^{(1)}$ and $\hat{\pi}_T^{(1)}$.

Theorem A.2 *Suppose that the assumptions of Theorem A.1 hold and that* $K\to\infty$ *in such a way that* $K^3/T\to 0$ *and* $T^{1/2}\sum_{|j|>K}\|\pi_j\|\to 0$. *Then,*

(i)

$$N^{1/2}\left(\hat{\theta}_T^{(1)}-\theta_0\right) \Rightarrow \left(\int_0^1 K\left(B_v^0(s),\theta_0\right)K\left(B_v^0(s),\theta_0\right)' ds\right)^{-1}\int_0^1 K\left(B_v^0(s),\theta_0\right)dB_e(s) \stackrel{def}{=} \chi\left(B_v^0,\theta_0\right),$$

where $B_e(s)$ *is a Brownian motion which is independent of* $B_v(s)$ *and has variance* ω_e^2.

(ii) $\left\|\hat{\pi}_T^{(1)}-\pi_0\right\|=O_p\left(K^{1/2}/N^{1/2}\right)$.

Proof: Theorem A.2 can be proven in the same manner as Theorem 3 of SC. Details are not reported here.

The independence of the Brownian motions $B_e(s)$ and $B_v(s)$ implies that the distribution in part (i) is mixed normal.

The following lemmas will be used to prove Lemma A.5.

Lemma A.3 *If the assumptions of Theorem A.1 hold,*

$$T^{-1/2}\sum_{t=1}^{[Ts]}\tilde{u}_t \Rightarrow B_u(s) - F\left(s, B_v^0, \theta_0\right)'\psi\left(B_v^0, \theta_0, \kappa_{vu}\right).$$

Proof: Since $\tilde{u}_t = u_t - \left(g\left(x_{tT}, \tilde{\theta}_T\right) - g\left(x_{tT}, \theta^0\right)\right)$, a second order Taylor series expansion of $g\left(x_{tT}, \tilde{\theta}_T\right)$ around θ^0 gives

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{[Ts]}\tilde{u}_t &= T^{-1/2}\sum_{t=1}^{[Ts]}u_t - T^{-1/2}\sum_{t=1}^{[Ts]}K\left(x_{tT}, \theta_0\right)'\left(\tilde{\theta}_T - \theta_0\right) \\ &\quad + T^{1/2}\left(\tilde{\theta}_T - \theta_0\right)'\left(T^{-1}\sum_{t=1}^{[Ts]}\partial^2 g\left(x_{tT}, \dot{\theta}_T\right)/\partial\theta\partial\theta'\right)\left(\tilde{\theta}_T - \theta_0\right), \end{aligned} \quad \text{(A.1)}$$

where $\left\|\dot{\theta}_T - \theta_0\right\| \le \left\|\tilde{\theta}_T - \theta_0\right\|$ and the notation $K(x, \theta_0) = \frac{\partial g(x,\theta)}{\partial\theta}_{|\theta=\theta_0}$ has again been used. Since the multivariate invariance principle, (5), implies that $\max_{1\le t\le T}\|x_{tT}\| = O_p(1)$ and since the function $\partial^2 g(\cdot,\cdot)/\partial\theta\partial\theta'$ is bounded on compact subsets of its domain, it follows from Lemma 1(i) of SC that the matrix in the middle of the third term on the right-hand side of the preceding equation is of order $O_p(1)$ uniformly in $0 \le s \le 1$. From this fact and Theorem A.1, we conclude that, uniformly in $0 \le s \le 1$,

$$T^{-1/2}\sum_{t=1}^{[Ts]}\tilde{u}_t = T^{-1/2}\sum_{t=1}^{[Ts]}u_t - T^{-1/2}\sum_{t=1}^{[Ts]}K\left(x_{tT}, \theta_0\right)'\left(\tilde{\theta}_T - \theta_0\right) + o_p(1).$$

The first term on the right-hand side converges weakly to $B_u(s)$. For the second term, use the multivariate invariance principle, (5), and the continuous mapping theorem to obtain $T^{-1}\sum_{t=1}^{[Ts]}K\left(x_{tT}, \theta_0\right) \Rightarrow F\left(s, B_v^0, \theta_0\right)$ whereas $T^{1/2}\left(\tilde{\theta}_T - \theta_0\right) \Rightarrow \psi\left(B_v^0, \theta_0, \kappa_{vu}\right)$ by Theorem A.1 (ii). Since these weak convergencies hold jointly, the stated result follows.

Lemma A.4 *If the assumptions of Theorem A.2 hold,*

$$N^{-1/2}\sum_{t=K+2}^{[(T-K)s]}\hat{e}_{Kt}\Rightarrow B_e(s)-F\left(s,B_v^0,\theta_0\right)'\chi\left(B_v^0,\theta_0\right).$$

Proof: From the definitions, it follows that

$$\hat{e}_{Kt}=e_t-\left(g(x_{tT},\hat{\theta}_T^{(1)})-g(x_{tT},\theta_0)\right)+\sum_{|j|>K}\pi_j'v_{t-j}-V_t'\left(\hat{\pi}_T^{(1)}-\pi_0\right). \tag{A.2}$$

We proceed by showing that the contribution of the last two terms on the right-hand side is asymptotically negligible. This is the case if

$$\max_{K+2\leq n\leq T-K}\left|N^{-1/2}\sum_{t=K+2}^{n}\left(\sum_{|j|>K}\pi_j'v_{t-j}\right)\right|=o_p(1) \tag{A.3}$$

and

$$\max_{K+2\leq n\leq T-K}\left|N^{-1/2}\sum_{t=K+2}^{n}V_t'\left(\hat{\pi}_T-\pi_0\right)\right|=o_p(1). \tag{A.4}$$

Let $\epsilon>0$ and note that

$$\begin{aligned}&P\left\{\max_{K+2\leq n\leq T-K}\left|N^{-1/2}\sum_{t=K+2}^{n}\left(\sum_{|j|>K}\pi_j'v_{t-j}\right)\right|>\epsilon\right\}\\ \leq\ &\sum_{n=K+2}^{T-K}P\left\{\left|\sum_{t=K+2}^{n}\left(\sum_{|j|>K}\pi_j'v_{t-j}\right)\right|>N^{1/2}\epsilon\right\}\\ \leq\ &\frac{1}{\epsilon^2N}\sum_{n=K+2}^{T-K}E\left(\sum_{t=K+2}^{n}\left(\sum_{|j|>K}\pi_j'v_{t-j}\right)\right)^2,\end{aligned}$$

where the latter relation follows from Markov's inequality. In the last expression we have

$$E\left(\sum_{t=K+2}^{n}\left(\sum_{|j|>K}\pi_j'v_{t-j}\right)\right)^2=\sum_{t_1=K+2}^{n}\sum_{t_2=K+2}^{n}\sum_{|j_1|>K}\sum_{|j_2|>K}\pi_{j_1}'\Gamma_v\left(t_2-j_2-t_1+j_1\right)\pi_{j_2},$$

where $\Gamma_v(l)=Ev_tv_{t+l}'$. Since our assumptions imply that $\sum_{l=-\infty}^{\infty}\|\Gamma_v(l)\|<\infty$ (see relation (4)) the modulus of the right-hand side is bounded by a constant times $n\left(\sum_{|j|>K}\|\pi_j\|\right)^2$. Thus, we can conclude that

$$P\left\{\max_{K+2\leq n\leq T-K}\left|N^{-1/2}\sum_{t=K+2}^{n}\left(\sum_{|j|>K}\pi_j'v_{t-j}\right)\right|>\epsilon\right\}\leq\frac{c}{\epsilon^2}\left(N^{-1/2}\sum_{|j|>K}\|\pi_j\|\right)^2$$

for some finite constant c. The right-hand side is of order $o(1)$ by assumption and (A.3) follows.

In order to prove (A.4), use first the Cauchy-Schwarz inequality to get

$$\max_{K+2\leq n\leq T-K}\left|N^{-1/2}\sum_{t=K+2}^{n}V_t'\left(\hat{\pi}_T-\pi_0\right)\right|\leq\max_{K+2\leq n\leq T-K}\left\|N^{-1/2}\sum_{t=K+2}^{n}V_t\right\|\left\|\hat{\pi}_T-\pi_0\right\|.$$

Since $\|\hat{\pi}_T-\pi_0\|=O_p\left(K^{1/2}/N^{1/2}\right)$ by Theorem A.2(ii), this implies that we need to establish

$$\max_{K+2\leq n\leq T-K}\frac{K^{1/2}}{N}\left\|\sum_{t=K+2}^{n}V_t\right\|=o_p\left(1\right). \tag{A.5}$$

By the definitions, $\left\|\sum_{t=K+2}^{n}V_t\right\|=\left(\sum_{l=-K}^{K}\left\|\sum_{t=K+2}^{n}v_{t-l}\right\|^2\right)^{1/2}$ and hence, for any $\epsilon>0$,

$$\begin{aligned}
&P\left\{\max_{K+2\leq n\leq T-K}\frac{K^{1/2}}{N}\left\|\sum_{t=K+2}^{n}V_t\right\|>\epsilon\right\}\\
=\;&P\left\{\max_{K+2\leq n\leq T-K}\sum_{l=-K}^{K}\left\|\sum_{t=K+2}^{n}v_{t-l}\right\|^2>\frac{N^2}{K}\epsilon^2\right\}\\
\leq\;&P\left\{\max_{K+2\leq n\leq T-K}\max_{-K\leq l\leq K}\left\|\sum_{t=K+2}^{n}v_{t-l}\right\|^2>\frac{N^2}{K\left(2K+1\right)}\epsilon^2\right\}\\
\leq\;&\sum_{l=-K}^{K}P\left\{\max_{K+2\leq n\leq T-K}\left\|\sum_{t=K+2}^{n}v_{t-l}\right\|^2>\frac{N^2}{K\left(2K+1\right)}\epsilon^2\right\}.
\end{aligned}$$

Markov's inequality and the stationarity of process v_t now give

$$\begin{aligned}
P\left\{\max_{K+2\leq n\leq T-K}\frac{K^{1/2}}{N}\left\|\sum_{t=K+2}^{n}V_t\right\|>\epsilon\right\}&\leq\frac{K\left(2K+1\right)^2}{\epsilon^2N^2}E\left(\max_{K+2\leq n\leq T-K}\left\|\sum_{t=K+2}^{n}v_t\right\|^2\right)\\
&\leq\frac{cK\left(2K+1\right)^2}{\epsilon^2N},
\end{aligned}$$

where c is a finite constant. Here the latter inequality is due to the maximal inequality given in Corollary 16.10 of Davidson (1994). The applicability of this corollary in the present context is demonstrated by Example 16.3 of the same reference and the discussion leading to it. Since $K^3/N\to0$ by assumption, the last two inequalities prove (A.5) and hence (A.4).

Now consider equation (A.2) and use (A.3) and (A.4) to obtain

$$N^{-1/2}\sum_{t=K+2}^{[(T-K)s]}\hat{e}_{Kt}=N^{-1/2}\sum_{t=K+2}^{[(T-K)s]}e_t-N^{-1/2}\sum_{t=K+2}^{[(T-K)s]}\left(g(x_{tT},\hat{\theta}_T^{(1)})-g(x_{tT},\theta_0)\right)+o_p(1)$$

where the last term on the right-hand side is uniform in $0\leq s\leq 1$. By a second-order Taylor series expansion, we can write this further as

$$\begin{aligned}N^{-1/2}\sum_{t=K+2}^{[(T-K)s]}\hat{e}_{Kt} &= N^{-1/2}\sum_{t=K+2}^{[(T-K)s]}e_t-N^{-1/2}\sum_{t=K+2}^{[(T-K)s]}K\left(x_{tT},\theta_0\right)'\left(\hat{\theta}_T^{(1)}-\theta_0\right)\\ &\quad +N^{1/2}\left(\hat{\theta}_T^{(1)}-\theta_0\right)'\left(N^{-1}\sum_{t=K+2}^{[(T-K)s]}\partial^2 g\left(x_{tT},\dot{\theta}_T\right)/\partial\theta\partial\theta'\right)\left(\hat{\theta}_T^{(1)}-\theta_0\right)\\ &\quad +o_p(1),\end{aligned}$$

where $\left\|\dot{\theta}_T-\theta_0\right\|\leq\left\|\hat{\theta}_T^{(1)}-\theta_0\right\|$. As in the proof of Lemma A.3,

$$N^{-1/2}\sum_{t=K+2}^{[(T-K)s]}\hat{e}_{Kt}=N^{-1/2}\sum_{t=K+2}^{[(T-K)s]}e_t-N^{-1/2}\sum_{t=K+2}^{[(T-K)s]}K\left(x_{tT},\theta_0\right)'\left(\hat{\theta}_T^{(1)}-\theta_0\right)+o_p(1)$$

uniformly in $0\leq s\leq 1$. The first term on the right-hand side converges weakly to $B_e(s)$, as discussed in the proof of Theorem 3 of SC. The weak limit of the second term is obtained by using the multivariate invariance principle (5), the continuous mapping theorem and part (i) of Theorem A.2 as in the proof of Lemma A.3. This completes the proof.

Lemma A.5 *(i) If the assumptions of Theorem A.1 hold and $\tilde{\omega}_u^2$ is a consistent estimator of ω_u^2,*

$$C_{NLLS}\Rightarrow\omega_u^{-2}\int_0^1\left[B_u(s)-F\left(s,B_v^0,\theta_0\right)'\psi\left(B_v^0,\theta_0,\kappa_{vu}\right)\right]^2 ds,$$

where $F\left(s,B_v^0,\theta_0\right)=\int_0^s K\left(B_v^0(r),\theta_0\right)dr$, and $\psi\left(B_v^0,\theta_0,\kappa_{vu}\right)$ is defined in Theorem A.1.

(ii) If the assumptions of Theorem A.2 hold and $\tilde{\omega}_e^2$ is a consistent estimator of ω_e^2,

$$C_{LL}\Rightarrow\omega_e^{-2}\int_0^1\left[B_e(s)-F\left(s,B_v^0,\theta_0\right)'\chi\left(B_v^0,\theta_0\right)\right]^2 ds,$$

where $\chi\left(B_v^0,\theta_0\right)$ is defined in Theorem A.2.

Proof: The results of the lemma follow from the consistency of the estimators $\tilde{\omega}_u^2$ and $\tilde{\omega}_e^2$, Lemmas A.3 and A.4, and the continuous mapping theorem.

Proof of Theorem 1: Using equation (A.1), we obtain

$$\begin{aligned} b^{-1/2} \sum_{t=\mathbf{i}}^{[bs]+\mathbf{i}-1} \tilde{u}_t &= b^{-1/2} \sum_{t=\mathbf{i}}^{[bs]+\mathbf{i}-1} u_t - b^{-1} \sum_{t=\mathbf{i}}^{[bs]+\mathbf{i}-1} K\left(x_{tT}, \theta_0\right)' \sqrt{T}\left(\tilde{\theta}_T - \theta_0\right) \sqrt{\frac{b}{T}} \quad \text{(A.6)} \\ &+ T\left(\tilde{\theta}_T - \theta_0\right)' \left(b^{-3/2} \sum_{t=\mathbf{i}}^{[bs]+\mathbf{i}-1} \partial^2 g\left(x_{tT}, \dot{\theta}_T\right) / \partial\theta\partial\theta'\right) \left(\tilde{\theta}_T - \theta_0\right) \frac{b}{T}. \end{aligned}$$

Since $\frac{b}{T} \to 0$, the same arguments used to prove Lemma A.3 give

$$b^{-1/2} \sum_{t=\mathbf{i}}^{[bs]+\mathbf{i}-1} \tilde{u}_t \Rightarrow B_u\left(s\right),$$

which, along with the continuous mapping theorem and the consistency of $\tilde{\omega}_{i,u}^2$, yields part (i). Part (ii) holds by the same arguments.

As the preceding proof shows, the same result is obtained if $\sqrt{T}\left(\tilde{\theta}_T - \theta_0\right) = O_p\left(1\right)$ and if the standardized sums in the second and third terms on the right hand side of (A.6) are similarly of order $O_p\left(1\right)$. The proof can also be modified to allow for different orders of consistency for the estimator $\tilde{\theta}_T$ (see Park and Phillips (2001) for such consistency results in a special case of model (1)). Similar remarks apply to part (ii).

Appendix II: Derivation of the cumulative distribution function of $\int_0^1 W^2\left(s\right) ds$

The cumulative distribution function of $\int_0^1 W^2\left(s\right) ds$ can be derived by using the approach of Anderson and Darling (1952). A similar approach is taken in Abadir (1990). Since the characteristic function of $\int_0^1 W^2\left(s\right) ds$ is $\phi(t) = \left(\cosh\sqrt{-2it}\right)^{-1/2}$ (see White, 1958, and Evans and Savin, 1981), the Laplace transform of the probability density function of $\int_0^1 W^2\left(s\right) ds$ ($pdf(z)$) is given as

$$\xi(t) = \int_0^\infty \exp(-zt) pdf(z) dz = \phi(it) = \left(\cosh\sqrt{2it}\right)^{-1/2}.$$

Using this relation and integration by parts, we obtain

$$\int_0^\infty \exp(-zt)cdf(z)dz = \frac{\xi(t)}{t} = \frac{\left(\cosh\sqrt{2it}\right)^{-1/2}}{t},$$

where $cdf(z)$ denotes the cumulative distribution function of $\int_0^1 W^2(s)\,ds$. Now, $cdf(z)$ is obtained by using the the Laplace inverse transform and binomial expansion. That is,

$$\begin{aligned}
cdf(z) &= L^{-1}\left(\frac{\left(\cosh\sqrt{2it}\right)^{-1/2}}{t}\right) \\
&= L^{-1}\left(\frac{1}{t}\left(\frac{\exp(\sqrt{2t})}{2}\right)^{-1/2}\left(1+\exp(-2\sqrt{2t})\right)^{-1/2}\right) \\
&= L^{-1}\left(\frac{1}{t}\left(\frac{\exp(\sqrt{2t})}{2}\right)^{-1/2}\sum_{n=0}^{\infty}\frac{\Gamma(n+1/2)}{n!\Gamma(1/2)}(-1)^n\exp(-2n\sqrt{2t})\right) \\
&= \sqrt{2}\sum_{n=0}^{\infty}\frac{\Gamma(n+1/2)}{n!\Gamma(1/2)}(-1)^n L^{-1}\left(\frac{1}{t}\exp(-ut)\right) \\
&= \sqrt{2}\sum_{n=0}^{\infty}\frac{\Gamma(n+1/2)}{n!\Gamma(1/2)}(-1)^n\left(1-Erf(\frac{u}{2\sqrt{z}})\right),\ z\geq 0,
\end{aligned}$$

where $u=\frac{1}{\sqrt{2}}+2n\sqrt{2}$ and $L^{-1}(\cdot)$ denotes the Laplace inverse transform operator.

References

[1] Abadir, K. (1990) The limiting distribution of a functional of a Wiener process. Mimeo, Department of Economics, The American University of Cairo.

[2] Anderson, T.W. and D.A. Darling (1952) Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. The Annals of Mathematical Statistics 23, 193-212.

[3] Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

[4] Andrews, D.W.K. and C.J. McDermott (1995) Nonlinear econometric models with deterministically trending variables. Review of Economic Studies 62, 343–360.

[5] Balke, N.S. and T.B. Fomby (1997) Threshold cointegration. International Economic Review 38, 627–645.

[6] Bec, F. and A. Rahbek (2004) Vector equilibrium correction models with nonlinear discontinuous adjustments. Econometrics Journal 7, 628-651.

[7] Caner, M. and L. Kilian (2001) Size distortions of the null hypothesis of stationarity: evidence and implications for the PPP debate. Journal of International Money and Finance 20, 639–657.

[8] Chang, Y. and J.Y. Park (2003) Index models with integrated time series. Journal of Econometrics 114, 73–106.

[9] Chang, Y., J.Y. Park and P.C.B. Phillips (2001) Nonlinear econometric models with cointegrated and deterministically trending regressors. Econometrics Journal 4, 1–36.

[10] Choi, I. (2005) Subsampling vector autoregressive tests of linear constraints. Journal of Econometrics 124, 55–89.

[11] Choi, I. and B.C. Ahn (1995) Testing for cointegration in a system of equations. Econometric Theory 11, 952–983.

[12] Choi, I. and P. Saikkonen (2004) Testing linearity in cointegrating smooth transition regressions. Econometrics Journal 7, 341–365.

[13] Davidson, J. (1994) Stochastic Limit Theory. Oxford University Press: Oxford.

[14] Engle, R. and C.W.J. Granger (1987) Cointegration and error correction: representation, estimation and testing. Econometrica 55, 251–276.

[15] Evans, G.B.A. and E. Savin (1981) Testing for unit roots: 1. Econometrica 49, 753–779.

[16] Granger, C.W.J. (1981) Some properties of time series data and their use in econometric model specification. Journal of Econometrics 16, 121–130.

[17] Hansen, B.E. (1992) Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 489–500.

[18] Hansen, B.E. and B. Seo (2002) Testing for two-regime threshold cointegration in vector error correction models. Journal of Econometrics 110, 293–318.

[19] Hoffman, D.L. and R.H. Rasche (1991) Long-run income and interest elasticities of money demand in the United States. Review of Economics and Statistics 78, 665–674.

[20] Hoffman, D.L., R.H. Rasche and M. A. Tieslau (1995) The stability of long-run money demand in five industrial countries. Journal of Monetary Economics 35, 317–339.

[21] Jin, S. (2004) Discrete choice modeling with nonstationary panels applied to exchange rate regime choice. Mimeo, Department of Economics, Yale University.

[22] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? Journal of Econometrics 54, 159–178.

[23] Lütkepohl, H. T. Teräsvirta and J. Wolters (1999) Investigating stability and linearity of a German M1 demand function,.Journal of Applied Econometrics 14, 511–525.

[24] Neftci, S.A. (1984) Are economic time series asymmetric over the business cycle? Journal of Political Economy 92, 307–328.

[25] Park, J. and P.C.B. Phillips (1999) Asymptotics for nonlinear transformations of time series. Econometric Theory 15, 269–298.

[26] Park, J. and P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. Econometrica 69, 117–161.

[27] Patterson, K. (2000) An Introduction to Applied Econometrics: A Time Series Approach. MacMillan Press: London.

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

[29] Pötscher, B. and I.R. Prucha (1997) Dynamic Nonlinear Econometric Models. New York: Springer.

[30] Romano, J.P. and M. Wolf (2001) Subsampling intervals in autoregressive models with linear time trend. Econometrica 69, 1283–1314.

[31] Saikkonen, P. (1991) Asymptotically efficient estimation of cointegration regressions. Econometric Theory 7, 1–21.

[32] Saikkonen, P. (2008) Stability of regime switching error correction models under linear cointegration. Econometric Theory 24, 294-318.

[33] Saikkonen, P. and I. Choi (2004) Cointegrating smooth transition regressions. Econometric Theory 20, 301–340.

[34] Shin, Y. (1994) A residual-based test of the null of cointegration against the alternative of no cointegration. Econometric Theory 10, 91–115.

[35] Stock, J.H. and M.W. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica 61, 783–820.

[36] White, J. (1958) The limiting distribution of the serial correlation in the explosive case. Annals of Mathematical Statistics 29, 1188–1197.

Table 1: Empirical size and power of subresidual-based tests:
Polynomial regression model

Notes: 1. Data were generated by using $y_t = x_t + x_t^2 + u_t,\ (t = -29, ..., T)$ with $u_t = \alpha u_{t-1} + \varepsilon_t$ and $\begin{pmatrix} \Delta x_t \\ \varepsilon_t \end{pmatrix} \sim iid\ N\left(0, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$. 2. The number of replications is 3,000. 3. The minimum volatility rule was used for the choice of block size. 4. Lag length for the long-run variance estimation is denoted here as l.

(1) $l = [4(b/100)^{0.25}]$

		5% level				10% level			
			$C_{LL}^{b,\max}$		$C_{NLLS}^{b,\max}$		$C_{LL}^{b,\max}$		$C_{NLLS}^{b,\max}$
	T	$K=1$	$K=2$	$K=3$		$K=1$	$K=2$	$K=3$	
	150	0.2	0.2	0.2	0.6	0.9	0.8	0.7	1.7
$\alpha = 0.5$	300	0.2	0.3	0.3	0.4	0.7	0.7	0.8	1.3
	600	0.2	0.2	0.3	0.2	0.9	0.6	0.6	1.1
	150	4.7	4.0	3.3	6.2	8.2	7.9	6.8	11.8
$\alpha = 0.8$	300	3.1	3.2	2.7	4.7	6.0	5.4	5.0	8.4
	600	4.1	3.7	3.2	4.9	6.5	5.7	5.6	8.1
	150	31.0	29.4	28.0	34.9	39.0	37.1	35.5	42.8
$\alpha = 0.95$	300	32.0	30.7	29.5	33.7	39.3	38.0	36.8	40.6
	600	32.5	31.7	30.5	33.8	39.9	39.3	38.2	41.4
	150	49.5	47.7	45.3	52.7	56.8	55.2	52.7	59.8
$\alpha = 1.0$	300	64.1	62.8	61.2	65.2	68.9	68.5	67.0	70.4
	600	79.9	79.4	79.0	80.6	84.5	84.4	83.4	85.1

(2) $l = [12(b/100)^{0.25}]$

		5% level				10% level			
		$C_{LL}^{b,\max}$			$C_{NLLS}^{b,\max}$	$C_{LL}^{b,\max}$			$C_{NLLS}^{b,\max}$
	T	$K=1$	$K=2$	$K=3$		$K=1$	$K=2$	$K=3$	
$\alpha=0.5$	150	0.0	0.0	0.0	0.0	0.2	0.1	0.1	0.3
	300	0.0	0.0	0.0	0.1	0.2	0.2	0.3	0.6
	600	0.0	0.0	0.0	0.0	0.2	0.1	0.1	0.4
$\alpha=0.8$	150	0.0	0.0	0.0	0.0	0.5	0.4	0.3	0.6
	300	0.2	0.2	0.2	0.2	0.8	0.8	0.8	0.9
	600	0.3	0.2	0.2	0.3	1.1	0.9	1.0	1.2
$\alpha=0.95$	150	0.7	0.5	0.5	0.7	4.6	4.2	3.7	4.9
	300	3.4	3.2	2.8	3.3	8.9	8.4	8.2	9.5
	600	5.2	5.3	5.3	5.6	10.9	10.3	10.2	11.9
$\alpha=1.0$	150	3.0	2.6	2.3	3.6	11.9	12.0	10.1	13.3
	300	20.3	19.8	19.2	21.1	35.1	33.8	32.7	36.0
	600	48.9	48.3	47.7	49.8	59.7	59.7	58.7	61.4

Table 2: Empirical size and power of subresidual-based tests:
Smooth transition regression model

Notes: 1. Data were generated by using $y_t = x_t + g(x_t) x_t + u_t$, $(t = -29, ..., T)$ with $u_t = \alpha u_{t-1} + \varepsilon_t$, $g(x_t) = \frac{1}{1+\exp(-(x_t - 5))}$ and $\begin{pmatrix} \Delta x_t \\ \varepsilon_t \end{pmatrix} \sim iid\ N\left(0, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$. 2. The number of replications is 500, 300 and 300 at T=150, 300, 600, respectively. 3. The minimum volatility rule was used for the choice of block size. 4. Lag length for the long-run variance estimation is denoted here as l.

(1) $l = [4(b/100)^{0.25}]$

		5% level				10% level			
		$C_{LL}^{b,\max}$			$C_{NLLS}^{b,\max}$	$C_{LL}^{b,\max}$			$C_{NLLS}^{b,\max}$
	T	$K=1$	$K=2$	$K=3$		$K=1$	$K=2$	$K=3$	
	150	0.6	0.1	0.1	0.1	1.6	2.2	2.6	1.6
$\alpha = 0.5$	300	0.0	0.3	0.7	0.0	0.3	0.3	0.7	0.7
	600	0.7	0.3	0.7	1.7	1.3	1.0	0.7	1.7
	150	6.0	6.4	8.2	5.0	9.4	11.0	11.8	9.6
$\alpha = 0.8$	300	4.3	4.0	2.3	6.3	7.7	6.7	6.7	10.3
	600	3.0	2.0	3.0	2.0	4.7	3.3	5.3	5.0
	150	24.0	24.2	24.0	28.2	31.2	30.0	30.8	36.2
$\alpha = 0.95$	300	20.0	20.7	20.7	19.7	28.0	26.7	28.3	28.3
	600	34.3	32.3	34.0	34.3	40.0	39.0	41.3	41.0
	150	28.0	29.8	31.2	42.8	35.0	36.2	35.0	51.8
$\alpha = 1.0$	300	47.3	47.3	47.3	64.0	54.0	55.3	54.0	71.0
	600	53.7	55.0	53.7	79.7	57.0	58.0	56.0	84.0

(2) $l = [12(b/100)^{0.25}]$

	T	5% level $C_{LL}^{b,\max}$ $K=1$	$K=2$	$K=3$	$C_{NLLS}^{b,\max}$	10% level $C_{LL}^{b,\max}$ $K=1$	$K=2$	$K=3$	$C_{NLLS}^{b,\max}$
$\alpha = 0.5$	150	0.0	0.0	0.0	0.0	0.8	0.4	0.6	0.0
	300	0.0	0.0	0.0	0.0	0.0	0.3	0.3	0.0
	600	0.0	0.0	0.0	0.3	1.3	0.7	1.0	1.3
$\alpha = 0.8$	150	0.4	0.4	0.8	0.2	1.8	1.6	2.0	1.4
	300	0.0	0.0	0.0	0.0	0.3	0.3	0.0	0.0
	600	0.3	0.3	0.0	0.0	0.7	0.7	1.0	0.0
$\alpha = 0.95$	150	0.4	0.6	0.2	0.2	3.0	3.4	3.4	4.8
	300	2.0	3.0	5.3	1.3	6.0	6.0	7.3	5.3
	600	4.7	4.3	2.3	5.7	10.0	9.3	10.0	9.7
$\alpha = 1.0$	150	1.6	1.4	1.0	1.8	7.2	6.4	4.6	9.8
	300	11.7	11.7	11.3	21.0	25.0	24.3	22.7	37.7
	600	37.0	38.0	35.0	52.3	43.3	43.7	40.3	61.0

Table 3: Empirical size and power of subresidual-based tests:
Linear regression model

Notes: 1. Data were generated by using $y_t = x_t + u_t$, $(t = -29, ..., T)$ with $u_t = \alpha u_{t-1} + \varepsilon_t$ and $\begin{pmatrix} \Delta x_t \\ \varepsilon_t \end{pmatrix} \sim iid\ N\left(0, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$. 2. The number of replications is 3,000. 3. The minimum volatility rule was used for the choice of block size. 4. Lag length for the long-run variance estimation is denoted here as l.

(1) $l = [4(b/100)^{0.25}]$

		5% level				10% level			
		$C_{LL}^{b,\max}$			$C_{NLLS}^{b,\max}$	$C_{LL}^{b,\max}$			$C_{NLLS}^{b,\max}$
	T	$K=1$	$K=2$	$K=3$		$K=1$	$K=2$	$K=3$	
	150	1.0	1.0	1.0	1.6	2.4	2.2	2.0	3.3
$\alpha = 0.5$	300	0.7	0.8	0.8	1.3	2.0	2.0	1.8	2.9
	600	0.8	0.7	0.6	1.2	1.5	1.5	1.4	2.1
	150	9.3	7.6	7.5	11.4	14.1	12.2	11.9	17.0
$\alpha = 0.8$	300	6.7	5.9	5.3	8.4	11.1	10.3	9.9	14.4
	600	6.3	5.7	5.4	7.7	10.0	9.6	9.0	12.4
	150	42.4	40.2	38.3	45.9	50.6	49.2	46.8	54.7
$\alpha = 0.95$	300	44.9	43.4	41.4	46.7	53.6	52.3	50.4	55.8
	600	42.9	41.3	41.0	44.7	51.0	49.8	49.2	53.0
	150	61.8	59.4	57.5	64.8	68.2	66.4	64.7	70.8
$\alpha = 1.0$	300	74.7	73.3	72.3	75.2	80.3	79.2	77.7	80.6
	600	86.5	86.0	85.4	87.1	91.9	91.5	91.0	92.1

(2) $l = [12(b/100)^{0.25}]$

		5% level				10% level			
		$C_{LL}^{b,\max}$			$C_{NLLS}^{b,\max}$	$C_{LL}^{b,\max}$			$C_{NLLS}^{b,\max}$
	T	$K=1$	$K=2$	$K=3$		$K=1$	$K=2$	$K=3$	
$\alpha = 0.5$	150	0.1	0.1	0.1	0.1	0.4	0.3	0.3	0.4
	300	0.1	0.0	0.0	0.1	0.2	0.3	0.3	0.6
	600	0.1	0.2	0.1	0.1	0.5	0.5	0.3	0.7
$\alpha = 0.8$	150	0.2	0.2	0.1	0.3	1.2	1.1	1.0	1.8
	300	0.6	0.5	0.4	0.7	1.7	1.3	1.1	2.1
	600	0.6	0.6	0.5	0.6	1.7	1.6	1.6	2.2
$\alpha = 0.95$	150	3.4	3.3	2.8	3.7	11.6	10.9	9.9	12.8
	300	10.4	10.1	9.4	11.2	19.7	19.0	18.2	21.0
	600	11.8	10.6	10.3	12.2	19.0	17.9	17.6	19.9
$\alpha = 1.0$	150	12.0	10.7	10.7	13.0	26.6	25.5	24.4	28.6
	300	40.3	40.0	38.5	40.5	53.0	52.7	50.5	53.8
	600	66.1	65.3	64.7	67.1	71.7	70.8	70.3	71.9

Table 4: Empirical size and power of the KPSS tests:
Linear regression model

Note: 1. Data were generated by using $y_t = x_t + u_t,\ (t = -29, ..., T)$ with $u_t = \alpha u_{t-1} + \varepsilon_t$ and $\begin{pmatrix} \Delta x_t \\ \varepsilon_t \end{pmatrix} \sim iid\ N\left(0, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$. 2. KPSS (LL) denotes the KPSS test using the leads-and-lags regression, and KPSS (OLS) that using the OLS regression. 3. The number of replications is 3,000. 4. Lag length for the long-run variance estimation is denoted here as l.

(1) $l = [4(T/100)^{0.25}]$

		5% level				10% level			
		$KPSS\ (LL)$			$KPSS\ (OLS)$	$KPSS\ (LL)$			$KPSS\ (OLS)$
	T	$K=1$	$K=2$	$K=3$		$K=1$	$K=2$	$K=3$	
$\alpha = 0.5$	150	11.1	10.2	10.2	17.0	21.2	19.5	18.7	28.8
	300	9.7	9.5	9.4	13.2	17.4	17.1	17.5	24.0
	600	9.3	9.0	9.0	13.9	17.2	17.3	17.1	24.3
$\alpha = 0.8$	150	35.2	33.3	31.8	42.5	52.3	50.2	47.6	59.6
	300	31.2	29.4	28.2	36.7	46.7	45.1	43.9	53.8
	600	28.0	26.3	25.5	34.2	43.1	40.9	39.5	49.2
$\alpha = 0.95$	150	78.8	77.3	75.1	80.7	88.9	87.7	86.2	90.8
	300	83.0	81.9	81.2	85.0	92.7	92.1	91.3	93.8
	600	83.3	82.4	81.9	84.7	92.6	92.0	92.0	93.6
$\alpha = 1.0$	150	89.9	88.9	88.1	91.2	95.8	95.4	95.0	96.7
	300	96.4	96.2	95.9	96.6	98.6	98.5	98.4	98.8
	600	99.3	99.3	99.3	99.4	99.8	99.8	99.8	99.9

(2) $l = [12(T/100)^{0.25}]$

		5% level				10% level			
		$KPSS\ (LL)$			$KPSS\ (OLS)$	$KPSS\ (LL)$			$KPSS\ (OLS)$
	T	$K=1$	$K=2$	$K=3$		$K=1$	$K=2$	$K=3$	
	150	6.9	6.3	6.3	11.0	15.4	14.4	13.3	21.0
$\alpha = 0.5$	300	6.4	6.3	6.2	8.5	12.7	12.3	11.9	17.3
	600	7.0	6.7	6.4	9.7	13.0	12.7	12.6	18.6
	150	12.2	11.8	11.3	14.8	23.4	22.1	21.4	28.3
$\alpha = 0.8$	300	12.3	11.6	10.7	14.2	21.8	20.2	19.4	26.4
	600	12.1	11.4	10.8	14.0	21.2	19.9	18.8	25.3
	150	34.8	32.8	31.3	36.8	51.7	49.7	47.7	54.4
$\alpha = 0.95$	300	39.1	38.4	37.4	40.9	56.2	55.8	54.5	58.1
	600	40.3	39.3	38.6	41.8	57.2	56.8	55.8	58.8
	150	53.7	51.9	49.9	56.0	69.5	68.4	66.7	71.1
$\alpha = 1.0$	300	74.1	73.4	72.4	75.3	85.4	84.8	84.1	85.9
	600	89.0	89.0	88.7	89.5	94.9	94.8	94.7	95.2

Table 5: Cointegration test results for the U.S. money demand equation

Notes: 1. Data were taken from the International Financial Statistics. The sampling period is 1959:Q1–2000:Q4, and the sample size is 168.
2. A smooth transition model

$$\begin{aligned} y_t &= \theta_0 + \theta_1 x_{1t} + \theta_2 x_{2t} + \theta_3 g\left(x_{2t}\right) x_{2t} + u_t; \\ g(x_t) &= \frac{1}{1+\exp(-\theta_4(x_{2t}-\theta_5))}, \end{aligned}$$

with $y = \ln(\text{M1}) - \ln(\text{GDP deflator})$, $x_1 = \ln\ (\text{GDP}) - \ln(\text{GDP deflator})$, $x_2 = \ln(\text{T-bill rate})$, was used.
3. Lag length for the long-run variance estimation is $[4(b/100)^{0.25}]$.
4. P-values were calculated using the cdf of Subsection 4.2, $\int_0^1 W^2\left(s\right) ds$.
5. For the minimum volatility rule, we used $m = 2$.
6. M denotes the number of subresidual-based tests used for the Bonferroni procedure.

	$C_{LL}^{b,\max}$			$C_{NLLS}^{b,\max}$
	$K=1$	$K=2$	$K=3$	
P-value	0.0604	0.1029	0.0782	0.2946
Block size	67	76	76	79
M	3	3	3	3

Table 6: Smooth transition regression results for the U.S. money demand equation

Notes: Numbers in parentheses are standard errors, the square root of the long-run variance estimated with Andrews' (1991) methods using an AR(4) approximation for the prefilter. In addition, the notes for Table 4 apply here.

	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5
LL ($K = 1$)	-0.0463	0.2391	-0.0743	-0.0169	20.68	2.193
	(0.0670)	(0.0185)	(0.0331)	(0.0192)	(31.83)	(0.2028)
LL ($K = 2$)	-0.0449	0.2424	-0.0785	-0.0196	15.16	2.211
	(0.0690)	(0.0189)	(0.0334)	(0.0198)	(33.01)	(0.2054)
LL ($K = 3$)	-0.0455	0.2451	-0.0818	-0.0205	11.20	2.211
	(0.0723)	(0.0194)	(0.0340)	(0.0201)	(33.61)	(0.2056)
NLLS	-0.0629	0.2319	-0.0548	-0.0184	12.83	2.151